Auditory feedback for automated driving

Bazilinskyy, Pavlo

DOI
10.4233/uuid:51dbba63-ba7c-4958-ab7d-d6dc182ec5b5

Publication date
2018

Document Version
Final published version

Citation (APA)
Bazilinskyy, P. (2018). *Auditory feedback for automated driving*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:51dbba63-ba7c-4958-ab7d-d6dc182ec5b5

Auditory feedback for automated driving

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.dr.ir. Tim van der Hagen;
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
vrijdag 14 december 2018 om 10:00 uur

door

Pavlo BAZILINSKYY
Erasmus Mundus Double MSc in Dependable Software Systems,
University of St Andrews, UK / Maynooth University, Ireland
geboren te Chernihiv, Oekraïne

Dit proefschrift is goedgekeurd door de
promotoren: Dr.ir. J. C. F. de Winter en Prof.dr. F. C. T. van der Helm

Samenstelling promotiecommissie bestaat uit:

Rector Magnificus	voorzitter
Prof.dr. F. C. T. van der Helm	TU Delft, *promotor*
Dr.ir. J. C. F. de Winter	TU Delft, *promotor*

onafhankelijke leden:

Dr. J. Hickman	Virginia Tech Transportation Institute
Prof.dr. M. H. Martens	Universiteit Twente
Dr. W. P. Vlakveld	SWOV
Prof.dr.ir. J. W. C. van Lint	TU Delft
Dr.ir. M. Mulder	TU Delft

The research presented in this dissertation was conducted in the project HFAuto – Human Factors of Automated Driving (PITN-GA-2013-605817) in the Marie Curie Initial Training Network. The contributions described in this thesis were made as a part of Work Package 2 of HFauto, the focus of which was to develop a human-machine interface supporting the operator of the future automated vehicle.

To my muse.
To Kateryna, Olena, Yevheniy, Oleksandr, Yakov, Jessey.

Contents

Summary

Automated driving may be a key to solving a number of problems that humanity faces today: large numbers of fatalities in traffic, traffic congestions, and increased gas emissions. However, unless the car drives itself fully automatically (such a car would not need to have a steering wheel, nor accelerator and brake pedals), the driver needs to receive information from the vehicle. Such information can be delivered by sound, visual displays, vibrotactile feedback, or a combination of two or three kinds of signals. Sound may be a particularly promising feedback modality, as sound can attract a driver's attention irrespective of his/her momentary visual attention.

Although ample research exists on warning systems and other types of auditory displays, what is less well known is how to design warning systems for automated driving specifically.

Taking over control from an automated car is a spatially demanding task that may involve a high level of urgency, and warning signals (also called 'take-over requests', TORs) need to be designed so that the driver reacts as quickly and safely as possible. Furthermore, little knowledge is available on how to support the situation awareness and mode awareness of drivers of automated cars. The goal of this thesis is to discover how the auditory modality should be used during automated driving and to contribute towards the development of design guidelines.

First, this thesis describes the state-of-the-art (Chapter 2) by examining and improving the current sound design process in the industry, and by examining the requirements of the future users of automated cars, the public (Chapter 2). Next, the thesis focuses on the design of discrete warnings/TORs (Chapter 3), the use of sound for supporting situation awareness (Chapter 4), and mode awareness (Chapter 5). Finally, Chapters 6 and 7 provide a future outlook, conclusions, and recommendations. The content of the thesis is described in more detail below.

Chapter 2 describes state of the art in the domain of the use of sound in the automotive industry. Section 2.1 presents a new sound design process for the automotive industry developed with Continental AG, consisting of 3 stages: description, design/creation, and verification. An evaluation of the process showed that it supports the more efficient creation of auditory assets than the unstructured process that was previously employed in the company.

To design good feedback is not enough, it also needs to be appreciated by users. To this end, Section 2.2 describes a crowdsourced online survey that was used to investigate peoples' opinion of 1,205 responses from 91 countries on auditory interfaces in modern cars and their readiness to have auditory feedback in automated vehicles. The study was continued in another crowdsourced online survey described in Section 2.3, where 1,692 people were surveyed on auditory, visual, and vibrotactile TORs in scenarios of varying levels of urgency. Based on the results, multimodal TORs were the most preferred option in scenarios associated with high urgency. Sound-based TORs were the most favored choice in scenarios with low urgency. Auditory feedback was also preferred for confirmation that the system is ready to switch from manual to automated mode. Speech-based feedback was more accepted than artificial

sounds, and the female voice was more preferred than the male voice as a take-over request.

To understand better how sound may be used during fully automated driving, it is crucial to acknowledge the opinion of potential end users of such vehicles on the technology. Section 2.4 investigates anonymous textual comments concerning fully automated driving by using data from three Internet-based surveys (including the surveys described in Sections 2.2 and 2.3) with 8,862 respondents from 112 countries. The opinion was split: 39% of the comments were positive towards automated driving and 23% were seen as such that express negative attitude towards automated driving.

Chapter 3 focuses on the use of the auditory modality to support TORs. Section 3.1 describes a crowdsourcing experiment on reaction times to audiovisual stimuli with different stimulus onset asynchrony (SOA). 1,823 participants each performed 176 reaction time trials consisting of 29 SOA levels and three visual intensity levels. The results replicated past research, with a V-shape of mean reaction time as a function of SOA. The study underlines the power of crowdsourced research, and shows that auditory and visual warnings need to be provided at exactly the same moment in order to generate optimally fast response times. The results also indicate large individual differences in reaction times to different SOA levels, a finding which implicates that multimodal feedback has important advantages as compared to unimodal feedback.

Then, in Section 3.2 focus was given to speech-based TORs. In a crowdsourced study, 2,669 participants from 95 countries listened to a random 10 out of 140 TORs, and rated each TOR on ease of understanding, pleasantness, urgency, and commandingness. Increased speech rate results in an increase of perceived urgency and commandingness. With high level of background noise, the female voice was preferred over the male voice, which contradicts the literature. Furthermore, a take-over request spoken by a person with Indian accent was easier to understand by participants from India compared to participants from other countries.

The results of the studies in Chapter 2 and Sections 3.1 and 3.2 were used to design a simulator-based study presented in Section 3.3. 24 participants took part in three sessions in a highly automated car (different TOR modality in each session: auditory, vibrotactile, and auditory-vibrotactile). TORs were played from the right, from the left, and from both left and right. The auditory TOR yielded comparatively low ratings of usefulness and satisfaction. Regardless of the directionality of the TOR, almost all drivers overtook the stationary vehicle on the left.

Section 3.4 summarizes results from survey research (Sections 2.2, 2.3, 3.1, 3.2) and driving simulator experiments (including Section 3.3) on TORs executed with one or multiple of the three modalities. Results showed that vibrotactile TORs in the driver's seat yielded relatively high ratings of self-reported usefulness and satisfaction. Auditory TORs in the form of beeps were regarded as useful but not satisfactory, and it was found that an increase of beep rate yields an increase of self-reported urgency. Visual-only feedback in the form of LEDs was seen by participants as neither useful nor satisfactory.

Chapter 4 draws attention to the use of auditory feedback for the situation awareness during manual and automated driving. Section 4.1 investigates how to represent distance information by means of sound. Three sonification approaches were tested: Beep Repetition Rate, Sound Intensity, and Sound

Fundamental Frequency. The three proposed methods produced a similar mean absolute distance error.

These results were used in three simulator-based experiments (Sections 4.2–4.4) to examine the idea whether it is possible to drive a car blindfolded with the use of continuous auditory feedback only. Different types of sonification (e.g., volume-based, beep-frequency based) were used, and the auditory feedback was provided when deviating more than 0.5 m from lane center. In all experiments, people drove on a track with sharp 90-degree corners while speed control was automated. Results showed no clear effects of sonification method on lane-keepign performance, but it was found that it is vital to not give feedback based on the current lateral position, but based on where the car will be about 2 seconds into the future. The predictor algorithm should consider the velocity vector of the car as well as the momentary steering wheel angle. Results showed that, with extensive practice and knowledge of the system, it is possible to drive on a track for 5 minutes without leaving the road. Drivers benefit from simple auditory feedback and additional stimuli add workload without improving performance.

Chapter 5 examines the use of sound for mode awareness during highly automated driving. An on-road experiment in a heavy truck equipped with low-level automated is described. I used continuous auditory feedback on the status of ACC, lane offset, and headway, which blends with the engine and wind sounds that are already present in the cabin. 23 truck drivers were presented with the additional sounds in isolation and in combination. Results showed that the sounds were easy to understand and that the lane-offset sound was regarded as somewhat useful. However, participants overall preferred a silent cabin and expressed displeasure with the idea of being presented with extra sounds on a continuous basis.

Chapter 6 provides an outlook on when fully automated driving may become a reality. In 12 crowdsourcing studies conducted between 2014 and 2017 (including the studies described in Sections 2.2, 2.3, 3.1, 3.2), 17,360 people from 129 countries were asked when they think that most cars will be able to drive fully automatically in their country of residence. The median reported year was 2030. Over the course of three years respondents have moderated their expectations regarding the penetration of fully automated cars. The respondents appear to be more optimistic than experts.

Chapter 7 presents a discussion and conclusions derived from all chapters in the thesis.

- The most preferred way to support a TOR is an auditory instruction in the form of a female voice.
- The preferences of people depend on the urgency of the situation.
- Reaction times are fastest when an auditory and a visual stimulus are presented at the same moment rather than with a temporal asynchrony.
- An increase of beep rate yields an increase of self-reported urgency.
- An increase in the speech rate results in an increase of perceived urgency and commandingness.
- If the goal is for drivers to react as quickly as possible, multimodal feedback should be used.

- It is important to use a preview controller (look-ahead time) for supporting drivers' situation awareness in a lane keeping task.
- Truck drivers are not favorable towards adding additional continuous feedback to the cabin, even though the feedback is easy to understand.

In summary, in this thesis I evaluated the use of sound as discrete warnings, but also as a means of continuous/spatial support for situation/mode awareness.

1 INTRODUCTION

This introduction gives a brief overview of the research gap that this thesis addresses. A more elaborate review of the literature can be found in Chapter 2 'State of the art'.

1.1 Motivation

The majority of traffic accidents (the scientific community has an on-going debate about the use of the words 'accident' versus 'crash' e.g., Blanchard et al., 2003; I use the words 'accident' and 'crash' through the manuscript) are caused by human error (National Highway Traffic Safety Administration, 2008). Automated driving has the potential to drastically reduce the number of fatalities on the roads. Moreover, the way people use cars today is not efficient. There are too many cars on the roads, which leads to traffic congestions, increased gas emissions, and fuel consumption.

Experts are speculating about the date when fully automated cars ('level 5 automation'; SAE International, 2014) will become available to the general public. It is not clear yet what the role of human-machine interfaces (HMIs) will be in such vehicles. Drivers in fully automated cars will not need to be warned about critical events, as, by definition, such a car must be able to handle all situations without any involvement from the driver. Instead, in fully automated cars, in-vehicle feedback may be tailored towards infotainment and entertainment.

Before fully automated cars become widespread, conditional ('level 3') and high ('level 4') automation will most likely be introduced. At these lower levels of automation, the automated driving system is not able to handle all situations. Therefore, the driver is sometimes requested to take back control. A warning issued when such a situation arises is called a take-over request (TOR). When the automated vehicle relinquishes control to the driver, the driver who was previously performing a non-driving task (e.g., reading a book) needs to build up situation awareness. In other words, the driver quickly 'needs to know what is going on so he/she can figure out what do to' (Adam, 1993). Accordingly, the driver could benefit from receiving feedback about the situation outside of the vehicle, and about whether the car is currently in the manual or automated mode (i.e., to facilitate the driver's mode awareness).

TORs during highly automated driving can be conveyed by auditory, visual, or vibrotactile displays. Of these modalities, auditory feedback has several important characteristics that make it suitable to be used as a warning system: (1) auditory information can be received at almost all times; (2) it is omnidirectional, such feedback can be transmitted from any direction; (3) humans can focus on one sound among multiple streams of sound; (4) sound is transient, sound is only available at that moment of creation (Bregman, 1990; Cooke & Ellis, 2001; Hermann, Hunt & Neuhoff, 2011; Wickens et al., 2012). Consequently, auditory feedback is used in a large variety of applications,

especially when there is a need to alert the user or when visual load needs to be avoided.

Auditory feedback can be both speech and non-speech based. Auditory feedback in combination with visual and/or vibrotactile feedback can also be a part of a multimodal display. Compared to unimodal displays, multimodal displays can generate more information in the same amount of time, which may result in better performance.

1.2 Research gap

Although ample research exists on warning systems and other types of auditory displays, what is less well known is how to design warning systems for automated driving specifically. Furthermore, little knowledge is available on how to support the situation awareness and mode awareness of drivers of automated cars. Crucial differences between auditory feedback for traditional applications (e.g., control rooms, aircraft) and automated driving are the following:

- In automated driving, situations can be highly urgent. For example, a TOR may be provided a few seconds before a collision, such as a situation where another vehicle is stranded on the road right before the driver. Even a few tenths of a second faster response can make the difference between crashing and not crashing, and may even save human lives. In contrast, in control rooms or aviation, warnings are usually meant to indicate to the operator that something is amiss and that a corrective action needs to be taken, but the operator still has multiple seconds or even minutes to respond and intervene. The high urgency in automated driving means that special care should be taken regarding the design of TORs, and critical questions should be answered, such as: (1) What should be the inter-stimulus interval (e.g., beep rate)? (2) Should one warning (e.g., auditory warning) or two combined warnings (e.g., vibrotactile-auditory) be provided? and (3) Should such multimodal stimuli be provided at the same moment or not? Here, the goal should be that drivers respond as quickly and safely as possible in safety-critical situations.
- In automated driving, the task is spatially demanding. For example, when the driver reclaims control of the vehicle, he/she has to take into account the position in the lane and usually perform a steering or braking action. (Continuous) information about spatially proximate objects and deviation from the lane centre could be mapped (i.e., sonified) to the driver. It is currently unknown how sonification should be applied for supporting drivers' situation awareness and decision-making.
- In automated driving, the user group is highly diverse. In contrast, in aviation or control rooms, operators are usually highly trained specialists for whom it is unlikely that the operators will misuse or ignore warning aids. Accordingly, in automated driving, it has to be ascertained that a variety of drivers accept the HMI. For example, it has to be examined whether drivers find specific characteristics of the warning (e.g., speaking accents, speaking tempo, speaking gender, beep rate, continuous vs. discrete warnings) pleasant and easy-to-understand. If drivers

reject/disuse or misunderstand the feedback from an HMI, the eventual effects on safety are likely to be negative.

1.3 Research goal

As indicated above, it is currently not known how sound should be used during automated driving. Although auditory warnings are already commonplace in e.g., control rooms, automated driving places special demands on the operator. Accordingly, the goal of this thesis is to understand how the auditory modality should be used during automated driving, equipped with either high or full automation, and to contribute towards the development of design guidelines. The use of the auditory and multimodal displays for supporting TORs during highly automated driving is the main focus of this work. Additionally, the guidelines on the use of sound for situation and mode awareness are given.

In this research, a spectrum of methods is used, ranging from online experiments and surveys (allowing for large-sample research), a driving simulator experiment (allowing for controlled research in an immersive environment), as well as an on-road study (allowing for testing concepts in a real environment among end users).

1.4 Thesis outline

Figure 1 shows the structure of the thesis. First, the current state of the use of sound in the automotive industry is investigated. Then, the topic of the use of discrete auditory feedback for supporting TORs during highly automated driving is highlighted. The focus is then switched to the use of continuous auditory feedback for situation and mode awareness. The chapters of this thesis are briefly introduced below.

Chapter 2 discusses the state of the art of the use of both auditory and multimodal feedback in in-vehicle interfaces. Section 2.1 describes the sound design process in the automotive industry and a way to improve it by introducing a software tool which structures the design process between client and supplier. This section does not necessarily have strong empirical value, but focuses on the design and utilization of my research in the industry. Conducting such a study, where I could enter the industry and assist a company with the design process was my obligation within my PhD. Then, in Sections 2.2 and 2.3, two online crowdsourced surveys are presented. They were conducted to gather the opinion of the public on auditory, visual, vibrotactile, and multimodal feedback for both highly and fully automated driving. The results outlined that auditory and multimodal feedback are attractive mediums for communication with people inside of a highly automated car. Conducting such online surveys allowed me to gather the requirements for the driving simulator and on-road studies presented in subsequent chapters. In Section 2.4, to receive a better insight into the requirements of the public towards future vehicles, a study on the opinion of people on fully automated driving is presented.

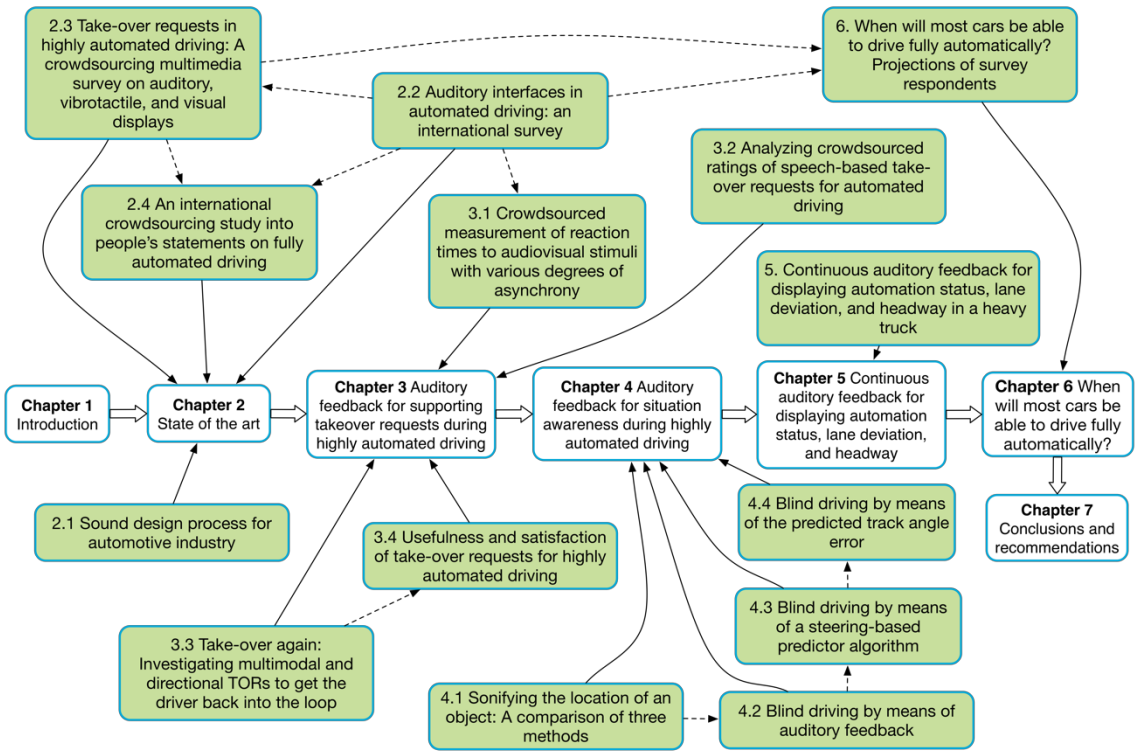


Figure 1. Outline of the thesis. White boxes show chapters of the thesis. Green boxes represent single studies within the work presented in the thesis, which are described in the form of a journal article or a conference paper. The numbers in boxes indicate the section or chapter of the thesis. Dotted lines represent the connections between studies, that is, when results and conclusions of one study were used in the preparation of another piece of work. Solid lines direct to a chapter of the thesis that each such study contributed to.

Chapter 3 introduces research that was conducted on the use of discrete sound for TORs during highly automated driving. The previous chapter outlines the importance of multimodal feedback for supporting highly automated driving. However, a number of questions need to be answered before designing such a feedback. Section 3.1 presents a fundamental study on measurement of reaction time to audio, visual, and audiovisual stimuli. In this section, attention is given to the intensity of a multimodal TOR and to whether the auditory and visual components of such feedback need to be presented simultaneously. Section 3.2 presents another crowdsourced experiment, where a large sample was asked to rank a selection of speech-based TORs on their urgency, commandingness, pleasantness, and ease of understanding. This study gives insight into the importance of gender, accent, and used phrase in a speech-based TOR. The use of crowdsourcing in this study allowed to replicate past findings with a much larger sample size, giving a clearer picture of the effects of the independent variables on participants' reaction times. Chapter 3 concludes with Section 3.4, which summarises results from survey research and driving simulator experiments conducted within the Work Package 2 of the HFAuto project on auditory, vibrotactile, and visual TORs in highly automated driving.

Following the discussion of the use of discrete sound in the previous chapter, *Chapter 4* discusses the use of continuous auditory feedback for situation awareness during highly automated driving. Physical processes in our world are essentially continuous and situations that arise during highly

automated driving develop in a continuous manner as well. To understand what type of artificial sound is most suitable to support continuous feedback, Section 4.1 studies three sonification approaches, where object distance information is mapped to a sound dimension. Next, a series of driving simulator studies in Sections 4.2–4.4 examine the idea whether it is possible to drive a car blindfolded with the use of auditory feedback only. The sections should be read as one research line and not as three separate studies. They showcase the power of research with a small number of participants but a large number of trials per participant (as was also documented by Smith & Little, 2018). These studies were conducted to put the auditory feedback to an ultimate test. It was examined whether a driver can keep a car in the lane with just auditory feedback ('blind driving'). These tests give insight into the usability of feedback for supporting a driver during highly automated driving in case of the absence of visual information.

Chapter 5 follows the discussion about continuous feedback and draws attention to the use of such sound for mode awareness during highly automated driving. The topic is examined in an on-road experiment on the use of continuous auditory feedback for the status of Automatic Cruise Control (ACC), lane deviation, and headway in a heavy truck. The continuous feedback presented in the study is based on the results of the more fundamental studies described in Chapter 4.

The results in Chapters 2–4 explore how sound could be used in future automated cars. However, the question still remains if/when (fully) automated cars will become widespread.

In *Chapter 6* the emphasis is on the question of when fully automated driving may become a reality, in particular, on when fully automated cars may be deployed. When fully automated cars will be available to the general public is a question that has attracted attention from futurists, car manufacturers, and academics. This question was asked to the public in a crowdsourced online survey, and the results of the study are explored in the chapter.

Chapter 7 presents a discussion and conclusions derived from all chapters in the thesis.

1.5 References

Adam, E. C. (1993). Fighter cockpits of the future. Proceedings of *the 12th IEEE/AIAA Digital Avionics Systems Conference (DASC)*, Fort Worth, TX, 318–323.

Blanchard, E. B., Hickling, E. J., & Kuhn, E. (2003). Of "crashes" and "accidents," a comment on Stewart and Lord. *Journal of Traumatic Stress,* 16, 527–528.

Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. Cambridge, MA: MIT Press.

Cooke, M., & Ellis, D. P. W. (2001). The auditory organization of speech and other sources in listeners and computational models. *Speech Communication, 35*, 141–177. http://dx.doi.org/10.1016/S0167-6393(00)00078-9

Hermann, T., Hunt, A., & Neuhoff, J. (2011). The sonification handbook. Berlin: Logos Publishing House.

SAE International. (2014). *Taxonomy and definitions for terms related to onroad*

motor vehicle automated driving systems (Standard No. J3016). Available at http://standards.sae.org/j3016_201401.

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review*, 1–19.

National Highway Traffic Safety Administration. (2008). *National motor vehicle crash causation survey: Report to congress*. National Highway Traffic Safety Administration Technical Report DOT HS, 811, 059.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2012). *Engineering psychology and human performance (Fourth edition)*. Psychology Press.

2 STATE OF THE ART

2.1 Sound design process for automotive industry

Corresponding article: Bazilinskyy, P., Cieler, S., & De Winter, J. C. F. (2018). Sound design process for automotive industry. Under review.

2.1.1 Abstract

In the automotive industry sounds often play a safety-critical role. The automotive industry is recognized as a challenging arena for sound design, as presented information not only needs to comply with safety regulations but also be pleasant to drive and match subjective expectations. By means of a structured interview with 10 employees of the company Continental, we collected requirements for the sound design process in an automotive industry setting. This study presents a new sound design process, consisting of 3 stages: description, design/creation, and verification. An evaluation of the process was performed, when 2 sound designers in the company design 3 sound assets with and without a prototype of a web application employing the new process. The created sound design process supports the more efficient creation of auditory assets than the unstructured process that was previously employed in the company.

2.1.2 Introduction

2.1.2.1 The emergence of sound design

It is generally recognized that 1933 was the year sound design emerged as a discipline when Murray Spivack created the sounds for the movie King Kong. The voice of Kong, the sounds of the jungle, and the voices of creatures in the jungle were created for the movie. Although the use of sound in movies became popular after its introduction in 1927, for many years through the 1930s, the dominant figure in the world of cinema of the early 20th century, Charlie Chaplin refused to add sounds to his movies (TIME, 1931). Ben Burtt, the creator of the soundtrack for Star Wars, made a decision to add sound to the scenes in space, even though sound does not propagate in a vacuum. He designed sounds to increase the entertainment value of the movie, contrary to the movie 2001: A Space Odyssey, where attention was given to accuracy, and no sound could be heard in scenes in space. Lucas said “Let’s go for what is emotionally right” for the soundtrack of Star Wars (Gould, 2012). Since the release of King Kong, we have seen a number of technological developments in sound design such as audio digitization, samplers, synthesizers, and digital signal processors (DSP). All these inventions allow manipulating sounds beyond recognition, but none of them has generated a revolution in sound design comparable to Spivack’s work.

Nowadays, sound design is employed in numerous domains, such as sport sciences (Schaffert, Mattes, & Effenberg, 2009), the video game industry (Collins, 2008; Grimshaw, Klinger, & Snavely, 2011), design of contact sounds

(Cook, 2002; Van Den Doel, Kry, & Pai, 2001), emotionally enriched product design, and feedback in hospital environments and aircraft cockpits (Patterson & Mayfield, 1990; Stanton & Edworthy, 1999). Sound design does not have well-established guidelines as visual design, where required products can be described by 2D illustrations (Kress & Van Leeuwen, 1996; Mullet & Sano, 1996; Watzman, 2002) or industrial design, where objects to be designed can be outlined by 3D models and(or) multiple 2D illustrations (Krishnan & Ulrich, 2001; Roozenburg & Eekels, 1995; Sokovic & Kopac, 2006; Urban & Hauser, 1980). Frauenberger and Stockman analysed 23 projects involving different aspects of the sound design process that were presented during the 13th International Conference on Auditory Display (2007) in Montreal, Canada (Frauenberger & Stockman, 2009). Only 2 out of 23 projects followed a well-defined sound design process, and 14 papers were driven by the needs of real-world applications.

2.1.2.2 Sounds and sound design in the automotive industry

In the automotive industry sounds often play a safety-critical role. In a visually complex task such as car driving, auditory information is particularly beneficial as a warning signal, as hearing is omnidirectional (Bjork, 1995; Haas & Edworthy, 2006; Salvendy, 1997). For example, it has been found that auditory warnings are preferred over visual-only warnings when the driver has to resume manual control of an automated car (Bazilinskyy & De Winter, 2015). In addition, it was demonstrated that directed sound, e.g. towards an obstacle, can reduce reaction times and improve the decision behaviour of the driver (Liu & Jhuang, 2012; Pfromm, Cieler & Bruder, 2015). A structured sound design is important in the creation of products that deal with high levels of emergency and urgency (Stanton & Edworthy, 1999).

In the automotive industry different types of auditory feedback are employed. Firstly, sonification, or the use of sound instead of visual data or completing visual data (Kramer, 1994). Secondly, earcons (i.e., sound aimed to represent a specific event or which conveys certain information) are often found in in-vehicle interfaces. Thirdly, auditory icons (i.e., sounds that are based on real events and provide a metaphoric or iconic structure for the mapping with information they provide) are useful (Demarey & Plénacoste, 2001).

The automotive industry is recognized as a challenging arena for sound design, as presented information not only needs to comply with safety regulations but also be pleasant to drive and match subjective expectations (Genuit, 1997; Sottek, Krebber, & Stanley, 2005). Nowadays, silence, or the absence of unwanted sounds, is valued heavily, especially in luxury cars, and simple manipulation of the loudness and intensity of sounds in cars may not be enough to offer driver acceptance (Sottek et al., 2005). The sound of the engine is not the only component for the evaluation of the sound environment of a car. The way the car reacts to events acoustically is important too. Moreover, sounds that used to be unnoticeable, such as the noise from electric motors and squeaks of mechanical components, can now be heard in many models of cars, especially electric cars.

In terms of the use of auditory feedback from in-vehicle interfaces, modern cars are not fundamentally different from models released decades ago. With the introduction of fully automated (i.e., ‘driverless’) vehicles on the roads, in-vehicle sound may take a role of infotainment (Bazilinskyy & De Winter, 2015).

Such a radical change may require a novel sound design process that is tailored towards automated driving.

Often, manufacturers and suppliers in the automotive industry employ no standardized and documented sound design process. At Continental Automotive, for example, there had been a previous attempt to implement a paper-based sheet for describing sound assets. The sheet was in French, and it had been used only locally in the French office of the company, and it never became popular inside of the company. At this company, the process of designing auditory artefacts, from the moment of request to the moment of release of the final version followed the following steps: 1) verbal or written description/definition of requirements; 2) iterative process of creation of required artefact with updated descriptions given mainly verbally; 3) validation and testing. We hypothesized that this process could be optimised by employing a classification of auditory artefacts that facilitates the descriptions of assets.

2.1.2.3 Classification of auditory artefacts

Sounds can be grouped into templates for further use (Misra, Cook, & Wang, 2006). According to Bisping (1997), sounds can be classified in the power/pleasantness space. For example, sounds of in-vehicle interfaces can have different levels of pleasantness (quiet, annoying, desirable, booming, rough, noisy, and friendly) and power (racy, fresh, dynamic, fast, and exciting) (Bisping, 1997; Västfjäll, 2003). Bisping mentioned that the sounds of luxury cars are mainly associated with being powerful/pleasant, while sounds from sports cars are often linked to the powerful/unpleasant quadrant. The powerless/pleasant quadrant contained the interior sounds from middle-sized cars, and trucks and small cars were in the powerless/unpleasant quadrant. Sounds could also be classified based on their purpose. For example, sounds could be made for entertainment, confirmation/acknowledgement, notification, notification of error, or warning.

2.1.2.4 Aim of the paper

This study aimed to develop a sound design process to guide the workflow of designing sounds in the automotive industry. No software to assist with the process of designing sounds existed at the start of the project. To define steps in the process, requirements were gathered from the employees of Continental during verbal interviews. A prototype of a software product and database to assist with the workflow of the sound design process were developed and deployed on the company's server. The database was populated with tagged and classified sound samples to be used for describing required auditory assets. An evaluation study, where participants were asked to design auditory assets with and without the developed software prototype was undertaken. The results of the study were used to claim the validly of the developed solution.

2.1.3 Interviews to assess the current situation – new approach

Ten employees of Continental participated in a structured interview to gather the requirements for the definition and implementation of the sound design process. Six participants were working in an office in Germany, two in France, and two in China. They were recruited by posting a message seeking for participants in the

internal network of the company. The interviewees were based in the office of the company in Germany, France, and China. The interviewer was the first author of the present article. The interviews were conducted between May 10, 2016 and May 19, 2016. The interviews were later transcribed and the transcripts may be found in the supplementary material. Together, the transcribed interviews were 34,125 words including answers and replies of the interviewer. The shortest interview had 1,810 words in its transcription and it lasted for 14 min 43 sec, and the longest interview had 5,835 words in its transcription and lasted for 47 min 12 sec. The average length of the interviews was 3,413 words. The average duration of the interviews was 30 min 56 sec. All participants were informed that their responses would be treated anonymously and the results will be publishable. The questions that were asked in each of the interviews (see Table 1) focused on receiving general information about the interviewee, their level of involvement in sound design at Continental, and their views on how the sound design process should be structured.

Not all participants were able to give answers to all questions. For example, if the participant said in response to Q11 that they never describe auditory artefacts by means of software, Q12 was skipped. The average age of the participants was 34.3 years old (SD = 11.1). There were nine males and one female. The background of the interviewees was diverse (UI and UX designer, two project managers, audio and speech quality analyst, two software developers, electronic engineer, two sound designers, and ergonomist), and the amount of experience of working with auditory assets varied heavily, from less than a year to 18 years of experience. All the interviewed persons had to, in some way, work with earcons, with seven interviewees stating that they used spearcons (i.e., speech-based earcons) in their work as well. Three persons stated that they had worked with more than 100 auditory artefacts prior to the interview, and four interviewees reported having worked with tens of auditory artefacts in their careers. Eight participants reported being involved in projects dealing with automated driving.

Table 1. Questions asked in the interviews

Q1	**What is your age?**
Q2	Can you tell me about your activities at Continental (your job description)?
Q3	Why do you (need to) work with auditory artefacts?
Q4	What type of auditory artefacts do you have to work with?
Q5	How many artefacts did you need to design so far?
Q6	For which types of scenarios do you give preference to the auditory modality over the visual and tactile modalities?
Q7	Are you involved in automated driving in any way?
Q8	How long have you been involved in the process of creating auditory artefacts?
Q9	What software do you use to design auditory artefacts?
Q10	What do you think is the best software for the design of auditory artefacts?
Q11	If you need to describe auditory artefacts, what software do you use for this?
Q12	What do you think is the best software for describing auditory artefacts?
Q13	What qualitative parameters of the auditory modality do you use to describe auditory artefacts?

Q14	Could you give an example of a verbal/textual description of an auditory artefact you used in one of your projects?
Q15	What quantitative parameters of the auditory modality do you use to describe auditory artefacts? (for example, frequency, duration, etc.).
Q16	We will build a database of auditory samples to be used in the auditory artefact creation process. All samples will be associated with tags and categories to support search. Could you name examples of the essential tags and categories to be used in such a database?
Q17	How many iterations do you (your team) normally go through before the final version of an auditory artefact is produced?
Q18	At which stage do things normally go 'wrong' in the design process of auditory artefacts? What would you name as a main drawback of the current design process?
Q19	Do you use any "agile principles" in the process of auditory artefacts creation?
Q20	Can you give an example of a situation where the auditory artefact design process 'went wrong'? When you did not receive an auditory artefact, you asked for.
Q21	Do you interact with GUI developers and designers during the process of creating auditory artefacts? How are auditory and visual outputs harmonized?
Q22	Are there any standardized "tests" for auditory artefacts? If yes, what tools do you use for such tests?
Q23	Which types of people participate in these tests?
Q24	Would you wish to see a standardized approach to producing auditory artefacts at Continental?
Q25	How would you improve the sound design process at Continental?

We asked the participants to talk about situations where they would prefer to use the auditory modality over visual and haptic modalities. Nine participants gave a response. Three interviewees replied that the auditory modality is beneficial during take-over requests in automated driving. Three persons indicated that it is beneficial when the information presented is not in the visual scope. One interviewee also stated that auditory feedback should be used when it is not needed to attract visual attention of the driver to a certain point in space. Furthermore, one participant expressed his opinion that speech output may be beneficial: "…if you really need to transport content and specific information"; he also stated that auditory output might be preferred in the context of transition of control where it is assumed that the driver will not be able to observe information in the cluster instrument due to visual distraction of doing a non-driving related task. Finally, two participants also stated that the auditory modality is beneficial for issuing confirmations.

The majority of participants describe their needed sound artefacts verbally or by means of simple text descriptions created with word processing tools, such as Microsoft Word or by plain text emails. Two participants reported being closely connected to the actual process of designing sounds as sound designers; they used Ableton Live, Logic Pro, Cubase, or Audacity software in their work. No clear opinion about the best software for the test of designing auditory artefacts was given.

In Q13, the interviewees were polled to list the qualitative parameters of the auditory modality that are used in the company for describing auditory artefacts. Six interviewees mentioned that the "mood" (happy, sad, etc.) of the artefact was important to mention. Four participants reported the "urgency" (not urgent, urgent) as an important parameter. Two interviewees said that the "value" (sounds cheap, sounds expensive or luxury) of the artefact has an

important role in the process. In Q15, the participants were asked to report the quantitative parameters of the auditory modality they used for describing auditory assets. The most commonly used parameters were frequency, duration/speed, and pitch. One of the sound designers reported that the duration of the required auditory artefacts was dependent on the animations that were used together with the artefacts. None of the interviewees was willing to provide examples of the descriptions of auditory assets from the projects in the company, due to confidentiality restrictions.

For Q16, tags and categories to be used in the software tool and design were mentioned. Figure 1 shows the word cloud generated based on the responses (at http://www.wordclouds.com). The most commonly mentioned tags are (mentioned more than once): spearcon, earcon, warning, metal, percussive, attention, emergency, indicator, and awareness.

Figure 1. Word cloud of suggested during interviews tags for the database.

The participants reported a diverse range of the number of iterations that happen when an auditory artefact is created in the company: from two to three iterations to having hundreds of iterations in one project. Seven participants replied affirmative to Q19, where they were asked if the principles of agile development (an iterative approach to software development (Collier, 2011)) are used for the creation of auditory artefacts in the company. Five of them reported the use of agile principles for the creation of auditory artefacts; a number of people said that they often face problems in the design process of auditory artefacts (Q18) during the description stage. All ten participants wanted to have a more structured sound design process in the company. The interviewees wished to have a simplified process featuring possibilities to verify the designed auditory artefacts in the company.

2.1.4 The newly developed sound design process

Figure 2 shows a new sound design process for developing auditory artefacts for the automotive industry. It is based on the workflow of creating auditory assets that was employed at Continental prior to the start of the project. During

this process, the auditory asset is created by following three main stages: 1) description; 2) design/creation; 3) verification. Hence, compared to the previously employed model, the new process digitalises the flow of information, adds descriptors to descriptions, and adds the verification stage.

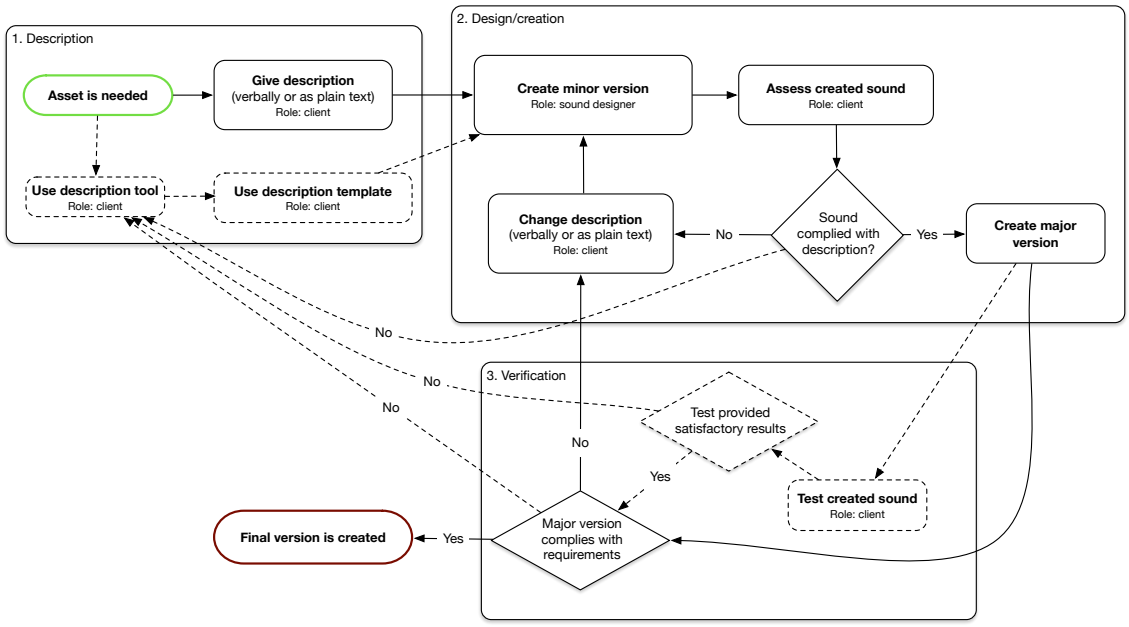


Figure 2. The newly developed sound design process. Rectangles represent actions; rhombuses represent decision to be taken. Shapes with solid borders and solid arrows represent steps and actions, respectively, that were already implemented in the existing process. Steps in boxes with dotted borders and dotted arrows indicate the newly developed steps and actions, respectively. The process assumes two roles of users: client (the person that needs the sound to be designed and created) and sound designer (the person that works with given descriptions to design and create the required sound).

The process starts in the description stage (Stage 1), when a person that requires the asset to be designed and created gives the initial description. Then, a number of iterations (one or more) are performed in the design/creation stage (Stage 2). During this stage, the designer uses descriptions given during Stage 1 (initial description) or during Stage 2 (updated descriptions) to create a minor version of the sound. The client assesses the created sounds and decides if the given version can be recognised as a major version. If it can be, the process enters Stage 3, where verification of the created sounds is undertaken. If the given sound satisfies all of the requirements, it is recognised as being final, the process comes to an end, and the final version is created. If the verification cannot be passed, the process goes back to Stage 2, and further modifications are undertaken in an iterative manner until a new major version is created that enters the verification stage (Stage 3).

The newly developed sound design process assumes the use of specialised software aided by a database with sounds that may be used as examples to aid descriptions. The tool and database help the process during Stages 1 and 3, and such steps are shown in green boxes. Such tooling removes the necessity to use plain text descriptions and verbal communication, which may not be precise enough.

2.1.5 Software tool and database for the description of sounds

The interviews showed that there was a need to design and implement the new sound design process in the company. To that end, a prototype of a software tool and accompanying database were created.

2.1.5.1 Method

The tool was developed with Flask library of Python 2.7 in the backend and Jinja2 in the frontend. The tool is available for testing at https://wordsforsound.herokuapp.com. The version is populated with auditory samples with CC licenses from the Internet, that is, it does not contain any confidential material belonging to Continental. The source code of the tool is available in the supplementary material. The prototype was developed between June 13, 2016 and July 27, 2016. Additional bug fixing and implementation of new features and improvements to existing functionality took place after July 27, 2016.

2.1.5.2 Design of tool and database

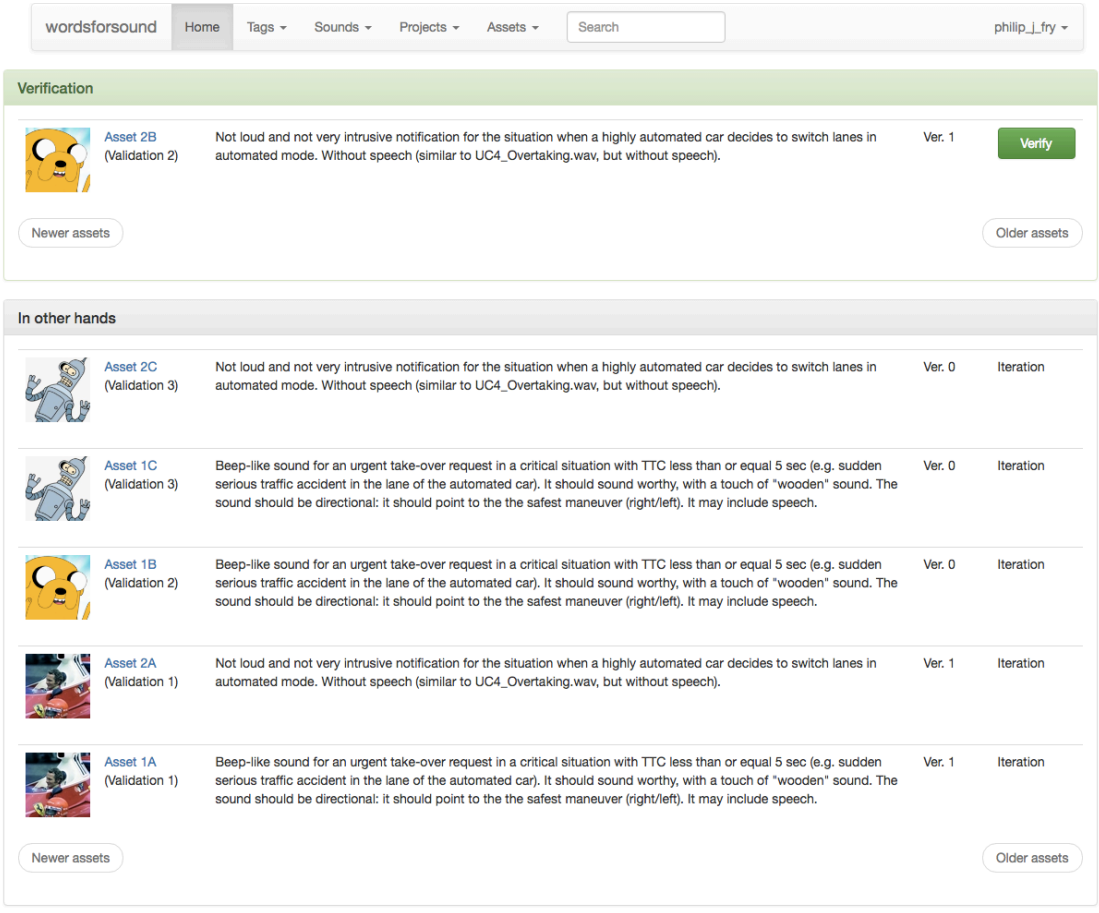


Figure 3. The homepage of the tool prototype.

The final version of the tool is based on the sound design framework proposed in Section 3. The tool features users of two types – 'clients' and 'suppliers' – which work on auditory assets by means of the tool. Clients are users that need auditory assets to be made. Suppliers are users that need to make (design/create) such requested assets. Multiple clients and multiple suppliers

can be involved in the creation of one asset. Assets need to belong to projects. Each project can have one or multiple assets. Figure 3 illustrates the home page view of the developed prototype. In this figure, the logged in user *philip_j_fry* is involved in six assets. Since he is a client, he needs to verify one of the assets ('Asset 2B' from the project 'Validation 3'), which is in iteration 1 (stage 2 in the process described in Fig. 2), i.e. the sound designer working on the asset has submitted one version of the asset as iteration 1. Five other assets are 'in other hands', i.e., *philip_j_fry* needs to wait until he can do any action for those assets as other users involved in the creation of those assets need to finish their tasks.

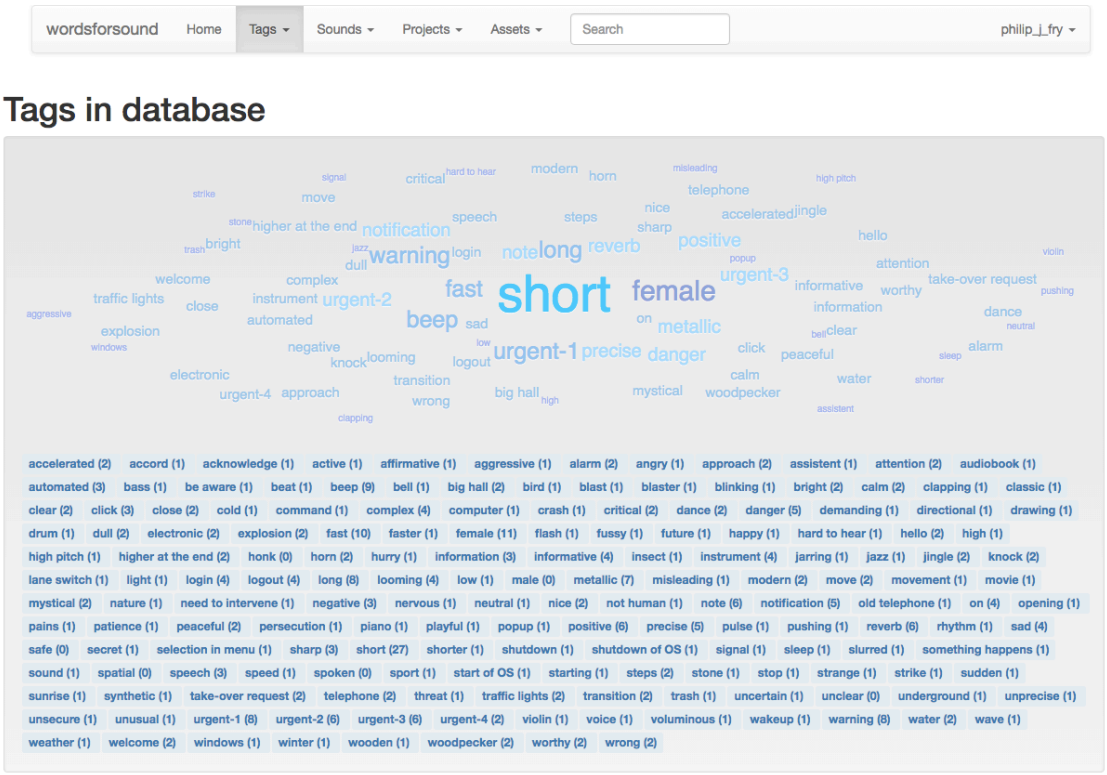


Figure 4. View of tags with word cloud in the tool prototype.

The tool is supported by a database. Besides being the platform supporting the process of creating auditory assets, it also serves as a company platform for storing new and finding already stored tags and auditory samples. It is a 'living' database, which is enriched by the employees of Continental. Figure 4 shows the view of tags in the database. The word cloud displays all tags in the system, where the size of the tag outlines how frequently it is used in the company. Figure 5 shows the view of the sound examples in the system. All sounds on the page are accompanied with embedded clickable previews of the sounds. All elements on both of these views are clickable. Figure 6 shows the view of an individual asset. On this page, users that have access to the asset can see all descriptions, iterations, and verifications for the asset.

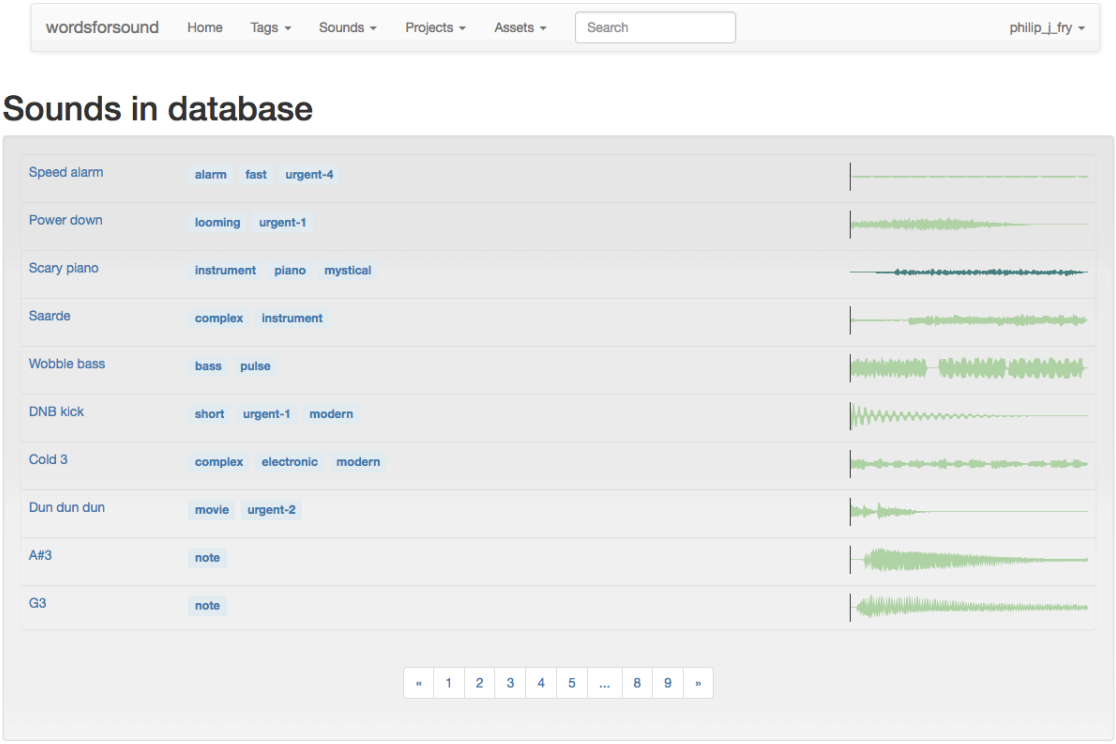


Figure 5. View of sound examples with clickable items in the tool prototype.

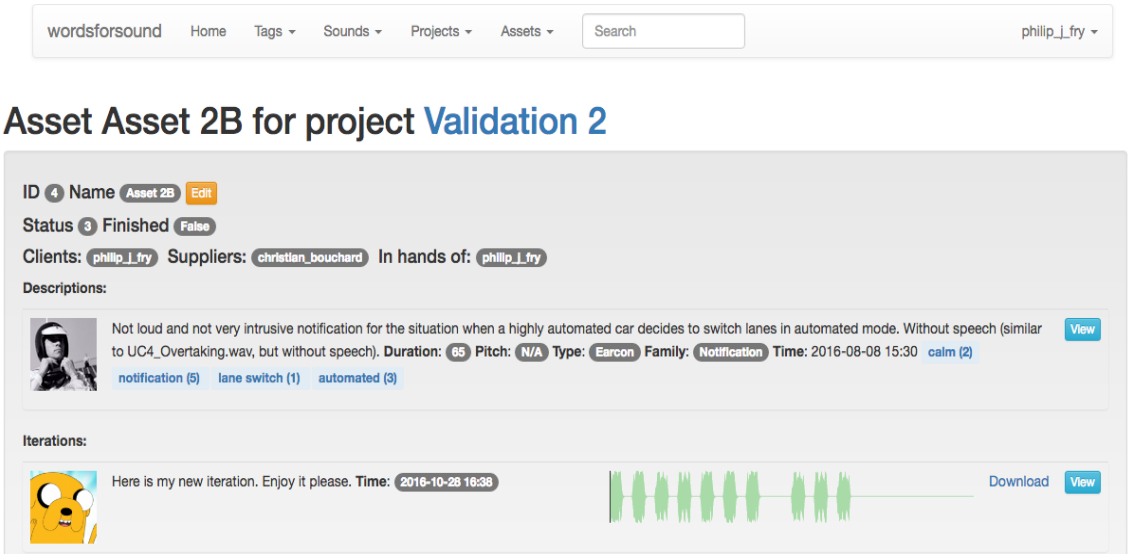


Figure 6. View of an asset with a history of changes.

2.1.5.3 Population of the database

The database was initially populated with 52 sounds from the previous and ongoing projects in the company. To enrich the database with tagged sounds, the employees at the company were asked to tag the sounds with as many tags for each sound as they could think of during a week. This allowed generating a first pool of tags to be used in the tool. Five people responded to the call and provided tags. In total, 310 tags were collected and added to the database, see Figure 4. The tags were diverse, indicating different needs of clients and designers of sounds, who are the stakeholders of the system. Words “short”, “female”, “fast”, “warning”, “reverb” were the most commonly used tags. Additionally, 32 sounds under CC license from the Internet were added to the

initial pool. This pool of sounds featured commonly used auditory assets, such as beeps, alarms, and alert sounds, and was tagged by the first author of the article.

2.1.6 Evaluation study

2.1.6.1 Method

In this experiment, two sound designers were asked to design three auditory assets without the tool prototype and database (Phase 1, baseline case with no structured sound design process) and with the tool prototype and database (Phase 2, a case with structured sound design process). One 'client' (i.e., second author of this work) represented an owner of the hypothetical project that needed the sounds to be designed. All assets described scenarios possible with automated driving to cater for the increasing number of projects within this topic.

Before the start of the first phase, the participants were asked to read instructions and complete a short introductory questionnaire. They were asked the following questions:

1. What is your gender?
2. What is your age?
3. How many years of experience in the area of design of auditory assets do you have?
4. Is design of auditory assets in the context of automated driving any different from regular manual driving?

The sound designers were then asked to design three assets that were described as follows:

- *Asset 1a:* Urgent take-over request for a situation where a car needs to give control back to the driver by means of a take-over request. The request is issued when the car is performing a lane-change, and it detects a fast-moving approaching vehicle from behind. The driver has less than 3 seconds to take back control. Furthermore, the sound should be directional and looming: it should give information on the location and speed of the approaching car. Input: speed of the automated car, speed of the car behind, TTC.
- *Asset 1b:* Take-over request for an urban environment. The car is driving in a city, and it faces construction works ahead. The car decides to give control back to the driver because it cannot turn around and it has no information about adjacent streets that would allow the car to go around the construction area. It should sound modern, precise, and not boring. High pitch. Speech-based.
- *Asset 1c:* Confirmation signal that the automation mode was successfully enabled. It should not be loud or annoying.

Assets 1b and 1c had no input parameters. The description of Asset 1a contained text "Urgent take-over request…", giving an explicit instruction that it had to be a sound baring a sense of high urgency. Descriptions of the other two assets were made to represent auditory warnings of medium (Asset 1b) and low

(Asset 1c) urgency. The participants were asked to work on the design of the assets as if it was a real project and deliver their version to the client as soon as possible. The client would then give feedback on the iterations and inform the designers if any further work was required.

When the participant finalised all three assets, he was asked to proceed to the second phase of the experiment. In this phase, they had to design three new assets. The descriptions of the assets in the second phase were comparable to descriptions of the corresponding assets in the first phase (Asset 2a was based on Asset 1a, Asset 2b on Asset 1b, and Asset 2c on Asset 1c). This time they were asked to organise the design process with the help of the tool prototype and database. The assets were described in the tool as follows:

- *Asset 2A*: Beep-like sound for an urgent take-over request in a critical situation with TTC less than or equal 5 sec (e.g., a sudden serious traffic accident in the lane of the automated car). It should sound worthy, with a touch of 'wooden' sound. The sound should be directional: it should point to the safest manoeuvre (right/left). It may include speech. **Input**: speed of the automated car, TTC, safest manoeuvre trajectory. **Tags**: beep, urgent, take-over request, critical, automated car, worthy, wooden, directional, safe.
- *Asset 2B*: Non-urgent take-over request (TTC less than 10 sec) with information on an object in the blind-spot in the left lane, behind the automated car (driving on the middle lane). It should sound non-intrusive, modern and electric. Could involve speech by a female actor with US English accent. **Tags**: non-urgent, take-over request, blind-spot, automated car, non-intrusive, worthy, modern, electric, speech, US English accent.
- *Asset 2C*: Not loud and not very intrusive notification for the situation when a highly automated car decides to switch lanes in automated mode. Without speech (similar to UC4_Overtaking.wav, but without speech). **Tags**: not loud, notification, switch lanes, highly automated car.

All of the assets in Phase 2 were accompanied by auditory examples from previous projects in the company. The auditory examples were added to the descriptions in the software tool from the database.

A new iteration was considered finished if the process "Create minor version" or "Create major version" was reached (Fig. 2). In the instructions for each Phase, the participants were asked to monitor the amount of time it took them to complete an iteration. Furthermore, after finishing the iterations after Phase 1 and Phase 2, they were asked to report the amount it took to produce the sounds.

2.1.6.2 Results

The participants were able to successfully deliver all three assets with the help of the tool and without the tool. Table 2 summarises self-reported number of iterations that were used and time spent to deliver the final versions of the assets by both participants. Two criteria were compared to estimate the effectiveness of the employment of the tool for the sound creation process: number of

iterations for producing the final version and the amount of time required to produce the final version.

Table 2. Numbers of iterations that were required to deliver the final version of the assets.

	Participant 1 - iterations	**Participant 1 - duration of work**	**Participant 2 - iterations**	**Participant 2 - duration of work**
Asset 1A	5	3 h 15 min	5	2 h 6 min
Asset 1B	2	2 h 30 min	2	50 min
Asset 1C	3	2 h 30 min	2	50 min
Asset 2A	3	45 min	1	25 min
Asset 2B	1	15 min	1	25 min
Asset 2C	1	15 min	1	25 min

As can be seen in Table 2, the time required for the creation an auditory asset was lower when the tool was used compared to when it was not used. ($M_{no\ tool} = 120$; $M_{with\ tool} = 25$). Furthermore, the average number of iterations required to produce the final version, as indicated in Stage 3 of the described process (Fig. 1), of the auditory asset was reduced by 1.84 iterations ($M_{no\ tool} = 3.17$; $M_{with\ tool} = 1.33$). The participants also commented that the process of creating assets with the software prototype was easier and more straightforward, compared to the process of designing sounds implemented before the project.

2.1.7 Discussion

In this study, we designed and implemented a new sound design process in an automotive industry setting. Multiple factors make the creation of auditory feedback for in-vehicle interfaces challenging: strict requirements for reliability, end users with requirements that are difficult to outline, etc. A well-defined process is needed to support the design of auditory assets in such an environment. Unlike the movie industry, sound design is not well-defined in the automotive industry. This results in a paradoxical situation, where auditory assets that need to be created have to adhere to well-defined requirements, but people that need to make such assets have no access to a structured process aimed to allow the creation of such assets. In this project, we offered a solution to this problematic situation.

By means of a structured interview with 10 employees of the company Continental, we collected requirements for the process of designing and creating auditory artefacts in an automotive setting. The results of the interviews showed that the needs and expectations of the employees of the company varied greatly. It was challenging to develop the sound design process to be used by both employees that request auditory feedback to be designed as well as employees that are asked to design such feedback. Based on the interviews, it was clear that the interviewed employees were interested in simplifying the process of designing auditory artefacts in the company. An easy to use software solution was needed.

Based on the results of the interviews and conducted literature survey, we created a sound design process tailored to research and development activities in the automotive industry. It consists of three stages: description, design/creation, and verification (Fig. 2). The designed process assumed two

roles of users – client and sound designer – which corresponded to the unstructured sound creation process employed by the company prior to the project. The verification stage was added to the process because multiple employees of the company reported problems arising from not having a structured process for verifying if created auditory assets correspond well to their requirements. The creation process is iterative, and both clients and sound designers can enter all three stages multiple times while working on one auditory asset.

To validate the sound design process, we developed a prototype of a web application for the use in the company. The prototype was supported by a database populated with sound examples and tags. We launched an evaluation study to compare the process of creating auditory assets with and without the tool. Two sound designers employed in the company were asked to design six auditory assets of varying complexity and associated urgency, three assets by following the unstructured process and three assets by following the newly introduced sound design process. For the creation of assets with the new sound design process, the participants were given sound examples and tags in addition to textual descriptions. Both participants were able to create all three assets with the new sound design process faster and by going through fewer iterations compared to creating similar auditory assets without the tool. They used the iterative design approach and utilised the ability to supplement textual descriptions with auditory examples and tags from the database of auditory examples and assets from previous projects. There may have been a learning effect in our evaluation study, since the assets were created by means of the newly introduced sound design process after working on the assets by employing the unstructured process. The software tool had to be learned as well, which may have affected the results too. We can conclude that the created sound design process supports the more efficient creation of auditory assets than the unstructured process that was previously employed in the company. The software solution will be improved in the future to optimise the sound design process further. The company has added the presented in this article sound design process to their workflow. A randomized validation study with a larger number of participants should be conducted to validate the learning effects.

2.1.8 Supplementary material

Anonymised transcripts of interviews and instructions with all questionnaires for the evaluation experiment may be found at https://www.dropbox.com/sh/i6klu4hn26y4mom/AACASF-_f-tCY85DMcEjNexia?dl=0

Source code of the developed software prototype is available at https://github.com/bazilinskyy/wordsforsound

2.1.9 References

Bazilinskyy, P., & De Winter, J. (2015). Auditory interfaces in automated driving: an international survey. *PeerJ Computer Science*, *1*, e13. https://doi.org/10.7717/peerj-cs.13

Bazilinskyy, P., Petermeijer, S. M., Petrovych, V., Dodou, D., & De Winter, J. C. F. (2018). Take-over requests in highly automated driving: A crowdsourcing

survey on auditory, vibrotactile, and visual displays. *Transportation Research Part F: Traffic Psychology and Behaviour*, *56*, 82–98.

Bisping, R. (1997). Car Interior Sound Quality: Experimental Analysis by Synthesis. *Acta Acustica United with Acustica*, *83*(January 1996), 813–818.

Bjork, E. A. (1995). Psychophysiological responses to some natural sounds. *Acta Acustica*, *3*(1), 83–88.

Collier, K. (2011). Introducing Agile Analytics: A Value- Driven Approach to Business Intelligence and Data Warehousing. Proceedings of *Agile Analytics: A Value-Driven Approach to Business Intelligence and Data Warehousing*.

Collins, K. (2008). Game Sound: An Introduction to the History, Theory, and Practice of Video Game Music and Sound Design. *Pelitutkimuksen Vuosikirja 2010*, *22*, 200. https://doi.org/10.1093/jdh/epp021

Cook, P. R. (2002). Modeling Bill's gait: Analysis and parametric synthesis of walking sounds. Proceedings of Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio.

Demarey, C., & Plénacoste, P. (2001). User, sound context and use context: What are their roles in 3D sound metaphors design? In *International Community for Auditory Display* (pp. 136–140).

Film Sound. (2008). The sounds in Star Wars. Available at http://filmsound.org/starwars/dramatic.htm

Frauenberger, C., & Stockman, T. (2009). Auditory display design-An investigation of a design pattern approach. *International Journal of Human Computer Studies*, *67*(11), 907–922. https://doi.org/10.1016/j.ijhcs.2009.05.008

Genuit, K. (1997). Background and practical examples of sound design. *Acta Acustica United with Acustica*, *83*(5), 805–812.

Gould, R. (2012). Concepts and Approaches in Sound Design. Available at http://audioblackholes.tumblr.com/post/28441657592/concepts-and-approaches-in-sound-design-ep1

Grimshaw, M., Klinger, K., & Snavely, J. (2011). Game Sound Technology and Player Interaction: Concepts and Developments. Russell The Journal Of The Bertrand Russell Archives. https://doi.org/10.4018/978-1-61692-828-5.ch017

Haas, E., & Edworthy, J. (2006). An introduction to auditory warnings and alarms. *Handbook of Warnings*, 189–198.

Kramer, G. (1994). Some organizing principles for representing data with sound. *Auditory Display-Sonification, Audification, and Auditory Interfaces*, 185–221.

Kress, G. R., & Van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. Psychology Press.

Krishnan, V., & Ulrich, K. T. (2001). Product Development Decisions: A Review of the Literature. *Management Science*, *47*(1), 1–21. https://doi.org/10.1287/mnsc.47.1.1.10668

Liu, Y.-C., & Jhuang, J.-W. (2012). Effects of in-vehicle warning information displays with or without spatial compatibility on driving behaviors and response performance. *Applied Ergonomics, 43*(4), 679–686.

Misra, A., Cook, P. R., & Wang, G. (2006). A new paradigm for sound design. Proceedings of *the International Conference on Digital Audio Effects (DAFx-06)*, 319–324.

Mullet, K., & Sano, D. (1996). Designing Visual Interfaces. *ACM SIGCHI Bulletin (Vol. 28)*. https://doi.org/10.1145/226650.570118

Patterson, R. D., & Mayfield, T. F. (1990). Auditory warning sounds in the work environment [and Discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 327*(1241), 485–492.

Pfromm, M., Cieler, S. & Bruder, R. (2015). Auslegung und Evaluation einer Mensch-Maschine-Schnittstelle für ein umfassendes Fahrerassistenzkonzept. *Zeitschrift für Arbeitswissenschaft, 69(2)*, 71–80.

Roozenburg, N. F. M., & Eekels, J. (1995). *Product design: fundamentals and methods* (Vol. 2). Wiley Chichester.

Salvendy, G. (1997). Handbook of human factors and ergonomics. *Handbook of Human Factors and Ergonomics*, *9*(3), 2138. https://doi.org/10.1002/9781444395150.ch13

Schaffert, N., Mattes, K., & Effenberg, A. O. (2009). A sound design for the purposes of movement optimisation in elite sport (using the example of rowing).

Sokovic, M., & Kopac, J. (2006). RE (reverse engineering) as necessary phase by rapid product development. *Journal of Materials Processing Technology*, *175*(1–3), 398–403. https://doi.org/10.1016/j.jmatprotec.2005.04.047

Sottek, R., Krebber, W., & Stanley, G. R. (2005). Tools and methods for product sound design of vehicles.

Stanton, N. A., & Edworthy, J. (1999). *Human factors in auditory warnings*. Ashgate.

TIME. (1931, February 9). Charles Spencer Chaplin. *TIME*.

Urban, G. L., & Hauser, J. R. (1980). *Design and marketing of new products*. Prentice hall.

Van Den Doel, K., Kry, P. G., & Pai, D. K. (2001). FoleyAutomatic: physically-based sound effects for interactive simulation and animation. Proceedings of *the 28th annual conference on Computer graphics and interactive techniques*, 537–544.

Västfjäll, D. (2003). Emotional Reactions to Product Sounds. *Human Engineering*.

Watzman, S. (2002). Visual design principles for usable interfaces. The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, 263–285.

2.2 Auditory interfaces in automated driving: an international survey

Corresponding article: Bazilinskyy, P., & De Winter, J. (2015). Auditory interfaces in automated driving: an international survey. PeerJ Computer Science, 1, e13.

2.2.1 Abstract

This study investigated peoples' opinion on auditory interfaces in contemporary cars and their willingness to be exposed to auditory feedback in automated driving. We used an Internet-based survey to collect 1,205 responses from 91 countries. The respondents stated their attitudes towards two existing auditory driver assistance systems, a parking assistant (PA) and a forward collision

warning system (FCWS), as well as towards a futuristic augmented sound system (FS) proposed for fully automated driving. The respondents were positive towards the PA and FCWS, and rated the willingness to have automated versions of these systems as 3.87 and 3.77, respectively (on a scale from 1 = disagree strongly to 5 = agree strongly). The respondents tolerated the FS (the mean willingness to use was 3.00 on the same scale). The results showed that among the available response options, the female voice was the most preferred feedback type for takeover requests in highly automated driving, regardless of whether the respondents' country was English speaking or not. The present results could be useful for designers of automated vehicles and other stakeholders.

2.2.2 Introduction

2.2.2.1 The development of automated driving systems

The development of automated driving technology is a key topic in modern transportation research. A transition to automated driving may have a large positive influence on society (European Commision, 2011). Each year more than 1,000,000 fatal accidents occur on roads worldwide, with the lower-income countries being overrepresented (Gururaj, 2008; World Health Organization, 2013). If automated driving systems are designed to be fully capable and reliable, a very large portion of—yet probably not all—road traffic accidents could be prevented (Goodall, 2014). Furthermore, traffic congestions, gas emissions, and fuel consumption may reduce considerably thanks to automated driving systems.

The control of vehicles can be represented as a spectrum consisting of five levels: (1) manual driving, (2) driver assistance, (3) partially automated driving, (4) highly automated driving, and (5) fully automated driving (Gasser & Westhoff, 2012). The introduction of driver assistance systems (i.e., level 2 automation) to the public took place in the 1990s with the release of Adaptive Cruise Control (ACC), a system that automates the longitudinal motion of the vehicle (Beiker, 2012). Advancements in cameras, radars, lasers, and artificial intelligence have led to the creation of systems that make partially automated driving possible. Partially automated driving systems not only control the longitudinal motion of a vehicle, but also its lateral motion. Examples of such systems are BMW's Traffic Jam Assistant (BMW, 2013), Volvo's ACC with steer assist (Volvo, 2013a), and Mercedes' Distronic Plus with Steering Assist (Daimler, 2013). In partially automated driving, drivers are usually required to keep their eyes focused on the road and intermittently touch the steering wheel.

Highly automated driving (HAD) is a next step. In HAD, the human can release the hands from the steering wheel and is no longer required to monitor the road permanently (e.g., Banks, Stanton, & Harvey, 2014). However, humans still have an important role in the control of highly automated vehicles (Alicandri & Moyer, 1992; Dingus, Hulse & Barfield, 1998; Levitan, Golembiewski & Bloomfield, 1998). In HAD, drivers can be asked to take over control of the vehicle when required, for example, when the vehicle automation cannot solve a task in a demanding traffic environment. The time between issuing a 'takeover request' and the required moment of transition of control from the vehicle to the human is a critical design parameter (Gold et al., 2013). If the driver spends too

much time on reclaiming the control of the vehicle, or if the driver does not comprehend the warning signal sent by the vehicle, an accident may result. Clearly, the design of appropriate feedback is essential for the successful introduction of HAD to the public roads. Indeed, inappropriate feedback is regarded as a primary cause of automation-induced accidents (Norman, 1990).

Fully automated driving (FAD) will be the next and final iteration in automated driving. People have been envisioning this step in the development of transportation for a long time. Almost half a millennium ago, Leonardo Da Vinci envisioned a pre-programmed clockwork cart (Weber, 2014). In 1939 during the New York World's Fair, General Motors presented their vision of the world 20 years into the future (1959–1960). In their Futurama exhibition, they introduced a concept of automated highways with trench-like lanes for separating traffic (Wetmore, 2003). In 1953, the futurist Isaac Asimov wrote a short story 'Sally' that pictured a situation where only cars that did not require a human driver were allowed on the roads.

FAD offers numerous potential benefits. It could reduce stress and allow the operator to engage in non-driving tasks such as working, using in-vehicle entertainment, or resting (e.g., Jamson et al., 2011; Llaneras, Salinger, & Green, 2013). Furthermore, FAD is a recommended solution for achieving an optimal traffic flow, for example by means of platooning on highways (Bergenhem et al., 2012; Varaiya, 1993). The Google Driverless Car is one of the existing prototypes of FAD (Markoff, 2010). However, this particular vehicle does not fully comply with the principles of FAD; in reality, the Google Driverless Car relies on accurate three-dimensional maps of the environment and currently cannot cope with all dynamic environments of high complexity. It requires considerable advances in sensing and artificial intelligence before FAD becomes practically feasible on all public roads. Continental, a leading German manufacturer specialising on components for automotive industry, predicts that FAD will be launched in the year 2025 (Continental, 2012), whereas some voices have argued that FAD will never happen (Gomes, 2014; Underwood, 2014; Yoshida, 2014).

Although automated driving systems are expected to improve safety, certain side effects may occur regarding the human factor (e.g., Bainbridge, 1983; Desmond, Hancock, & Monette, 1998; Merat et al., 2012; Brandenburg & Skottke, 2014). A degraded reaction time to critical events has been found among drivers exposed to ACC (Stanton, Young, & McCaulder, 1997; Stanton et al., 2001; Rudin-Brown & Parker, 2004; Larsson, Kircher, & Hultgren, 2014), and this issue is likely to be aggravated in higher levels of automated driving (De Winter et al., 2014; Strand et al., 2014). Furthermore, it is expected that people who will be driving highly and fully automated cars will suffer from a reduction of their manual control skills, similar to pilots in highly automated airplanes (Ebbatson, 2009; Scallen, Hancock & Duley, 1995). The development of effective feedback systems is considered important in supporting operator's sustained attention, also called vigilance (Heikoop et al., 2015).

2.2.2.2 Auditory displays

As mentioned above, unless the driving task is fully automated, an appropriate feedback system is required that warns and/or informs the human when automation mode changes are required. The present study investigated the potential of auditory feedback in automated driving. The auditory modality has

several important characteristics: (1) it is omnidirectional, that is, unlike visual cues, auditory cues can be received from any direction. This is especially important in automated driving, during which the driver may not be attending to the road and dashboard; (2) the auditory sense can receive information at almost all times; (3) sound is transient, that is, unlike visual information which can be continuously available, information passed in the form of sound is only available at that particular moment; (4) although auditory cues may be masked by other sounds, humans have the ability to selectively focus on one sound when multiple streams of sound are available, also known as the cocktail party effect (Bregman, 1990; Cooke & Ellis, 2001; Hermann, Hunt & Neuhoff, 2011; Wickens & Hollands, 2013).

An advantage of sound is that it is possible to use language, which may be more informative as compared to the information conveyed with haptic or visual interfaces. Because of the aforementioned qualities of sound, auditory displays are used in a variety of applications, especially in those cases where the user needs to be alerted or where additional visual load has to be avoided. For example, the majority of present route navigation devices use voice and sound messages to give directions to their users (Holland, Morse & Gedenryd, 2002), and flight crews use auditory signals to get informed about proximate aircraft or to obtain directional information (e.g., Begault, 1993; Bronkhorst, Veltman & Van Breda, 1996). An auditory interface in combination with tactile feedback was suggested in a driving simulator study (Ho, Reed & Spence, 2007) as an optimal warning system for collision avoidance. The auditory modality has potential not only as a warning method, but also for providing inputs to the machine (e.g., speech interfaces). Literature reviews (Stanton & Edworthy, 1999; Barón & Green, 2006) suggest that people drive 'better' (i.e., lower lane variation, steadier speed) when auditory interfaces are employed in a manually driven car.

Auditory feedback can be delivered as a pre-recorded voice or as an artificial sound warning/message. The term earcon refers to a brief auditory message (e.g., a tune or a sound of a bell) that represents a certain event or object. Earcons have been introduced to desktop computers to complement visual icons (Mynatt, 1990; Belz, Robinson & Casali, 1999; Hermann, Hunt & Neuhoff, 2011). Previous research has shown that a female voice is favoured over a male voice in route navigation devices (Large & Burnett, 2013). However, national or cultural differences seem to exist, where in some cases, the male voice is preferred over the female voice. In 2010, BMW supposedly had to recall its navigating system in Germany because male drivers disliked the idea of following orders communicated via a female voice (Takayama & Nass, 2008), and Apple recently added the option of a male voice to their voice control system Siri (Bosker, 2013). In a driving simulator study by Jonsson and Dahlbäck (2011), non-native speakers of English responded more accurately to route instructions provided by a female voice than to route instructions provided by a male voice.

2.2.2.3 Auditory systems in current vehicles: parking assistant and forward collision warning system

Modern vehicles often include systems that assist in driving and increase road safety. Such systems support drivers by providing auditory/visual/haptic warning messages and by taking over control of some of the driving tasks. In the present survey, we investigated the opinion of people on two existing auditory

systems: a parking assistant (PA) and a forward collision warning system (FCWS).

The first generations of PAs were so-called parking sensors, which produce warning sounds (beeps) when the car gets too close to a nearby object while parking, using ultrasonic or electromagnetic sensors (BMW, 2013; Toyota, 2014; Volkswagen, 2014). Some recent PA systems take over the positioning of the vehicle during parking, leaving the control of acceleration and deceleration to the driver (Volkswagen, 2014). Other PAs take over control of the parking process entirely, as can be seen in the Toyota Prius 2015 and BMW X5 (BMW, 2014; Toyota, 2014).

A FCWS is a system that provides a warning sound when a vehicle is rapidly approaching a vehicle in front. FCWSs have the potential to prevent a large portion of rear-end collisions (Jamson, Lai & Carsten, 2008; Kingsley, 2009; Kessler et al., 2012). If a potential accident is detected by the FCWS, the system either gives a warning to a driver (Honda, 2014) or engages in emergency braking and/or steers way from the object (Volvo, 2013b). Most FCWS detect vehicles with the help of computer vision (Srinivasa, 2002; Dagan et al., 2004), an approach that is used by companies like Honda and BMW (BMW, 2013; Volvo, 2013b; Honda, 2014) and/or radars (Volvo, 2013b; Ford, 2014; Honda, 2014; Mercedes-Benz, 2014). Both approaches have limitations, and the system may not issue warnings or stop the vehicle in bad weather or in other situations where the sensors are obscured by external factors. The introduction of vehicle-to-vehicle (V2V) communication may increase the efficiency and capabilities of collision warning systems (e.g., Miller & Huang, 2002). Eighty-eight percent of owners of Volvo cars surveyed by Braitman et al. (2010) reported always having the FCWS turned on.

It is expected that both PAs and FCWSs will remain in future partially and highly automated vehicles. However, these technologies will become obsolete with the introduction of FAD because both parking and collision avoidance will be handled without any input from the human driver.

2.2.2.4 ‘Augmented / spatial’ sound system for fully automated driving

Auditory warning signals will not be required in FAD, because in FAD the automation by definition takes care of all possible emergency conditions. This study proposes an experimental setup aimed at the three-dimensional augmentation of sound surrounding a vehicle, hereafter referred to as the ‘future system’ (FS), which could be used in FAD for entertainment and comfort. Three-dimensional sound is being developed as a means for providing feedback to humans (Lumbreras, Sánchez & Barcia, 1996; Garas, 2000; Rozier, 2000; Godinho, António & Tadeu, 2001; Dobler, Haller & Stampfl, 2002).

Our proposed FS filters out unwanted sounds (e.g., tire/engine noise coming from vehicles in the vicinity) and amplifies desired sounds (e.g., sound of birds singing in a park). We envision that such a system could be used in future fully automated vehicles. Vehicles driving fully automatically have full control of the vehicle and must have reliable detection capabilities of the environment. Drivers of such vehicles will not be required to pay attention to the processes that take place in the environment surrounding the car. Hence, a spatial augmentation of sounds that a driver prefers to hear and simultaneous cancelation of unwanted sounds may enhance the pleasure of being engaged in

FAD. Such system will probably have to be configurable: drivers must have the option to select which sounds they want to augment and which sounds they wish to filter out, as well as to adjust the volume of these sounds.

2.2.2.5 The aim of the present survey study

As mentioned above, feedback is important in HAD, especially regarding transitions of control. It is relevant for the development of automated driving systems to know what types of interfaces people want and need. Because automated cars do not exist yet on the consumer market, it is impossible to test such research questions in an ecologically valid environment, except through driving simulator research.

The present study was undertaken from a different point of view. We proceeded on the basis that respondents were asked to *imagine* automated driving scenarios. The aim of the present study was to investigate the opinion of people on two existing auditory displays (PA & FCWS) as well as the augmented sound system 'FS'. The respondents were asked to judge two qualities of the systems—helpfulness and annoyance—and state whether they would consider using such systems in the future. In addition, we asked people to report their preferred type of feedback for takeover requests in HAD. Statistical associations between self-reported driving style as measured with the Driver Behaviour Questionnaire (DBQ), yearly mileage, number of accidents, and opinions of respondents on the qualities of the proposed systems were assessed.

The hypothesis that people from non-English speaking countries prefer a female voice to a male voice in automated driving systems was also tested. Additionally, the respondents were asked to provide their general thoughts on the concept of automated driving in a free-response question. Finally, the respondents provided their opinion on the year of introduction of fully automated driving in their country of residence. Results of these analyses were compared with findings from two previous surveys that asked questions related to other aspects of automated vehicles (De Winter et al., 2015; Kyriakidis, Happee, & De Winter, 2015).

2.2.3 Methods

2.2.3.1 Survey

A survey containing 31 questions was developed with the online tool CrowdFlower (www.crowdflower.com). Table 1 shows the questions of the survey as well as the corresponding coding. The full survey is included in the supplementary material. The survey was targeted towards reasonably educated persons without knowledge of automated driving. A previous survey indicated that people who work on CrowdFlower-based surveys have mostly undergraduate degrees (Kyriakidis et al., 2015).

The present survey introduced in plain language three levels of driving: manual driving, partially automated driving, and fully automated driving. Manual driving was referred to as "normal (non automated) cars". The explanation of partially automated driving was provided as follows: "Imagine again that you are driving in an automated car (that can perform certain tasks without any interaction from the humans in the car). However, the automation cannot handle all possible situations, and you sometimes have to take over control".

Respondents were asked to imagine fully automated driving as follows: “Imagine a fully automated car (no steering wheel) that drives completely on its own with no manual interaction”.

The survey contained questions on the person’s age, gender, driving frequency, mileage, and accident involvement. The questions asking participants to provide information on their driving style were based on the violations scale of the DBQ, as used by De Winter (2013).

The respondents were asked to express their opinion on two currently existing systems and one proposed setup that could be used during fully automated driving. Specifically, we asked respondents about (1) a parking assistant (PA) in a manually driven car that produces warning sounds (beeps) when the car gets too close to a nearby object while parking, (2) a forward collision warning system (FCWS) in a manually driven car that provides a warning sound when a car is rapidly approaching another car in front, and (3) a future augmented surround sound system in a fully automated vehicle (FS). The FS was described as follows: “Now imagine that this fully automated car records what is happening outside and plays it via speakers inside the car, informing the occupants about the outside environment. In other words, those who sit in the car can hear what is happening outside even when their windows are closed. Sound volume in such system could be adjusted; particular noise (for example sound coming from another vehicle) could be filtered out. Such system could, for example, be used during a leisure drive through a park on a hot day”. Illustrations belonging to the three scenarios (i.e., PA, FCWS, FS) were used in the survey (Fig. 1). No auditory examples were used. The illustrations were uploaded to a remote site in order to be embedded to the survey. Supplementary material contains the XML code used to create the survey. If one wishes to add images to a CrowdFlower survey, the suggested method could be used.

Figure 1. Illustrations belonging to the three scenarios presented to the respondents. (A) Parking assistant (PA); (B) Forward collision warning system (FCWS); (C) Future system (FS).

The respondents were asked to indicate disadvantages of the PA (Q17) and FCWS (Q23) and to indicate advantages and disadvantages of the FS (Q26) by means of textual responses. The respondents also had the opportunity to indicate the preferred mode of feedback for receiving a takeover request (Q27 & Q28). In the last question (Q31), they were asked to “provide any suggestions, which could help engineers to build safe and enjoyable automated cars”. Giving a response to this last free-response question was optional. All examples of given comments shown in this article are direct quotes from the responses; no grammatical or syntactic errors were corrected. The respondents had to complete all questions (except Q28 & Q31), and each question had an *I prefer not to respond* response option.

Table 1. All survey items.

Variable	Question	Full question as reported in the survey	Used coding
Instr	Q1	Have you read and understood the above instructions?	1 = Yes, 2 = No
Gender	Q2	What is your gender? (1 = female, 2 = male)	-1 = I prefer not to respond, 1 = Female, 2 = Male
Age	Q3	What is your age?	Positive integer value
DriveFreq	Q4	On average, how often did you drive a vehicle in the last 12 months?	-1 = I prefer not to respond, 1 = Never, 2 = Less than once a month, 3 = Once a month to once a week, 4 = 1 to 3 days a week, 5 = 4 to 6 days a week, 6 = Every day
KmYear	Q5	About how many kilometres (miles) did you drive in the last 12 months?	-1 = I prefer not to respond, 1 = 0, 2 = 1 - 1,000, 3 = 1,001 - 5,000, 4 = 5,001 - 15,000, 5 = 15,001 - 20,000, 6 = 20,001 - 25,000, 7 = 25,001 - 35,000, 8 = 35,001 - 50,000, 9 = 50,001 - 100,000, 10 = more than 100,000
NrAcc	Q6	How many accidents were you involved in when driving a car in the last 3 years? (please include all accidents, regardless of how they were caused, how slight they were, or where they happened)?	-1 = I prefer not to respond, 1 = 0, 2 = 1, 3 = 2, 4 = 3, 5 = 4, 6 = 5, 7 = More than 5
Vangered	Q7	How often do you do the following?: Becoming angered by a particular type of driver, and indicate your hostility by whatever means you can.	-1 = I prefer not to respond, 1 = 0 times per month, 2 = 1 to 3 times per month, 3 = 4 to 6 times per month, 4 = 7 to 9 times per month, 5 = 10 or more times per month
Vmotorway	Q8	How often do you do the following?: Disregarding the speed limit on a motorway.	-1 = I prefer not to respond, 1 = 0 times per month, 2 = 1 to 3 times per month, 3 = 4 to 6 times per month, 4 = 7 to 9 times per month, 5 = 10 or more times per month
Vresident	Q9	How often do you do the following?: Disregarding the speed limit on a residential road.	-1 = I prefer not to respond, 1 = 0 times per month, 2 = 1 to 3 times per month, 3 = 4 to 6 times per month, 4 = 7 to 9 times per month, 5 = 10 or more times per month

Vfollowing	Q10	How often do you do the following?: Driving so close to the car in front that it would be difficult to stop in an emergency.	-1 = I prefer not to respond, 1 = 0 times per month, 2 = 1 to 3 times per month, 3 = 4 to 6 times per month, 4 = 7 to 9 times per month, 5 = 10 or more times per month
Vrace	Q11	How often do you do the following?: Racing away from traffic lights with the intention of beating the driver next to you.	-1 = I prefer not to respond, 1 = 0 times per month, 2 = 1 to 3 times per month, 3 = 4 to 6 times per month, 4 = 7 to 9 times per month, 5 = 10 or more times per month
Vhorn	Q12	How often do you do the following?: Sounding your horn to indicate your annoyance with another road user.	-1 = I prefer not to respond, 1 = 0 times per month, 2 = 1 to 3 times per month, 3 = 4 to 6 times per month, 4 = 7 to 9 times per month, 5 = 10 or more times per month
Vphone	Q13	How often do you do the following?: Using a mobile phone without a hands free kit.	-1 = I prefer not to respond, 1 = 0 times per month, 2 = 1 to 3 times per month, 3 = 4 to 6 times per month, 4 = 7 to 9 times per month, 5 = 10 or more times per month
Vmean	N/A	Mean for Q7-12	Numeric value
PApast	Q14	In the past month, did you drive a car with a parking assistant?	-1 = I prefer not to respond, 1 = I do not know, 2 = No, 3 = Yes
PAhelp	Q15	A parking assistant is helpful.	-1 = I prefer not to respond, 1 = Disagree strongly, 2 = Disagree a little, 3 = Neither agree nor disagree, 4 = Agree a little, 5 = Agree strongly
PAannoy	Q16	A parking assistant is annoying.	-1 = I prefer not to respond, 1 = Disagree strongly, 2 = Disagree a little, 3 = Neither agree nor disagree, 4 = Agree a little, 5 = Agree strongly
PAopin	Q17	What do you think are the disadvantages of a parking assistant?	Textual response
PAfut	Q18	I would like to have a system in my car that can park the car automatically, just by pressing a button.	-1 = I prefer not to respond, 1 = Disagree strongly, 2 = Disagree a little, 3 = Neither agree nor disagree, 4 = Agree a little, 5 = Agree strongly
FCWSpast	Q19	In the past month, did you drive a car with a forward collision warning system?	-1 = I prefer not to respond, 1 = I do not know, 2 = No, 3 = Yes

FCWShelp	Q20	A forward collision warning system is helpful.	-1 = I prefer not to respond, 1 = Disagree strongly, 2 = Disagree a little, 3 = Neither agree nor disagree, 4 = Agree a little, 5 = Agree strongly
FCWSannoy	Q21	A forward collision warning system is annoying.	-1 = I prefer not to respond, 1 = Disagree strongly, 2 = Disagree a little, 3 = Neither agree nor disagree, 4 = Agree a little, 5 = Agree strongly
FCWSfut	Q22	I would you like to have a system in my car that brakes automatically to avoid collisions (Autonomous Emergency Braking).	-1 = I prefer not to respond, 1 = Disagree strongly, 2 = Disagree a little, 3 = Neither agree nor disagree, 4 = Agree a little, 5 = Agree strongly
FCWSopin	Q23	What do you think are the disadvantages of a forward collision warning system?	Textual response
FSannoy	Q24	I believe that this type of surround sound system would be annoying.	-1 = I prefer not to respond, 1 = Disagree strongly, 2 = Disagree a little, 3 = Neither agree nor disagree, 4 = Agree a little, 5 = Agree strongly
FSfut	Q25	I would prefer to use such a sound system instead of opening the window, when driving through a scenic place (for example, a national park).	-1 = I prefer not to respond, 1 = Disagree strongly, 2 = Disagree a little, 3 = Neither agree nor disagree, 4 = Agree a little, 5 = Agree strongly
FSopin	Q26	What would be the advantages and the disadvantages of such sound system?	Textual response
TORint	Q27	Now imagine again that you are driving in an automated car (that can perform certain tasks without any interaction from the humans in the car). However, the automation cannot handle all possible situations, and you sometimes have to take over control. What type of warning signal would you like to receive in case manual take over is required?	1 = Warning sound: one beep, 2 = Warning sound: two beeps, 3 = Warning sound: horn sound, 4 = Warning sound: bell sound, 5 = Warning light, 6 = Visual warning message projected on windscreen ‘Take over please’, 7 = Vibrations in your seat, 8 = Vibrations in your steering wheel, 9 = Vibrations in your seatbelt, 10 = Vibrations in the floor, 11 = Female voice: ‘Take over please’, 12 = Male voice: ‘Take over please’, 13 = Other, 14 = None of the above

TORintot	Q28	If you answered 'Other' in the previous question, please specify what type of warning signal you would like to receive in the described scenario.	Textual response
FACpref	Q29	I would prefer to drive in a fully automated car rather than a normal (non automated) car.	-1 = I prefer not to respond, 1 = Disagree strongly, 2 = Disagree a little, 3 = Neither agree nor disagree, 4 = Agree a little, 5 = Agree strongly
YearAuto	Q30	In which year do you think that most cars will be able to drive fully automatically in your country of residence?	Year
Comm	Q31	Please provide any suggestions which could help engineers to build safe and enjoyable automated cars.	Textual response
SurvTime		Survey time (derived from results generated by Crowdflower)	Seconds

2.2.3.2 Configuration of CrowdFlower

In the instructions, the respondents were informed that they would need approximately 10 min to complete the survey. The task expiration time was set to 30 min. Contributors from all countries were allowed to participate in the survey, in order to collect data from an as large and diverse population as possible. Moreover, the lowest level of experience of contributors 'Level 1 contributors' was selected. This level of experience accounts for 60% of completed work on CrowdFlower. As a result, the survey was available to a large number of workers, which allowed reaching a relatively diverse group of users of the platform. Completing the survey more than once from the same IP address was allowed (note, however, that responses from the same IP address were filtered out in our analyses, see results section). For the completion of the survey a payment of $0.15 was offered, and 2,000 responses were collected. The study was preceded by a pilot test with 10 respondents. The pilot test did not lead to any changes in the survey, and these 10 respondents were not included in the analysis.

2.2.3.3 Analyses

Descriptive statistics (i.e., mean, median, standard deviation, skewness, and number of responses) were calculated for each of the variables. The skewness was calculated as the third central moment divided by the cube of the standard deviation. A Spearman correlation matrix among the variables was created. The first author manually performed the analysis of textual responses (Q17, Q23, Q26, Q28, & Q31).

CrowdFlower automatically provides the respondent's country based on his/her IP address. We analysed the preferences of people from English

speaking countries, as defined by the UK government (UK Visas & Immigration, 2014): Antigua and Barbuda, Australia, Bahamas, Barbados, Belize, Canada, Dominica, Grenada, Ireland, Jamaica, New Zealand, Saint Lucia, Trinidad and Tobago, United Kingdom, and the United States) versus non-English speaking countries regarding the use of a male or female voice for supporting takeover requests during highly automated driving. Supplementary material contains the MATLAB script used to analyse the data.

2.2.3.4 Ethics statement

All data were collected anonymously. The research was approved by the Human Research Ethics Committee (HREC) of the Delft University of Technology. Documented informed consent was obtained via a dedicated survey item asking whether the respondent had read and understood the survey instructions.

2.2.4 Results

2.2.4.1 Number of respondents and respondent satisfaction

In total, 2,000 surveys were completed. The responses were gathered on 2 September 2014 between 15:00 and 20:15 (CET). The survey received an overall satisfaction rating of 4.4 out of 5.0. Additionally, the respondents ranked clearness of the instructions as 4.4 / 5.0, fairness of the questions as 4.2 / 5.0, easiness of the survey as 4.2 / 5.0, and the offered payment as 4.1 / 5.0.

2.2.4.2 Data filtering

The respondents who indicated they had not read the instructions (N = 10), who indicated they were under 18 and thereby did not adhere to the survey instructions (N = 6), who chose the *I prefer not to respond* or *I do not know* options in one or more of the multiple choice questions (N = 231), who indicated they never drive (N = 193), or who indicated they drive 0 km per year (N = 191) were excluded from the analyses. Since no limitations were applied on the number of responses that could be generated per IP address, some people completed the survey more than once. Such behaviour was seen as an indication that these persons participated in the survey primarily because of monetary gain. Thus, we applied a strict filter, and all data generated from non-unique IP addresses were removed (N = 465). In total, 795 surveys were removed, leaving 1,205 completed surveys for further analysis.

For the question “In which year do you think that most cars will be able to drive fully automatically in your country of residence?”, non-numeric responses (e.g., a year complemented by words such as “maybe 2030”, or “never”) and answers before the year 2014 were excluded, leaving 1,082 numeric responses.

2.2.4.3 Analyses at the individual level

The 1,205 respondents were from 91 countries (all 2,000 responses were associated with 95 countries). Descriptive statistics for all variables are listed in Table 2. The respondents took on average 9.2 minutes to complete the survey (SD = 5.6 min, median = 7.7 min). The supplementary material contains the entire Spearman correlation matrix. The correlation coefficients between variables that

related to questions about the PA, FCWS, and FS (PApast, PAhelp, PAannoy, PAfut, FCWSpast, FCWShelp, FCWSannoy, FCWSfut, FSannoy, & FSfut) on the one hand, and Age, DriveFreq, KmYear, NrAcc, the DBQ variables (Vangered, Vmotorway, Vresident, Vfollowing, Vrace, Vhorn, & Vphone), YearAuto, and SurvTime, on the other, were overall small, between −0.15 and 0.25.

Table 2. Descriptive statistics for the survey items (*N* = 1,205).

Variable	Mean	Median	SD	Skewness	Min	Max
Gender	1.75	2	0.43	-1.17	1	2
Age	31.94	30	10.49	1.04	18	73
DriveFreq	4.72	5	1.21	-0.66	2	6
KmYear	4.09	4	1.78	0.92	2	10
NrAcc	1.47	1	0.94	2.88	1	7
Vangered	1.86	2	0.86	1.46	1	5
Vmotorway	1.85	2	1.05	1.54	1	5
Vresident	1.70	1	1.01	1.79	1	5
Vfollowing	1.45	1	0.77	2.07	1	5
Vrace	1.32	1	0.69	2.62	1	5
Vhorn	1.86	2	1	1.41	1	5
Vphone	1.64	1	1.01	1.84	1	5
Vmean	1.67	1.57	0.57	1.36	1	4.71
PApast	2.27	2	0.45	1.03	2	3
PAhelp	4.33	5	0.88	-1.38	1	5
PAannoy	2.35	2	1.18	0.39	1	5
PAfut	3.87	4	1.24	-0.93	1	5
FCWSpast	2.10	2	0.30	2.65	2	3
FCWShelp	4.11	4	1.04	-1.14	1	5
FCWSannoy	2.56	3	1.26	0.27	1	5
FCWSfut	3.77	4	1.22	-0.80	1	5
FSannoy	3.21	3	1.22	-0.18	1	5
FSfut	3.00	3	1.29	-0.09	1	5
FACpref	3.01	3	1.33	-0.05	1	5
YearAuto	2078.33	2030	713.77	30.73	2014	25000
SurvTime	553.95	462	338.41	1.34	58	1810

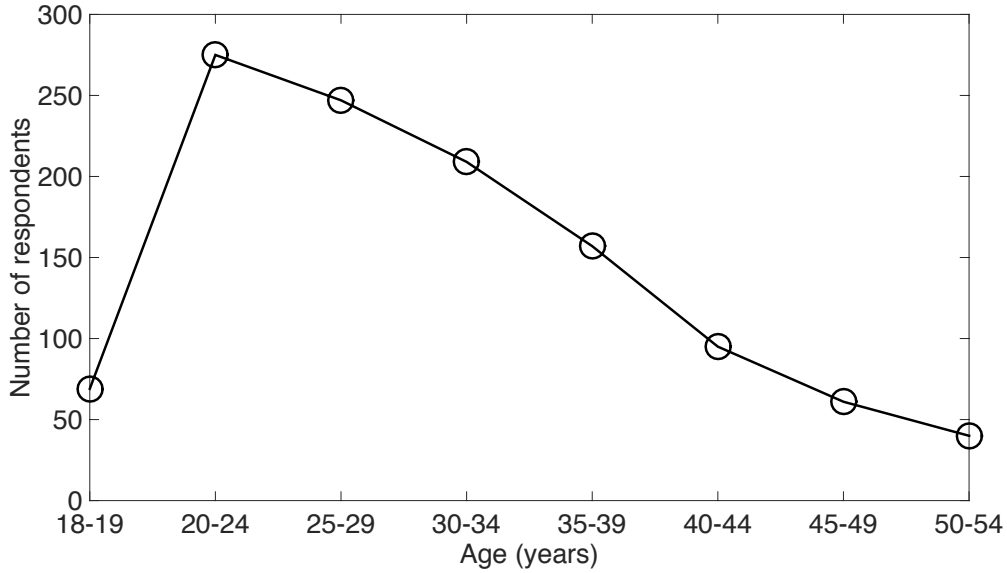


Figure 2. Distribution of the age of the respondents aged between 18 and 54 years.

The respondents' mean and median age were 31.9 and 30 years, respectively. Figure 2 shows the distribution of the respondents in 5-year wide age groups. 75.2% of the respondents were male (906 men vs. 299 women). The frequencies of the answers are provided in Table 3.

Table 3. Frequencies of answers.

Variable	1	2	3	4	5	6	7	8	9	10
Gender	299	906								
DriveFreq	0	79	108	293	313	412				
KmYear	0	250	245	337	148	76	77	49	16	7
NrAcc	855	226	79	19	13	6	7			
Vangered	417	629	96	33	30					
Vmotorway	541	466	97	39	62					
Vresident	663	379	78	32	53					
Vfollowing	807	295	70	20	13					
Vrace	939	182	60	15	9					
Vhorn	513	480	123	44	45					
Vphone	730	304	89	34	48					
PApast	17	865	323							
PAhelp	14	39	134	367	651					
PAannoy	383	297	290	196	39					
PAfut	86	109	175	338	497					
FCWSpast	28	1058	119							
FCWShelp	34	68	178	373	552					
FCWSannoy	331	254	324	208	88					
FCWSfut	82	123	203	377	420					
FSannoy	126	218	346	309	206					
FSfut	204	222	312	301	166					
FACpref	204	252	264	292	193					

Figure 3 shows that the respondents expected most cars to be able to drive in fully automated mode in their countries of residence around 2030 (median response), with a highly skewed distribution.

The respondents were asked to provide their opinion on two characteristics of the PA and FCWS systems, annoyance and helpfulness, and whether they would be willing to have automated versions of such systems in their own cars (Q18 for the PA & Q22 for the FCWS), all questions on a scale from 1 (*disagree strongly*) to 5 (*agree strongly*). Figure 4 shows the results for these questions.

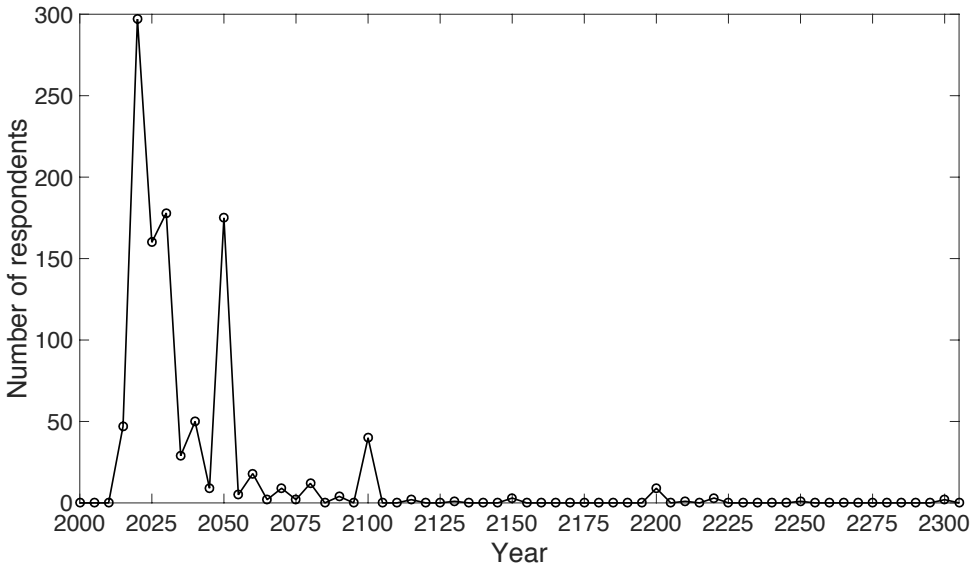


Figure 3. Distribution of responses for the question: "In which year do you think that most cars will be able to drive fully automatically in your country of residence?" (Q30). Years were divided into 5-year-wide bins.

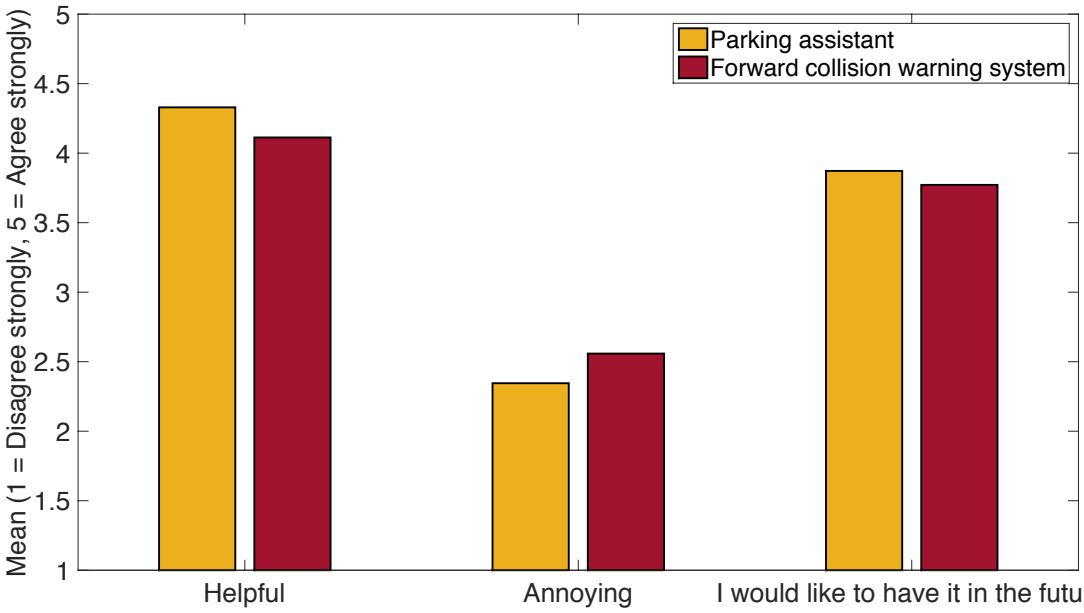


Figure 4. Opinion of the respondents on whether a PA and FCWS are helpful and annoying, and whether they would like to have automated versions of such systems in their cars in the future.

Figure 5 shows associations between the opinion of the respondents on annoyance and helpfulness of the PA and FCWS and their age divided into 5-year wide bins. Figure 5a shows that younger respondents found that both the PA ($\rho = -0.05$, $p = .103$) and the FCWS ($\rho = -0.14$, $p < .001$) were more annoying, but these effects were weak. The Spearman correlation between the respondents' age and the reported annoyance of the FS was weak as well ($\rho = 0.06$, $p = .035$). Figure 5b shows that the perceived helpfulness of the FCWS ($\rho = 0.12$, $p < .001$) slightly increased with age. People who found the PA annoying typically indicated that the FCWS was also annoying ($\rho = 0.47$, $p < .001$), and

respondents who thought that the PA was helpful, considered the FCWS to be helpful as well (ρ = 0.34, $p < .001$).

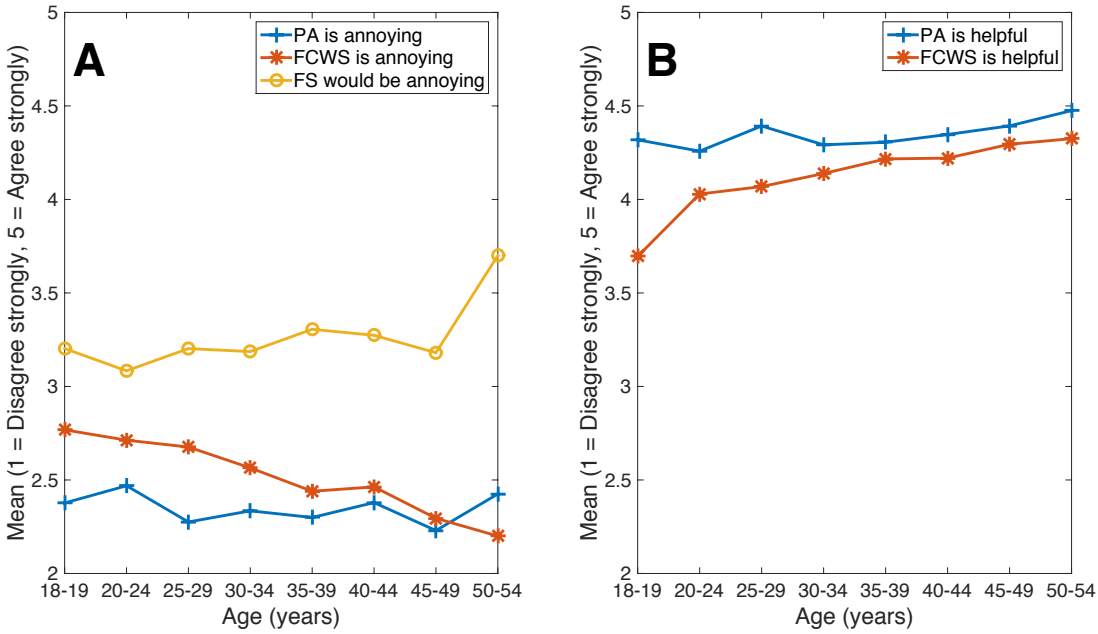


Figure 5. (A) Opinion on the annoyance of the parking assistant (PA, Q16), forward collision warning system (FCWS, Q21), and future system (FS, Q24) as a function of age; (B) Opinion on the helpfulness of the PA (Q15) and FCWS (Q20) as a function of age. Age was divided into 5-year-wide bins.

Figure 6 shows the respondents' opinion on the proposed future system. The respondents were asked whether they would find such system annoying (Q24) and whether they would prefer to use such system instead of opening windows while driving in a fully automated car through a scenic place (Q25). A large portion of the respondents was neutral in their responses: 346 people chose the option *Neither agree nor disagree* in Q24, and 312 persons chose the same option in Q25.

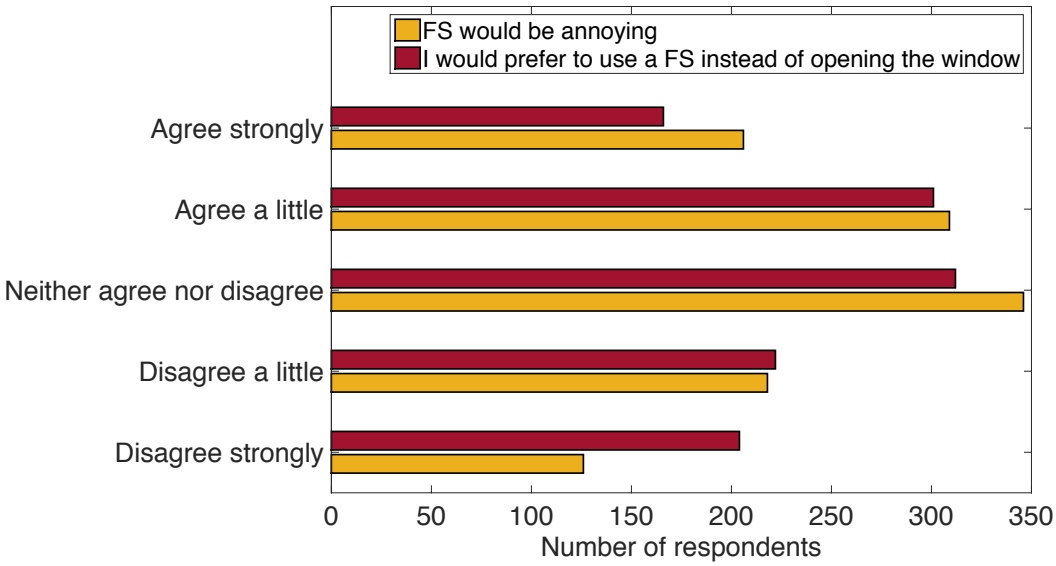


Figure 6. Distribution of opinions on whether the proposed future system (FS) would be annoying (Q24) and whether the respondents would prefer the system to opening windows in fully automated cars (Q25).

In Q27 the respondents were asked to report on the types of feedback that they would like to be supported by in case of a takeover request during highly automated driving. The respondents were allowed to select multiple options. Figure 7 shows that a large number of people preferred auditory feedback provided by a female voice saying 'Take over please' (*N* = 514). The number of respondents who chose the option with the male voice was considerably lower (*N* = 244). Figure 7 makes a distinction between the numbers of female and male respondents. It is apparent that both female and male respondents preferred the female over the male voice.

Other types of auditory feedback were reported in the following order: two beeps (*N* = 375), one beep (*N* = 195), a bell sound (*N* = 194), and a horn sound (*N* = 135). The respondents indicated a high level of support for both visual signals offered in the question: a warning message projected on the windscreen 'Take over please' (*N* = 429) and a warning light (*N* = 406). However, respondents showed a relatively low level of acceptance of the offered variations of a vibration interface: vibrations in the seat (*N* = 341), vibrations in the steering wheel (*N* = 179), vibrations in the seatbelt (*N* = 117), and vibrations in the floor (*N* = 64). Furthermore, the results seem to suggest that female respondents were less likely than male respondents to prefer a female voice.

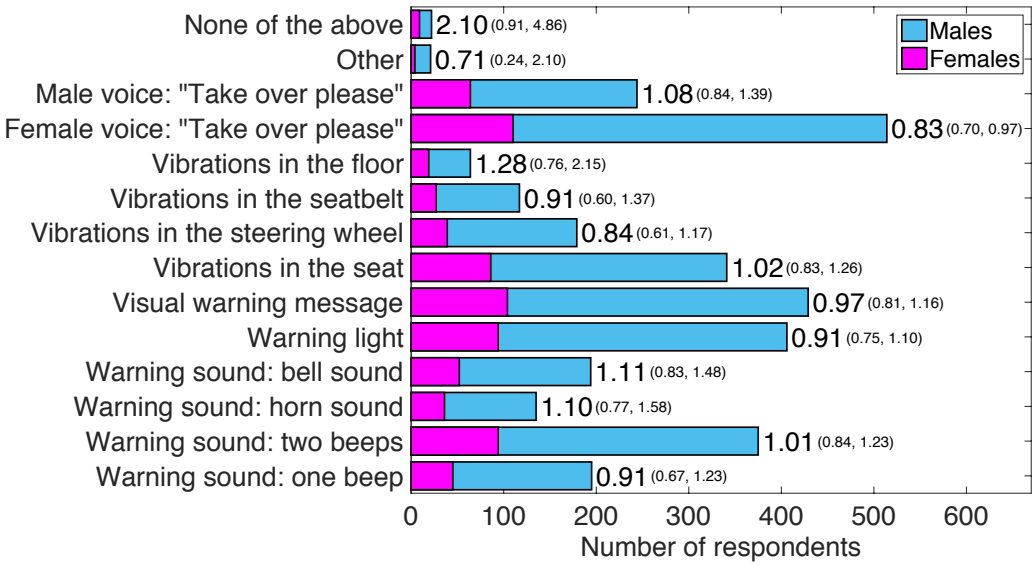


Figure 7. Numbers of respondents who indicated a preference for a particular takeover request during highly automated driving in the question: "*Now imagine again that you are driving in an automated car (that can perform certain tasks without any interaction from the humans in the car). However, the automation cannot handle all possible situations, and you sometimes have to take over control. What type of warning signal would you like to receive in case manual take over is required?"* (Q27). Each bar is supplemented by the corresponding 'risk ratios' of female respondents, calculated as the proportion of females who indicated this answer divided by the number of males who indicated this answer. If the risk ratio is greater than 1 females are overrepresented. Conversely, if the risk ratio is smaller than 1, females are underrepresented. 95% confidence intervals are shown in parentheses.

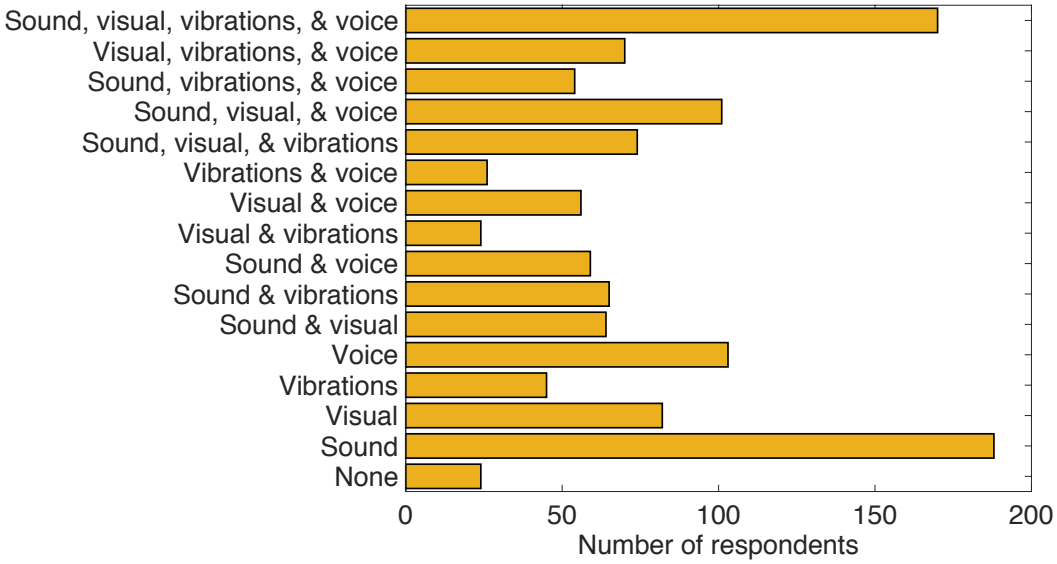


Figure 8. Preference to combinations of types of signals for aiding takeover requests during highly automated driving (Q27). All possible combinations are listed. Hence, the total number of respondents adds up to 1,205.

Figure 8 shows the opinion of the respondents on the combinations of warning signals. The figure shows that most people (*N* = 188) preferred a sound signal (i.e., one or two beeps, a horn, or bell) without additional information. A large number of people indicated that they would like to receive a combination of all four modalities (*N* = 170) or the combination of a sound signal, a visual message, and a voice (*N* = 101).

2.2.4.4 Cross-national differences in opinion for feedback for takeover requests

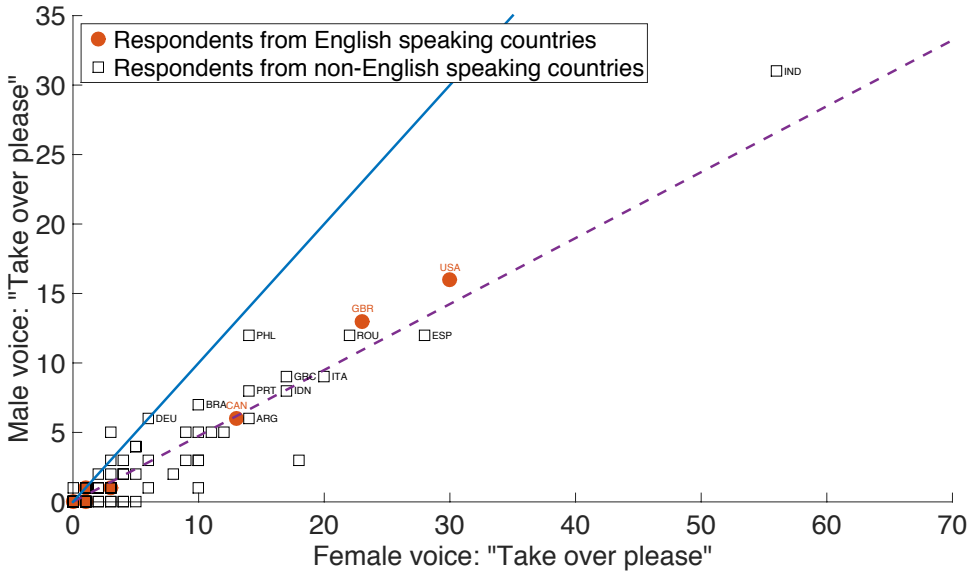


Figure 9. Numbers of respondents from English and non-English speaking countries who indicated a preference for a male voice and a female voice for a takeover request during highly automated driving (Q27). The dashed line represents the ratio between the number of respondents who preferred a female voice and the number of respondents who preferred a male voice. The solid line is the line of unity. No labels for shown for countries with five or less respondents indicating a male voice, to support the clarity of the figure. Country abbreviations are listed according to ISO 3166-1 alpha-3.

Next, we tested the hypothesis whether peoples' preference for a female and male voice in supporting takeover requests in highly automated driving was different between English and non-English speaking countries. Figure 9 presents a scatter plot, showing the numbers of respondents per country who indicated that they would like to receive a female or a male voice. The overall percentage of respondents who expressed preference for a female voice was 43% (514 / 1,205), and the overall percentage of people who expressed preference for a male voice was 20% (244 / 1,205). The corresponding percentages were 42% (71 / 168) and 22% (37 / 168) for English speaking countries, and they were 43% (443 / 1,037) and 20% (207 / 1,037) for non-English speaking countries. The differences between English speaking countries and non-English speaking countries were not statistically significant (female voice: RR = 0.99, 95% CI = [0.82, 1.20]; male voice: RR = 1.10, 95% CI = [0.81, 1.50]).

2.2.4.5 Analyses of textual comments

The respondents provided their feedback on the disadvantages of the PA in Q17. The responses that were less than five characters long (*N* = 181) or that were not written in English (*N* = 39) were ignored. Comments were processed before data filtering and were hence based on all 2,000 responses. 12.4% of the respondents (*N* = 151) provided negative feedback on the auditory interfaces in parking assistants. Many people (*N* = 135) indicated that PA systems were annoying, for example: "*Sound should not be too loud and annoying*" and *"I think it could be annoying especially when your focusing".* Thirty-seven respondents pointed out that the PA used overly loud sounds. Several answers to the question contained comments that the PA sounds can be distracting (*N* = 21) and inaccurate (*N* = 48). Five respondents indicated that they would prefer feedback in other types of modalities, for example: "*annoying, use something else instead of the constant loud beeping sounds"* and *"The sound, a voice message would be better".* Five respondents indicated that the PA systems cannot be used by deaf people.

The respondents indicated their opinion on the disadvantages of the FCWS in Q23. The responses that were less than five characters long (*N* = 276) or that were not written in English (*N* = 35) were not included in the analysis. Sixteen respondents indicated that the auditory feedback used in FCWS was annoying, for example: "*This situation might come up too often so the warning sound may get annoying fast*" and "*The beeps might feel annoying*".

Next, the respondents were asked to comment on possible advantages and disadvantages of the FS in Q26. The responses that were less than five characters long (*N* = 138) or that were not written in English (*N* = 46) were not included in the analysis. In total, 1,249 comments were analysed. A collection of mixed responses was received. Overall, more comments were classified as positive (*N* = 132) than negative (*N* = 52) to the FS. However, the respondents also pointed out concerns about a number of characteristics that they associated with the system: annoyance to both the driver and to other road users in the traffic (*N* = 101), distraction (*N* = 47), and loudness (*N* = 28). Fifty-five respondents expressed their concerns that the system would be impractical; however, most of such concerns could be associated with the lack of understanding of the concept of a fully automated car. Certain respondents showed a high level of negativity caused by an apparent lack of understanding

the concept of filtering only specific sounds coming from the outside environment. Examples are: "*You can not hear some bells or signal from other cars", "Main disadvantage: makes driver unaware of any dangers*", "*If car noises are filtered out how would you hear if another car is incoming*", and "*I feel that filtering other car noise may be dangerous*".

In Q27 the respondents were asked to indicate their preference for types of interfaces to be used for takeover requests in HAD. One of the options in that question was "*Other*". If respondents selected this option, they were invited to provide further comments in Q28. The responses that were less than five characters long (N = 32) or that were not written in English (N = 1) were ignored. In total, 22 responses were analysed. One respondent indicated that he/she would prefer to be aided by continuous beeps until he/she reclaimed control. Another respondent stated "*steering wheel up or down motion to signal steering wheel usage needed, accompanied by a specific message*". One respondent mentioned that interfaces used in such scenario need to be adaptive depending on the urgency of the request "*It honestly depends on the situation the car needs me to take over for. Does it affect anyone's safety at all? Does it actually /need/ to be done straight away? Is it critically important in any other way? In those cases I'd obviously like a very noticeable signal however 'annoying' it may be. In other situations however I'd prefer a decent text message or a gentle reminder*".

2.2.5 Discussion

The aim of this study was to obtain opinions on preferred feedback types for takeover requests in HAD from a large number of people coming from all over the globe. Additionally, the aim was to measure perceived helpfulness and annoyance of auditory interfaces for three systems. Specifically, the respondents who participated in the survey were presented with two existing systems used in modern vehicles (a parking assistant [PA] & a forward collision warning system [FCWS]) and one futuristic setup (FS) envisioned for FAD. Respondents were asked whether they would consider using the proposed FS in future automated vehicles. Our survey helped us to gather opinions from people before technology is actually available.

Previous research suggests that the modality of aiding systems in automated cars should be chosen carefully to avoid frustration of people who will be using such vehicles and to increase safety of automation on public roads. Stanton, Young, and McCaulder (1997) expressed concerns that interfaces currently employed in ACC do not support the understanding of the behaviour and limitations of the system. A driving simulator study by Adell et al. (2008) provided a comprehensive analysis of combinations of interfaces for supporting safe driving. Participants in that study were most positive about the haptic interface, while the auditory warning signals were not highly appreciated, which may be explained by the nature of the experiment that exposed the participants to a high urgency scenario of avoiding rear-end collisions (Adell et al., 2008).

Findings from the aviation field show that the female voice is more difficult to understand in noisy environments (Nixon et al., 1998a). It has also been argued that the female voice has the advantage that it stands out more in a predominantly male environment, such as the military (Noyes, Hellier, & Edworthy, 2006). However, differences in speech intelligibility and perceived urgency between male and female voices are generally small and findings have

been mixed (e.g., Arrabito, 2009; Edworthy, Hellier, & Rivers, 2003; Nixon et al., 1998b). However, it has been found that most people normally use a female voice when using their route navigation device (Large & Burnett, 2013). In the present research, respondents were asked to select the types of interfaces they are willing to be guided by during a takeover request. The results of our survey further showed that the female voice is preferred in both English and non-English speaking countries. Thus, our findings reinforce the idea that the overall most preferred way to support the transition of control is an auditory instruction performed with a female voice. Note that determining the language of respondents based on their IP address cannot guarantee accurate results in all cases. In future surveys adding a question prompting for the participant's spoken language may yield more accurate results.

It was found that the participants showed a relatively low level of appreciation of vibratory interfaces, which contrasts with the findings in Adell et al. (2008). This could be due to the fact that only a small number of systems that feature vibratory feedback are available in modern vehicles. A relatively large number of people indicated that they would like to be aided by all four proposed modalities. In addition, a large number of respondents indicated that the combination of a sound signal, a visual message, and a vibration signal would be preferable during takeover requests in highly automated driving. This is a surprising finding as such a combination is not common in current cars. A possible explanation of this finding could be that the respondents misinterpreted the question and instead of indicating their preference for multimodal feedback, expressed their preferences for the types of feedback that can be used separately from each other during takeover requests in highly automated driving. Another limitation of the present study is that we did not vary possible parameters of the feedback signals, including the quality, intensity, timing, and speed of delivery of the take-over requests. Future experimental research could investigate the effects of such parameters.

The existing systems—the PA and FCWS—received favourable ratings, which may not be surprising, since these systems have already been tested and are already available on the market. One limitation in this context is that the participants relied on a narrative description, complemented with a visual illustration; the survey did not contain actual examples of auditory cues. Before the initiation of the survey, it was believed that the proposed FS would be seen as a way to enhance the enjoyment of driving a car through a scenic place. The results showed that the participants were rather sceptical about such a concept: the system was perceived as somewhat annoying, with a mean score of 3.21 to question Q24 on the scale from disagree strongly (1) to agree strongly (5). The proposed FS was not highly rated, possibly because the concept was perceived as a bad idea, because of a lack of previous experience with such system, or because people could not envision it due to the lack of a realistic representation (see also the analysis of the textual comments). It should also be noted that a large proportion of respondents selected the middle option *Neither agree nor disagree,* possibly indicating difficulties with understanding the concept of the proposed system (for studies into middle category endorsement, see Kulas, Stachowski & Haynes, 2008; Kulas & Stachowski, 2009; Sturgis, Roberts & Smith, 2012).

Small effects of age on the acceptance of FAD were previously reported by Payre, Cestac and Delhomme (2014). In the present study, we also observed

small age effects regarding the self-reported annoyance of the three proposed systems: younger participants saw the PA and FCWS as more annoying than older respondents did. However, young respondents perceived the FS as less annoying than the older respondents. It is known that younger people are more likely to accept new technologies (Lee, 2007; Tacken et al., 2005), and thus may be more successful at envisioning such abstract concepts as the FS. A somewhat stronger age effect was observed regarding helpfulness: older respondents found the FCWS more helpful than the younger participants. It is known that young people feel more confident behind the wheel (Matthews & Moran, 1986; Lee et al., 2002; Lee, 2007; Clarke, Ward & Truman, 2005), and therefore may think they need less external help than older drivers.

CrowdFlower offers a platform that supports full anonymity of participants. This anonymity may have encouraged respondents to express their thoughts freely, without the fear of being judged by the organizers of the survey. All but the last free-response items required people to enter at least one character. A large number of respondents did not provide meaningful comments. However, a substantial portion of respondents did provide valuable answers, facilitating the understanding of what people think about not only the use of auditory interfaces in future highly and fully automated cars, but also about the concept of automated driving in general. Numerous respondents expressed their concerns about the qualities of current PA and FCWS systems. Some participants suggested that they want to be aided by visual and vibratory feedback in addition to auditory feedback. A number of people indicated the inaccessibility of modern PAs and FCWSs to deaf users. However, current systems also provide haptic and/or visual cues (BMW, 2013; Volvo, 2013b; Ford, 2014; Honda, 2014), and so people with a hearing impairment could still benefit from such multimodal feedback. Some respondents were sceptical about the introduction of highly and fully automated vehicles, which may be related to general consumer scepticism about new technologies. Respondents expected that most cars would drive fully automatically by the year 2030 (median value), a result that matches findings in previously published research (Sommer, 2013; De Winter et al., 2015; Kyriakidis et al., 2015).

The total cost of the study performed by means of a crowdsourced online-based survey was lower than what is offered by companies that conduct similar surveys with help of classic recruitment methods (De Winter et al., 2015). A group of people filled in the survey more than once, and we reasoned that their responses ought not to be trusted. Thus, we applied a strict filter, and removed all respondents who filled out the survey more than once. We also excluded all people who had one or more missing items. With appropriate data quality control mechanisms, crowdsourcing is known to be a powerful research tool (Howe, 2006; Kittur, Chi & Suh, 2008; Mason & Suri, 2012; Crump, McDonnell & Gureckis, 2013). Nonetheless, as with any self-report questionnaire, the validity of the results is limited to what people can imagine or retrieve from their memory. Furthermore, CrowdFlower respondents are not representative of the entire population of stakeholders of future HAD cars. It is likely that highly automated vehicles will initially be purchased by wealthy people, while projects on CrowdFlower are often completed by people from low-income countries (Kyriakidis et al., 2015).

In conclusion, the present survey study showed that the PA and FCWS were well appreciated by respondents, whereas the proposed future system (FS)

was not rated highly. A second conclusion is that the female voice is the most preferred takeover request among the offered options. The scientific community and the automotive industry may be able to use the information gathered in the present survey for the development of automated driving systems, in particular future iterations of parking assistants and forward collision warning systems, as well as for the design of human-machine interfaces for automated driving.

2.2.6 Acknowledgements

We would like to express our special gratitude to Daria Nikulina for designing the illustrations used in the survey.

2.2.7 Supplementary material

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.13#supplemental-information

2.2.8 References

Adell, E., Varhelyi, A., Fontana, M. D., & Bruel, L. (2008). Test of HMI alternatives for driver support to keep safe speed and safe distance – a simulator study. *The Open Transportation Journal,* 2, 53–64. https://doi.org/10.2174/1874447800802010053

Alicandri, E., & Moyer, J. (1992). Human factors and the automated highway system. Proceedings of *the Human Factors and Ergonomics Society Annual Meeting*, 36, 1064–1067. https://doi.org/10.1518/107118192786749793

Arrabito, G. R. (2009). Effects of talker sex and voice style of verbal cockpit warnings on performance. *Human Factors, 51*, 3–20. https://doi.org/10.1177/0018720808333411

Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19, 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

Banks, V. A., Stanton, N. A., & Harvey, C. (2014). Sub-systems on the road to vehicle automation: Hands and feet free but not 'mind' free driving. *Safety Science,* 62, 505–514. https://doi.org/10.1016/j.ssci.2013.10.014

Barón, A., & Green, P. (2006). *Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review* (Technical Report No. UMTRI-2006-5). Ann Arbor, MI: Transportation Research Institute, University of Michigan. *Available at http://www.umich.edu/~driving/publications/UMTRI-2006-5a.pdf*

Begault, D. R. (1993). Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation. *Human Factors, 35*, 707–717. https://doi.org/10.1177/001872089303500409

Beiker, S. A. (2012). Legal aspects of autonomous driving: The need for a legal infrastructure that permits autonomous driving in public to maximize safety and consumer benefit. *Santa Clara Law Review*, *52*, 1145–1561.

Belz, S. M., Robinson, G. S., & Casali, J. G. (1999). A new class of auditory warning signals for complex systems: Auditory icons. *Human Factors, 41*, 608–618. https://doi.org/10.1518/001872099779656734

Bergenhem, C., Pettersson, H., Coelingh, E., Englund, C., Shladover, S., & Tsugawa, S. (2012). Overview of platooning systems. Proceedings of *the*

19th ITS World Congress, Vienna, Austria. Available at http://publications.lib.chalmers.se/records/fulltext/174621/local_174621.pdf

BMW. (2013). BMW ConnectedDrive from A to Z. Available at http://www.bmw.com/com/en/insights/technology/connecteddrive/2013/a_to_z/index.html

BMW. (2014). BMW X5: Park Assistant. Available at http://www.bmw.com/com/en/newvehicles/x/x5/2013/showroom/driver_assistance/park_assistant.html#t=l

Bosker B. (2013). Why Siri's voice is now a man (and a woman). Available at http://www.huffingtonpost.com/2013/06/11/siri-voice-man-woman_n_3423245.html

Braitman, K. A., McCartt, A. T., Zuby, D. S., & Singer, J. (2010). Volvo and Infiniti drivers' experiences with select crash avoidance technologies. *Traffic Injury Prevention, 11*, 270–278. https://doi.org/10.1080/15389581003735600

Brandenburg, S., & Skottke, E. (2014). Switching from manual to automated driving and reverse: Are drivers behaving more risky after highly automated driving? Proceedings of *the IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), Qingdao, China*, 2978–2983. https://doi.org/10.1109/ITSC.2014.6958168

Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. Cambridge, MA: MIT Press.

Bronkhorst, A. W., Veltman, J. A., & Van Breda, L. (1996). Application of a three-dimensional auditory display in a flight task. *Human Factors, 38*, 23–33. https://doi.org/10.1518/001872096778940859

Clarke, D. D., Ward, P., & Truman, W. (2005). Voluntary risk taking and skill deficits in young driver accidents in the UK. *Accident Analysis & Prevention, 37*, 523–529. https://doi.org/10.1016/j.aap.2005.01.007

Continental. (2012). Continental strategy focuses on automated driving. Available at https://www.conti-online.com/www/pressportal_com_en/themes/press_releases/1_topics/automated_driving_en/pr_2012_12_18_automated_driving_en.html

Cooke, M., & Ellis, D. P. W. (2001). The auditory organization of speech and other sources in listeners and computational models. *Speech Communication, 35*, 141–177. https://doi.org/10.1016/S0167-6393(00)00078-9

Crump, M. J. C., McDonnell, J. V, & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE, 8*, e57410. https://doi.org/10.1371/journal.pone.0057410

Dagan, E., Mano, O., Stein, G. P., & Shashua, A. (2004). Forward collision warning with a single camera. Proceedings of *IEEE Intelligent Vehicles Symposium, Parma, Italy*, 71–79. https://doi.org/10.1109/IVS.2004.1336352

Daimler. (2013). DISTRONIC PLUS: warns and assists the driver. Available at http://www.daimler.com/dccom/0-5-1210218-1-1210321-1-0-0-1210228-0-0-135-0-0-0-0-0-0-0-0.html

De Winter, J. C. F. (2013). Predicting self-reported violations among novice license drivers using pre-license simulator measures. *Accident Analysis & Prevention*, 52, 71–79. https://doi.org/10.1016/j.aap.2012.12.018

De Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 196–217. https://doi.org/10.1016/j.trf.2014.06.016

De Winter, J. C. F., Kyriakidis, M., Dodou, D., & Happee, R. (2015). Using CrowdFlower to study the relationship between self-reported violations and traffic accidents. Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics (AHFE), Las Vegas, NV*.

Desmond, P., Hancock, P., & Monette, J. (1998). Fatigue and automation-induced impairments in simulated driving performance. *Transportation Research Record: Journal of the Transportation Research Board*, 1628, 8–14. https://doi.org/10.3141/1628-02

Dingus, T. A., Hulse, M. C., & Barfield, W. (1998). Human-system interface issues in the design and use of advanced traveler information systems. In Barfield W, Dingus TA, eds. *Human Factors in Intelligent Transportation Systems*. Mahwah, NJ: Lawrence Erlbaum Associates, 359–395.

Dobler, D., Haller, M., & Stampfl, P. (2002). ASR – augmented sound reality [abstract]. *Conference abstracts and applications on – SIGGRAPH '02*. New York, NY: ACM Press, 148.

Ebbatson, M. (2009). The loss of manual flying skills in pilots of highly automated airliners. PhD Thesis, Cranfield University.

Edworthy, J., Hellier, E., & Rivers, J. (2003). The use of male or female voices in warnings systems: a question of acoustics. *Noise & Health, 6*, 39–50.

European Commision. (2011). Roadmap to a single European transport area – towards a competitive and resource efficient transport system (White Paper, COM(2011) 144 final), Brussels.

Ford. (2014). Collision warning with brake support. Available at http://support.ford.com/maintenance/collision-warning-brake-support

Garas, J. (2000). *Adaptive 3D sound systems*. Kluwer Academic Publishers. https://doi.org/10.1007/978-1-4419-8776-1

Gasser, T., & Westhoff, D. (2012). BASt-study: definitions of automation and legal issues in Germany. Presented at the 2012 Road Vehicle Automation Workshop, Transportation Research Board. Available at http://onlinepubs.trb.org/onlinepubs/conferences/2012/Automation/presentations/Gasser.pdf

Godinho, L., António, J., & Tadeu, A. (2001). 3D sound scattering by rigid barriers in the vicinity of tall buildings. *Applied Acoustics, 62*, 1229–1248. https://doi.org/10.1016/S0003-682X(01)00004-4

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*, 1938–1942. https://doi.org/10.1177/1541931213571433

Gomes, L. (2014). Google self-driving car: It may never actually happen. Available at http://www.slate.com/articles/technology/technology/2014/10/google_self_driving_car_it_may_never_actually_happen.html

Goodall, N. J. (2014). Machine ethics and automated vehicles. In Meyer G, Beiker S, eds. *Road Vehicle Automation*. Springer International Publishing, 93–102. https://doi.org/10.1007/978-3-319-05990-7_9

Gururaj, G. (2008). Road traffic deaths, injuries and disabilities in India: Current scenario. *National Medical Journal of India*, *21*, 14–20.

Heikoop, D., De Winter, J. C. F., Van Arem, B., & Stanton, N. A. (2015). Psychological constructs in driving automation: A consensus model and critical comment on construct proliferation. Manuscript submitted for publication.

Hermann, T., Hunt, A., & Neuhoff J. (2011). *The sonification handbook*. Berlin: Logos Publishing House.

Ho, C., Reed, N., & Spence, C. (2007). Multisensory in-car warning signals for collision avoidance. *Human Factors, 49*, 1107–1114. https://doi.org/10.1518/001872007X249965

Holland, S., Morse, D. R., & Gedenryd, H. (2002). AudioGPS: spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous Computing*, 6, 253–259. https://doi.org/10.1007/s007790200025

Honda. (2014). Forward Collision Warning. Available at http://owners.honda.com/vehicles/information/2014/Accord-Coupe/features/Forward-Collision-Warning/3

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, *14,* 1–4.

Jamson, A. H., Merat, N., Carsten, O., & Lai, F. C. H. (2011). Fully-automated driving: The road to future vehicles. *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Olympic Valley - Lake Tahoe, CA. 2–9.

Jamson, A. H., Lai, F. C. H., & Carsten O. M. J. (2008). Potential benefits of an adaptive forward collision warning system. *Transportation Research Part C: Emerging Technologies, 16*, 471–484. https://doi.org/10.1016/j.trc.2007.09.003

Jonsson, I.-M., & Dahlbäck, N. (2011). I can't hear you? drivers interacting with male or female voices in native or non-native language. In Stephanidis C, ed. *Universal access in human-computer interaction. Context diversity*. Berlin Heidelberg: Springer-Verlag, 298–305.

Kessler, C., Etemad, A., Alessandretti, G., Heinig, K., Selpi, Brouwer, R., Cserpinszky, Hagleitner, W., & Benmimoun, M. (2012). *euroFOT – final report* (Deliverable D11.3). Aachen, Germany. Available at http://www.eurofot-ip.eu/download/library/deliverables/eurofotsp120121212v11dld113_final_report.pdf

Kingsley, K. (2009). Evaluating crash avoidance countermeasures using data from FMCSA/NHTSA's large truck crash causation study. Proceedings of the *21st International Technical Conference on the Enhanced Safety of Vehicles Conference, Stuttgart, Germany.* Available at http://www-nrd.nhtsa.dot.gov/pdf/ESV/esv21/09-0460.pdf

Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY: ACM Press, 453–456.

Kulas, J. T., & Stachowski, A. A. (2009). Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality, 43*, 489–493. https://doi.org/10.1016/j.jrp.2008.12.005

Kulas, J. T., Stachowski, A. A., & Haynes, B. A. (2008). Middle response functioning in likert-responses to personality items. *Journal of Business*

and Psychology, 22, 251–259. https://doi.org/10.1007/s10869-008-9064-2
Kyriakidis, M., Happee, R., & De Winter, J. C. F. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour, 32*, 127–140. https://doi.org/10.1016/j.trf.2015.04.014
Large, D. R., & Burnett, G. E. (2013). Drivers' preferences and emotional responses to satellite navigation voices. *International Journal of Vehicle Noise and Vibration, 9*, 28–46.
Larsson, A. F., Kircher, K., & Hultgren, J. A. (2014). Learning from experience: Familiarity with ACC and responding to a cut-in situation in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 229–237. https://doi.org/10.1016/j.trf.2014.05.008
Lee, A. H., Stevenson, M. R., Wang, K., & Yau, K. K. W. (2002). Modeling young driver motor vehicle crashes: data with extra zeros. *Accident Analysis & Prevention, 34*, 515–521. https://doi.org/10.1016/S0001-4575(01)00049-5
Lee, J. D. (2007). Technology and teen drivers. *Journal of Safety Research, 38*, 203–213. https://doi.org/10.1016/j.jsr.2007.02.008
Levitan, L., Golembiewski, G., & Bloomfield, J. R. (1998). Human factors issues for automated highway systems. *ITS Journal, 4*, 21–47. https://doi.org/10.1080/10248079808903735
Llaneras, R. E., Salinger, J. A., & Green, C. A. (2013). Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention to the forward roadway. Proceedings of *the 7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Bolton Landing, NY*. 92–98.
Lumbreras, M., Sánchez, J., & Barcia, M. (1996). A 3D sound hypermedial system for the blind. Proceedings of *the First European Conference on Disability, Virtual Reality and Associated Technologies*, Maidenhead, UK. 187–191. Available at http://www.icdvrat.org/1996/papers/1996_22.pdf
Markoff, J. (2010). Google cars drive themselves, in traffic. The New York Times. Available at http://www.nytimes.com/2010/10/10/science/10google.html
Mason, W., & Suri S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44*, 1–23. https://doi.org/10.3758/s13428-011-0124-6
Matthews, M. L., & Moran, A. R. (1986). Age differences in male drivers' perception of accident risk: The role of perceived driving ability. *Accident Analysis & Prevention, 18*, 299–313. https://doi.org/10.1016/0001-4575(86)90044-8
Merat, N., Jamson, A. H., Lai, F. C., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors, 54*, 762–771. https://doi.org/10.1177/0018720812442087
Mercedes-Benz. (2014). Pre-Safe. Available at http://www2.mercedes-benz.co.uk/content/unitedkingdom/mpc/mpc_unitedkingdom_website/en/home_mpc/passengercars/home/corporate_sales0/fleet/leasing/our_advantages/Safety.0009.html
Miller, R., & Huang, Q. (2002). An adaptive peer-to-peer collision warning system. Proceedings of *the IEEE 55th Vehicular Technology Conference, 2002, Birmingham, AL*, 317–321.

https://doi.org/10.1109/VTC.2002.1002718
Mynatt, E. D. (1990). Auditory presentation of graphical user interfaces. Available at https://smartech.gatech.edu/bitstream/handle/1853/3681/92-26.pdf
Nixon, C. W., Morris, L. J., McCavitt, A. R., McKinley, R. L., Anderson, T. R., McDaniel, M. P., & Yeager, D. G. (1998a). Female voice communications in high levels of aircraft cockpit noises – part I: spectra, levels, and microphones. *Aviation, Space, and Environmental Medicine, 69*, 675–683.
Nixon, C. W., Anderson, T. R., Morris, L. J., McCavitt, A., McKinley, R. L., Yeager, D. G., & McDaniel, M. P. (1998b). Female voice communications in high level aircraft cockpit noises – part II: vocoder and automatic speech recognition systems. *Aviation, Space, and Environmental Medicine, 69*, 1087–1094
Noyes, J. M., Hellier, E., & Edworthy, J. (2006). Speech warnings: a review. *Theoretical Issues in Ergonomics Science, 7*, 551–571. https://doi.org/10.1080/14639220600731123
Norman, D. A. (1990). The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 327*, 585–593. https://doi.org/10.1098/rstb.1990.0101
Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 252–263. https://doi.org/10.1016/j.trf.2014.04.009
Rozier, J. M. (2000). Hear&There: An augmented reality system of linked audio. MSc Thesis, Cambridge, MA: MIT.
Rudin-Brown, C. M., & Parker, H. A. (2004). Behavioural adaptation to adaptive cruise control (ACC): implications for preventive strategies. *Transportation Research Part F: Traffic Psychology and Behaviour, 7*, 59–76. https://doi.org/10.1016/j.trf.2004.02.001
Scallen, S. F., Hancock, P. A., & Duley, J. A. (1995). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Applied Ergonomics, 26*, 397–403. https://doi.org/10.1016/0003-6870(95)00054-2
Sommer, K. (2013). Mobility study 2013. Available at https://www.conti-online.com/www/pressportal_com_en/themes/initiatives/channel_mobility_study_en/ov_mobility_study2013_en
Srinivasa, N. (2002). Vision-based vehicle detection and tracking method for forward collision warning in automobiles. Proceedings of *IEEE Intelligent Vehicles Symposium, 2002*. 626–631. https://doi.org/10.1109/IVS.2002.1188021
Stanton, N. A., Young, M. S., Walker, G. H., Turner, H., & Randle, S. (2001). Automating the driver's control tasks. *International Journal of Cognitive Ergonomics, 5,* 221–236. https://doi.org/10.1207/S15327566IJCE0503_5
Stanton, N. A., & Edworthy, J. (1999). *Human factors in auditory warnings*. Aldershot, GB: Ashgate.
Stanton, N. A., Young, M. S., & McCaulder, B. (1997). Drive-by-wire: The case of driver workload and reclaiming control with adaptive cruise control. *Safety Science, 27*, 149–159. https://doi.org/10.1016/S0925-7535(97)00054-4

Strand, N., Nilsson, J., Karlsson, I. C. M., & Nilsson, L. (2014). Semi-automated versus highly automated driving in critical situations caused by automation failures. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 218–228. https://doi.org/10.1016/j.trf.2014.04.005

Sturgis, P., Roberts, S. C., & Smith, P. (2012). Middle alternatives revisited: How the neither/nor response acts as a way of saying "I Don't Know"? *Sociological Methods & Research, 43*, 15–38. https://doi.org/10.1177/0049124112452527

Tacken, M., Marcellini, F., Mollenkopf, H., Ruoppila, I., & Széman, Z. (2005). Use and acceptance of new technology by older people. Findings of the international MOBILATE survey: 'Enhancing mobility in later life'. *Gerontechnology, 3,* 126–137. https://doi.org/10.4017/gt.2005.03.03.002.00

Takayama, L., & Nass, C. (2008). Driver safety and information from afar: An experimental driving simulator study of wireless vs. in-car information services. *International Journal of Human Computer Studies, 66,* 173–184. https://doi.org/10.1016/j.ijhcs.2006.06.005

Toyota. (2014). Parking. Available at http://www.toyota-global.com/innovation/safety_technology/safety_technology/parking

UK Visas & Immigration. (2014). Tier 4 of the Points Based System – policy guidance. Available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/370866/T4_Guidance_11-14.pdf

Underwood, S. (2014). Automated vehicles forecast: Vehicle symposium opinion survey. Paper presented at *the Automated Vehicles Symposium*, San Francisco, CA. Available at https://drive.google.com/file/d/0B8gGx-CYkV-wREVMTEhHQUxjOWM

Varaiya, P. (1993). Smart cars on smart roads. Problems of control. *IEEE Transactions on Automatic Control, 38*, 195–207. https://doi.org/10.1109/9.250509

Volkswagen. (2014). Park assist. Available at http://www.volkswagen.co.uk/technology/parking-and-manoeuvring/park-assist

Volvo. (2013a). Volvo Car Group reveals world-class safety and support features that will be introduced in the all-new XC9. Available at https://www.media.volvocars.com/global/en-gb/media/pressreleases/49875/volvo-cars-reveals-world-class-safety-and-support-features-to-be-introduced-in-the-all-new-xc90-in-2

Volvo. (2013b). One million cars with pioneering auto brake technology sold: Volvo Car Group reaches landmark safety milestone*.* Available at https://www.media.volvocars.com/uk/en-gb/media/pressreleases/49817

Weber, M. (2014). Where to? A history of autonomous vehicles. Available at http://www.computerhistory.org/atchm/where-to-a-history-of-autonomous-vehicles

Wetmore, J. (2003). Driving the dream. The history and motivations behind 60 years of automated highway systems in America. *Automotive History Review, 7*, 4–19.

Wickens, C. D., & Hollands, J. G. (2013). *Engineering psychology and human performance*. New York, NY: HarperCollins Publishers.

World Health Organization. (2013). Global status report on road safety 2013,

World Health Organization. Available at http://www.who.int/iris/bitstream/10665/78256/1/9789241564564_eng.pdf?ua=1

Yoshida, J. (2014). Fully autonomous car? Don't buy shotgun yet. Available at *http://www.eetimes.com/author.asp?section_id=36&doc_id=1322246*

2.3 Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays

Corresponding article: Bazilinskyy, P., Petermeijer, S. M., Petrovych, V., Dodou, D., & De Winter, J. C. F. (2018). Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays. Transportation Research Part F: Traffic Psychology and Behaviour, 56, 82–98.

2.3.1 Abstract

An important research question in the domain of highly automated driving is how to aid drivers in transitions between manual and automated control. Until highly automated cars are available, knowledge on this topic has to be obtained via simulators and self-report questionnaires. Using crowdsourcing, we surveyed 1,692 people on auditory, visual, and vibrotactile take-over requests (TORs) in highly automated driving. The survey presented recordings of auditory messages and illustrations of visual and vibrational messages in traffic scenarios of various urgency levels. Multimodal TORs were the most preferred option in high-urgency scenarios. Auditory TORs were the most preferred option in low-urgency scenarios and as a confirmation message that the system is ready to switch from manual to automated mode. For low-urgency scenarios, visual-only TORs were more preferred than vibration-only TORs. Beeps with shorter interpulse intervals were perceived as more urgent, with Stevens' power law yielding an accurate fit to the data. Spoken messages were more accepted than abstract sounds, and the female voice was more preferred than the male voice. Preferences and perceived urgency ratings were similar in middle- and high-income countries. In summary, this international survey showed that people's preferences for TOR types in highly automated driving depend on the urgency of the situation.

2.3.2 Introduction

2.3.2.1 Highly automated driving and take-over requests

Now that partially automated driving systems are in serial production, it is foreseen that highly automated driving will be deployed on public roads in the next one or two decades (see Begg, 2014; Underwood, 2014 for predictive surveys). Highly automated driving, a term introduced in 2006 (Flemisch et al., 2006), is defined as a technology that takes over both longitudinal and lateral control, and in which the driver is no longer required to permanently monitor the machine (Gasser & Westhoff, 2012). In highly automated driving, the driver is permitted to take hands and feet off the steering wheel and pedals and may engage in non-driving tasks such as checking the phone, reading a book, or

resting. When the automation reaches its operational limit in a given traffic situation, the automation issues a so-called take-over request (TOR), asking the driver to take back control of the vehicle (Gasser & Westhoff, 2012; Hoeger et al., 2011).

The level of urgency of the situation with which the automation cannot cope is a critical parameter of the take-over process. Several scenarios are imaginable, such as (in order of increasing urgency):

1. The automated vehicle is reaching a target highway exit.
2. The automated vehicle has to make a lane change because it is approaching a slow-moving vehicle.
3. There are construction works on the road ahead, and at least one of the lanes is closed.
4. A technical failure prevents the automated vehicle from working properly.
5. An accident has just happened right in front of the automated vehicle.

In the first scenario, the automation can issue the TOR long in advance, so that the driver has ample time to resume manual control, whereas in the last scenario the traffic situation has changed abruptly, leaving the driver with little time for taking over. Recent studies (Gold, Damböck, Lorenz, & Bengler, 2013; Gold, Lorenz, Damböck, & Bengler, 2013; Mok et al., 2015; You et al., 2017) have quantified the effect of the urgency of the take-over (sometimes called “time budget”; Gold, Lorenz et al., 2013, “time buffer”; Gasser & Westhoff, 2012, or “lead time”; Society of Automotive Engineers [SAE], 2014) on the driver’s response time and on the quality of the take-over. For example, Gold, Damböck, et al. (2013) found that for shorter time budgets after an audio-visual TOR, drivers responded faster, but the take-over was of lower quality (hard braking, swerving, and inappropriate full stops). Choosing the right display for providing a TOR is important, especially in urgent scenarios, where a delay of a few tenths of a second in brake reaction could mean the difference between colliding and not colliding. Note that in this article, the term ‘display’ does not necessarily refer to a visual instrument, but to “any instrument or device that presents information to any sense organ (visual, auditory, or other)” (Swain & Guttmann, 1983, p. 2–3).

2.3.2.2 Displays for take-over requests in highly automated driving

2.3.2.2.1 Visual displays

Manual driving is primarily a visual task (Green, 1999; Sivak & Owens, 1996). Traditionally, visual information about the vehicle state (e.g., speed, RPM) as well as warnings (e.g., low fuel, high engine temperature) are provided on the dashboard. It is well established that the appropriate use of colour, saliency, and spatial positioning according to the principles of moving part and proximity compatibility can make a visual display easy to understand (e.g., Fitts & Jones, 1947; Grether, 1949). Accordingly, visual TORs during highly automated driving could be issued by lighting up an icon/region on the dashboard (e.g., Flemisch, Kaussner, Petermann, Schieben, & Schöming, 2011; Naujoks, Mai, & Neukum, 2014) or, more innovatively, on a head-up display (Kim et al., 2017; Langlois & Soualmi, 2016; see Manca, De Winter, & Happee, 2015, for an overview).

While visual information presented on the dashboard during manual driving may be detectable using peripheral vision (Lamble, Laakso, & Summala, 1999), such information might be missed if the driver is engaged in a non-driving task. A promising approach for conveying visual messages is to use ambient TORs, for example by lighting up a LED strip or bar under the windshield, along the A-pillar, or around the cabin (e.g., Kelsch & Dziennus, 2015; Löcken, Heuten, & Boll, 2015; Meschtscherjakov, Döttlinger, Rödel, & Tscheligi, 2015; Pfromm, Cieler, & Bruder, 2013; Winkler et al., 2018), or by using spatially directed LED strips (Dettman & Bullinger, 2017). Such ambient TORs might be more easily detected by the distracted driver than localized visual cues.

2.3.2.2.2 Auditory displays

Auditory displays have the advantage of being "gaze-free" (Meng & Spence, 2015; Stokes, Wickens, & Kite, 1990), meaning that the stimulus does not have to be in the field of view of the driver in order to be detected (for an overview of the advantages of auditory over visual displays, see Sanders & McCormick, 1987, as summarized by Stanton & Edworthy, 1999). Furthermore, it has been argued that the distinctive meaning of multiple auditory warnings can be easily learned (Blattner, Sumikawa, & Greenberg, 1989; Gaver, 1986). For example, Patterson (1982; see also Patterson & Mayfield, 1990) showed that listeners were able to learn between four and six auditory warnings in only a few practice trials.

Auditory displays are widely used in consumer cars for warning about hazards in the outside environment, such as approaching objects during parking (BMW, 2013; Toyota, 2014; Volkswagen, 2014) or vehicles on a collision path (Graham, 1999; Honda, 2014; Ramkissoon, 2001; for reviews, see Nees & Walker, 2011; Wickens & Seppelt, 2002). A TOR in a highly automated vehicle could be conveyed with similar abstract (i.e., non-verbal) warning sounds, such as beeps and tones, whereas voice messages could be used to transfer higher levels of semantics (Naujoks et al., 2016; Politis, Brewster, & Pollick, 2015a). In complex tasks, abstract auditory warnings may give faster initial responses than voice messages, because the duration of the latter is longer and humans tend to wait until the completion of the entire voice message before taking meaningful action (e.g., Bate, 1969).

Whether voice gender is important in auditory warnings has been a subject of discussion for several decades. A previous online survey showed that the female voice was strongly favoured over the male voice when used as a TOR (Bazilinskyy & De Winter, 2015), and a questionnaire on satellite navigation devices showed that a female voice is more often used for providing directions than a male voice (Large & Burnett, 2013). However, the results of studies on perceived urgency of a male versus a female voice are inconclusive. Park and Jang (1999) found that a male voice was perceived as more urgent than a female voice, whereas others reported the opposite effect (e.g., Hellier, Edworthy, Weedon, Walters, & Adams, see Edworthy & Hellier, 2003 for a review). Jang (2007) tested male and female voices of equal fundamental frequency and loudness and found that the male voice was perceived as more urgent. Jang argued that voice characteristics such as smoothness and timbre may explain the differences in perceived urgency between male and female voices. In a study investigating verbal cockpit warnings, Arrabito (2009) found that a male voice (either monotone or urgent) led to a larger number of correct responses and faster response times than a female voice. Nixon et al. (1998) assessed the

intelligibility of the male and female voice in military aircraft cockpits with noise levels ranging between 95 dB and 115 dB and found that the female voice tended to be less intelligible than the male voice, especially at the highest noise level (Nixon et al., 1998; see Noyes, Hellier, & Edworthy, 2006 for a review on speech warnings).

Looming sounds, which are sounds that increase in intensity with the criticality of the situation, are another type of auditory displays that might be useful in automated driving. Studies measuring brain activity with event-related functional magnetic resonance imaging (fMRI) have indicated that a sound that increases in intensity is associated with activation of the right amygdala and left temporal regions, suggesting that looming sounds are an intrinsic warning cue (Bach et al., 2008). Moreover, according to the principle of 'auditory tau' (Shaw, McGowan, & Turvey, 1991), it is theoretically possible to estimate time-to-contact based on the looming intensity of an acoustic signal (Gray, 2011; Silva et al., 2017). In a driving simulator study, Gray reported that a looming auditory warning led to brake reaction times that were 130 ms faster than the brake reaction times for an auditory warning with constant intensity.

2.3.2.2.3 Vibrotactile displays

Compared to visual and auditory displays, tactile displays are underused in the automotive domain (Jones & Sarter, 2008), but the interest in employing vibrations to convey information to the driver is growing rapidly (e.g., Birrell, Young, & Weldon, 2013; De Groot, De Winter, García, Mulder, & Wieringa, 2011; Grah et al., 2015; Meng, Ho, Gray, & Spence, 2015; for a review, see Petermeijer, De Winter, & Bengler, 2016). A specific advantage of vibrotactile displays is that they can provide information in a private manner (Petermeijer et al., 2016), whereas a disadvantage is that they can capture only a limited amount of information compared to auditory displays (Lu, Wickens, Sarter & Sebok, 2011) and may not be suitable for issuing multiple alerts. For example, Fitch, Hankey, Kleiner, and Dingus (2011) showed that while drivers could respond well to a display presenting three vibrational messages (conveyed by different combinations of pairs of four tactors at the seat pan), a seat pan display with seven distinctive vibrational alerts led to erroneous and delayed responses by the driver.

For the driver to perceive vibrotactile warnings, the tactors and the human have to be in physical contact with each other. Recently, BMW and Citroën, among other car manufacturers, introduced a lane departure warning system that alerts the driver by vibrating the steering wheel (BMW, 2013; Spence & Ho, 2008). In highly automated mode, in which the driver is likely to have his hands off the steering wheel, alternative locations to provide vibrations to the driver should be considered, such as the seat back, seat pan, or seat belt (e.g., Schwalk, Kalogerakis, & Maier, 2015; Telpaz, Rhindress, Zelman, & Tsimhoni, 2015).

2.3.2.2.4 Multimodal displays

Auditory, visual, and vibrotactile displays can be combined into a multimodal display (e.g., Lee & Spence, 2008; Liu, 2001; Oviatt, 2003; Reeves et al., 2004). Multimodal displays can output more information per quantum of time, resulting in better task performance, compared to unimodal displays (Selcon, Taylor, & McKenna, 1995; Wickens, Hollands, Banbury, & Parasuraman, 2012; and see

Burke et al., 2006, for a meta-analysis). For example, De Groot et al. (2013) showed that audio-visual displays were more effective than auditory-only displays for giving turn left/right instructions. A meta-analysis by Prewett, Elliott, Walvoord, and Coovert (2012) showed that task effectiveness (defined as an aggregate of error rate, task completion time, and reaction time effect sizes) was higher for visual-vibrotactile displays than for visual-only displays.

Multimodal displays incorporating all three modalities (i.e., visual, auditory, and vibrotactile) are rare in car driving (Petermeijer et al., 2016). In a series of driving simulator studies, Politis, Brewster, and Pollick (2014, 2015a, 2015b) investigated the effectiveness of tri-modal (visual, auditory, and vibrotactile) displays on driver behaviour and found that these led to increased perceived urgency and perceived alerting effectiveness as compared to unimodal displays. On the other hand, in a study investigating warnings for TORs in Adaptive Cruise Control (ACC), Lee, McGehee, Brown and Marshall (2006) found that a combination of a visual warning, an auditory warning, a vibratory seat, and pulsation at the brake pedal led to 400 ms slower reaction times than a visual-auditory warning.

The difference in the effects of multimodal warnings in past research can be explained by the fact that the efficacy of a multimodal display depends on whether or not they are semantically, temporally, and spatially congruent, leading to redundancy (Diaconescu, Alain,& McIntosh, 2011; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). It should also be noted that the benefit of multimodal over unimodal displays is not necessarily due to sensory integration (i.e., the beneficial interaction of redundant signals) but could also be caused by one of the modalities substituting another modality that is unavailable or overloaded. As Hancock et al. (2015) noted: “multisensory audio-tactile cuing may be superior to audio cuing alone in noisy environments, but this effect is not necessarily contingent on multisensory integration of auditory with tactile cues, per se” (p. 7).

2.3.2.3 Perceived urgency as a function of the display characteristics and the operator’s characteristics

Extensive work has been conducted on the relationship between display characteristics and evoked perceived urgency. As Hellier, Edworthy, and Dennis (1993), Hellier and Edworthy (1999), Park and Jang (1999), and Baldwin et al. (2012) showed, Stevens’ (1957) power law, which associates the objective magnitude of a physical stimulus with its perceived intensity, can be used to predict the perceived urgency of a warning. It has been found that perceived urgency increases with smaller interpulse intervals and that this effect holds for each of the modalities (visual, auditory, & vibrotactile) and combinations thereof (Haas & Casali, 1995; Haas & Edworthy, 1999; Van Erp, Toet, & Janssen, 2015). In Van Erp, Toet, and Janssen, who measured perceived urgency on a scale from 1 (*not urgent*) to 7 (*very urgent*) for 25 combinations of pulse and interpulse interval durations (both ranging between 100 and 1600 ms), perceived urgency increased with decreasing interpulse interval (e.g., for a pulse duration of 100 ms, perceived urgency increased from 2.3 for interpulse intervals of 1600 ms to 6.0 for interpulse intervals of 100 ms). Moreover, for the same interpulse interval, the highest perceived urgency was reported for pulses with a duration equal to the duration of the interpulse interval (i.e., symmetric pulse profiles). In verbal

warnings, the semantics of the signal words is an additional parameter that affects perceived urgency, with words such as “deadly” and “danger” being perceived as more urgent than “warning”, “caution”, or “note” (Hellier et al., 2002).

2.3.2.4 Aim of the paper

The aim of the present study was to investigate which unimodal or multimodal display is the most preferred for issuing a TOR during highly automated driving in traffic scenarios of various levels of urgency. Although preference may not coincide with the effectiveness of a display (e.g., Scott & Gray, 2008), the former is important in designing automated driving systems, as dissatisfaction with the display may lead to disuse (Parasuraman & Riley, 1997). It was hypothesized that in low-urgency scenarios drivers opt for unimodal TORs via the traditional visual and/or auditory sensory channels, whereas in high-urgency scenarios a combination of auditory, visual, and vibrotactile TORs is preferred. Additionally, we expected that the shorter the duration of looming sounds and the shorter the intervals between beeps, the more urgent the warning is perceived to be.

We also explored whether TOR preferences correlate with trust in automation and self-reported driving violations. Trust in automation is an important predictor variable in the development of highly automated driving systems, as low levels of trust may lead to disuse of automation, whereas high levels of trust may lead to misuse of automation and compliance (Parasuraman & Riley, 1997). Self-reported driving violations have been included as they are an important predictor of on-road driving speed (De Winter, Dodou, & Stanton, 2015), which in turn determines the criticality of driving situations and may therefore interact with TOR preferences.

Additionally, we aimed to gain additional insight into the results of Bazilinskyy and De Winter (2015), in which the female voice was preferred over the male voice for supporting TORs during highly automated driving. Bazilinskyy and De Winter included a textual question on this matter. In the current survey, we also produced a synthesized male and female voice and asked people to rate these voices.

Finally, we investigated whether preferences are consistent between countries with different income levels, in light of the view that income is likely to be a strong predictor of road safety (World Health Organization, 2015), and because previous cross-national questionnaire research suggests that respondents from higher-income countries are more likely to be critical towards aspects of automated driving (Bazilinskyy, Kyriakidis, & De Winter, 2015; Kyriakidis et al., 2015).

Nowadays, by means of the Internet, researchers can cost-effectively reach a large and diverse pool of participants (Gosling, Vazire, Srivastava, & John, 2004). Moreover, Internet and crowdsourcing platforms allow for the creation of media-rich surveys, containing audio snippets and videos. Media-rich surveys have been shown to be useful for investigating driving attitudes and opinions. For example, Eriksson, Solis Marcos, Kircher, Västfjäll, and Stanton (2015) used an online questionnaire with still images to investigate the type of information people would like to receive during a TOR as a function of the available time and the complexity of the traffic situation. The present survey was conducted with CrowdFlower, which is a platform that allows academic

researchers to access an online community to complete a dedicated task, such as filling out a survey (Kyriakidis et al., 2015), categorizing data, or annotating images or videos (Cabrall et al., 2016). Researchers ("Customers" in CrowdFlower terminology) upload the tasks, which are then completed by respondents ("Contributors") in return for a small monetary reward.

2.3.3 Methods

2.3.3.1 Survey

A survey consisting of 67 questions was developed with CrowdFlower (www.crowdflower.com). Table S1 in the supplementary material shows the questions of the survey as well as the corresponding response options. Earlier surveys indicated that people who work on CrowdFlower surveys have mostly an undergraduate degree (Bazilinskyy & De Winter, 2015; Kyriakidis, Happee, & De Winter, 2015; see also Behrend, Sharek, Meade, & Wiebe, 2011, for similar conclusions regarding Amazon Mechanical Turk). Therefore we used everyday language and avoided technical terms and definitions. A payment of $0.20 (USD) was offered to each respondent for completing the survey. Information about payment appears next to each task when a contributor browses through the list of available tasks. The research was approved by the TU Delft Human Research Ethics Committee.

2.3.3.2 Survey structure

At the beginning of the survey, contact information of the researchers was provided, and the purpose of the survey was described as "to explore the public opinion on the use of sound, vibration, and visual interfaces during highly automated car driving". Respondents were informed that the survey would take approximately 15 minutes of their time. Highly automated driving was then defined as "The automated driving car controls both speed and steering. The driver is not required to look at the road. If the automation cannot handle a situation, it provides a take-over request, and the driver must take over control", The respondents were informed that they could contact the investigators to ask questions about the study and that they had to be at least 18 years old to participate. Information about anonymity and voluntary participation was also given.

The questions were divided into six parts. First, general questions on the respondent's age, gender, age at which the respondent obtained their first licence for driving a car or motorcycle, and driving habits and behaviour were posed (Q2–Q15). The questions about driving style (Q9–Q15) were based on the violations scale of the Driver Behaviour Questionnaire (DBQ) as presented by De Winter (2013) and used in previous CrowdFlower surveys (Bazilinskyy & De Winter, 2015; De Winter, Kyriakidis, Dodou, & Happee, 2015).

The second part of the survey (Q16–Q24) focused on general trust in automation, of which Q19–Q24 were previously used by Merritt, Heimbaugh, LaChapell, and Lee (2012) and De Winter and Hancock (2015).

The third part (Q25–Q46) consisted of questions on auditory TORs. Auditory examples were presented in the form of MP3 files stored on an external server. Before answering the questions, the respondents were asked to click on links directing to the recordings and listen to the sound samples. Questions

checking whether the respondent actually listened to the samples were incorporated, where the respondents had to select which sound they had just listened to (Q25–Q29). The following auditory TORs were provided:

(1) female voice “Please take over!”;
(2) male voice “Please take over!”. The female and male voices saying “Please take over!” were created with the ‘Free online voice generator’ (http://onlinetonegenerator.com/voice-generator.html); *Google UK English Female* and *Google UK English Male* voices were used;
(3) four pairs of beeps with long (2 s) interpulse intervals, 6 pairs of beeps with medium (1 s) interpulse intervals, 8 pairs of beeps with short (750 ms) interpulse intervals, and 11 pairs of beeps with very short (430 ms) interpulse intervals (each pair consisted of two 240-ms long beep tones separated by a 100-ms interpulse interval, with a frequency of 1,840 Hz);
(4) bell sound;
(5) horn sound; and
(6) looming sounds with a duration of 1 s, 2.5 s, and 5 s. The digital volume of the three looming sounds increased from 0 at the beginning to 0.8 at the end of the sample, with the volume defined on a scale from 0 to 1. Volume increased linearly for the 1 s and 2.5 s samples and quadratically for the 5 s sample. All sound samples were pre-recorded and presented without context (e.g., the increase in volume was a function of time, and was not based on time-to-contact or any other parameter). The frequency of the looming sounds was 440 Hz. The looming sounds were generated with Audacity software.

Additionally, the respondents were asked on a five-point Likert scale from *Disagree strongly* to *Agree strongly* whether a TOR should be provided by means of the female voice (Q30), the male voice (Q31), the beeps (Q32), the bell sound (Q33), and the horn sound (Q34). Next, they were asked which of these five sound messages they considered as the most urgent (Q35) and the most annoying (Q36). Questions Q37 (multiple-choice question) and Q38 (free-response question) asked the respondents to indicate their opinion on why the female voice is often seen as the most preferred type of auditory message to be used in cars. In questions Q39–Q45, the respondents were asked on a five-point Likert scale from *Disagree strongly* to *Agree strongly* whether they considered each of the provided beep sounds and looming sounds as urgent, and in question Q46 the respondents’ opinion was polled on whether sounds are a good way to get their attention back to the road, again using a scale from *Disagree strongly* to *Agree strongly*.

The fourth part of the survey (Q47–Q52) focused on visual TORs. Respondents were presented with illustrations of (1) a green icon on the dashboard, (2) a strip of lights at the bottom of the windshield, (3) a head-up display with a green icon, and (4) a brighter dashboard (Fig. 1), and were asked on a five-point Likert scale from *Disagree strongly* to *Agree strongly* whether a take-over request should be provided by means of each of these four visual displays (Q47–Q50). Next, they were asked whether they would like the automation to take over control when they were not looking at the road for over 5 seconds (Q51), and whether visual messages are a good way to get their

attention back to the road (Q52), both questions on a scale from *Disagree strongly* to *Agree strongly*.

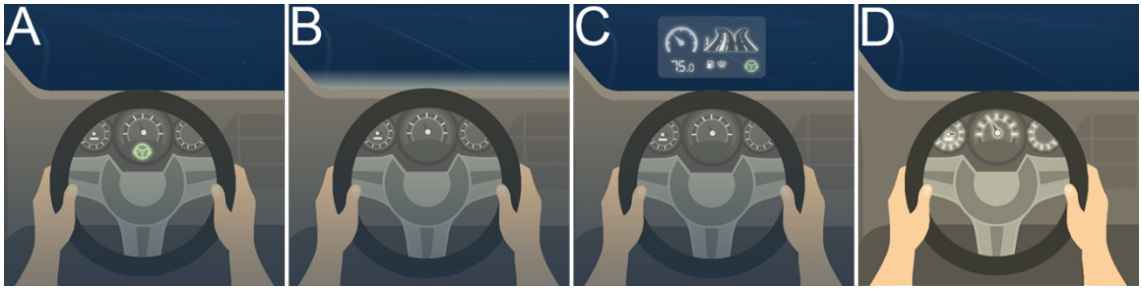


Figure 1. Illustrations for visual take-over requests. (A) A green icon on the dashboard (Q47); (B) A strip of lights at the bottom of the windshield (Q48); (C) A head-up display with a green icon (Q49); (D) A brighter dashboard (Q50).

The fifth part of the survey (Q53–Q57) posed questions on vibrotactile TORs. Figure 2 shows the illustrations that were provided to the respondents to indicate four proposed locations of vibrations, namely: (1) the seat back, (2) the seat pan, (3) the seat belt, and (4) the steering wheel. The respondents were again asked on a five-point Likert scale from *Disagree strongly* to *Agree strongly* whether a take-over request should be provided by means of each of these four vibrotactile displays (Q53–Q56). In Question Q57 the respondents' were asked whether vibrations are a good way to get their attention back to the road on a scale from *Disagree strongly* to *Agree strongly*.

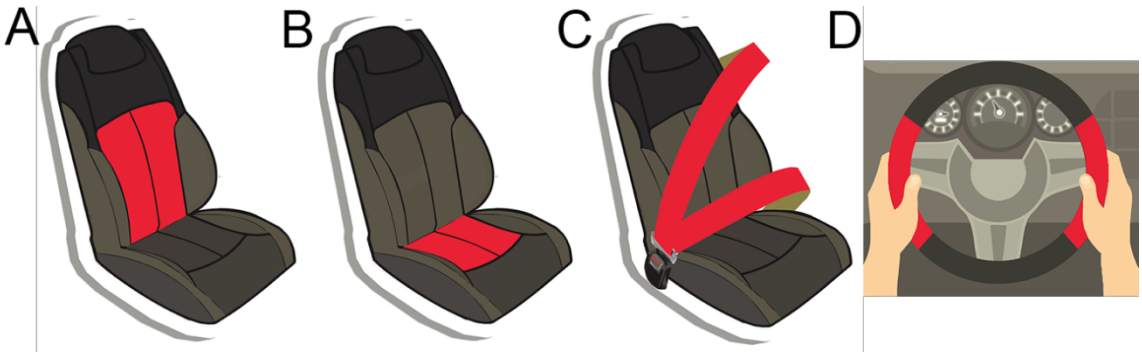


Figure 2. Illustrations for vibrotactile take-over requests. (A) Vibrations in the seat back (Q53); (2) Vibrations in the seat pan (Q54); (C) Vibrations in the seat belt (Q55); (D) Vibrations in the steering wheel (Q56).

The last part of the survey (Q58–Q65) presented five take-over scenarios of various levels of urgency: (1) construction works, (2) exit highway, (3) changing lanes, (4) automation failure, and (5) traffic accident ahead. For each scenario, the respondents were asked with which display they would like to receive a TOR as well as get informed that the automation was ready to take back control again (Response options: sound message; vibrations; visual message; sound message and vibrations (in any order); sound message and visual message (in any order); visual message and vibrations (in any order); sound message, visual message, and vibrations (in any order)). The scenarios were supplemented with illustrations (Fig. 3). Finally, to poll the public view on the future of transportation, the respondents were asked in which year they think that most cars will be able to drive fully automatically in their country of residence (Q66).

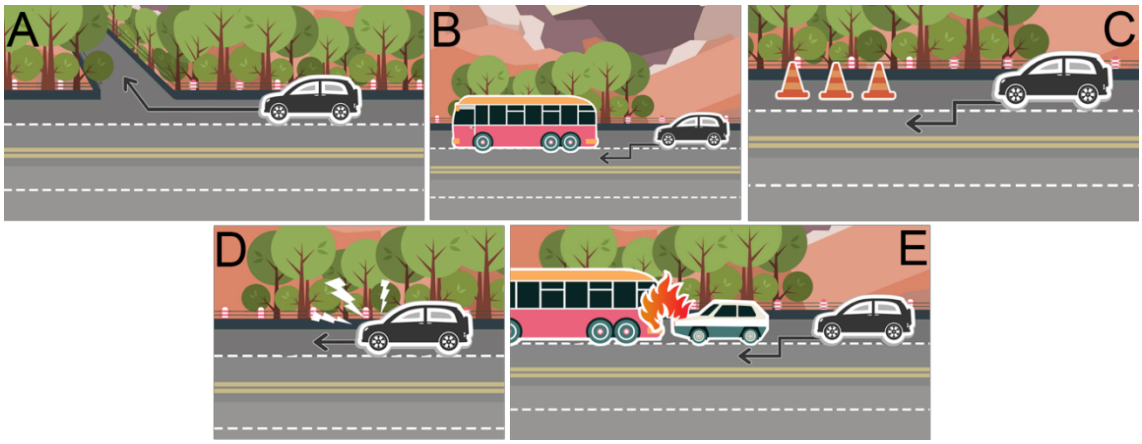


Figure 3. Illustrations for take-over scenarios, presented in order of increasing urgency. (A) ExitHighway (Q60); (B) ChangeLanes (Q61, Q62); (C) ConstWorks (Q58, Q59); (D) Failure (Q63); (E) Accident (Q64, Q65).

The respondents had to complete all questions, but each question had an "I prefer not to respond" option. The last question (Q67) was the only optional one, asking respondents to "provide any suggestions, which could help engineers to build safe and enjoyable automated cars".

2.3.3.3 Configuration of CrowdFlower

In CrowdFlower, the researcher can specify in which countries the target contributors reside. We allowed contributors from all countries to participate in the survey. CrowdFlower contributors are further classified in various performance categories, depending on how trustworthy their former contributions have been. The 'Highest speed–Level 1' contributors were selected. This level of experience accounts for 60% of completed work on CrowdFlower (CrowdFlower, 2015), allowing to reach a diverse group of users of the platform. Completing the survey more than once from the same IP address was not permitted. A payment of $0.20 (USD) was offered to each respondent for completing the survey. We collected 3,000 surveys, at a total cost of $798.00. The study was preceded by a pilot test with 10 respondents. These respondents were not included in the analysis. The pilot test did not lead to any changes in the survey.

2.3.3.4 Statistical analysis

Descriptive statistics (i.e., mean, median, standard deviation, skewness, and numbers of responses) were calculated for each of the variables. The skewness was calculated as the third central moment divided by the cube of the standard deviation. A Spearman correlation matrix of selected predictor variables (gender, age, driving and accident history, mean of the DBQ violations items Q9–Q15, a trust score calculated as the average of *z*-transformed responses to Q19–Q24, year in which the respondent thought that most cars would be able to drive fully automatically in his/her country of residence (Q66), and survey completion time) and outcome variables (all ordinal variables related to auditory, visual, and vibrotactile messages) was created at the level of respondents. A distinction was made between respondents located in middle-income countries (gross national income [GNI] per capita: $1,026–$12,475) and respondents located in high-income countries (GNI per capita: $12,476 or higher; The World Bank, 2016). We

did not perform statistical tests for each possible comparison because the sample size was large, and our concern rests more with practical significance than with statistical significance. As an indication, at our sample size of 1,692, a correlation coefficient of 0.05 is statistically significant from zero (p = 0.04). For all questions regarding the use of male and female voice, differences between male and female respondents were investigated using Fisher's exact test. 95% confidence intervals around the means per item were calculated assuming a normal distribution, whereas 95% confidence intervals of proportions of respondents were calculated assuming a binomial distribution. At the national level, correlations were calculated between GNI per capita in 2013 (The World Bank, 2015) and the outcome variables, only taking into account those countries with 25 or more respondents (see also Kyriakidis et al., 2015).

2.3.4 Results

2.3.4.1 Number of respondents and respondent satisfaction

In total, 3,000 respondents located in 102 countries were surveyed in this study. The responses were collected between 31 March 2015, 14:01 and 1 April 2015, 4:49 (GMT). The survey received an overall satisfaction rating of 4.4 on a scale from 1 (very dissatisfied) to 5 (very satisfied). The respondents took on average 16.9 min to complete the survey (SD = 11.1 min, median = 13.8 min).

2.3.4.2 Data filtering

Respondents who indicated they had not read the instructions (N = 18) and respondents who indicated they were under 18 and thereby did not adhere to the survey instructions (N = 5) were excluded. Additionally, respondents who selected incorrect auditory sounds in Q25–Q29 were excluded (N_{Q25} = 204, N_{Q26} = 281, N_{Q27} = 513, N_{Q28} = 597, N_{Q29} = 508). We also excluded 25 respondents whose country of residence was not identified by CrowdFlower. Finally, we excluded respondents who chose the "I prefer not to respond" or "I did not hear the recording clearly" options in one or more of the multiple-choice questions, as well as participants who did not report their age or age of obtaining a driving license (N = 711; which implies that retained participants all reported to have a driving license). In total, 1,308 surveys were removed, leaving 1,692 completed surveys for further analysis. For the question "In which year do you think that most cars will be able to drive fully automatically in your country of residence?", 35 of the 1,692 answers contained years before 2014 and were excluded from the analysis. Descriptive statistics for all variables are shown in Table S2 of the supplementary material. Table S3 shows histogram counts for all variables.

Note that by the listwise deletion of respondents with missing data (i.e., "I prefer not to respond" or "I did not hear the recording clearly" in one or more questions), the sample size across items becomes constant, and the analysis and presentation of our results are simplified at the expense of some statistical power. The mean responses to the questions were hardly affected by including or excluding these respondents. Specifically, the mean absolute difference of the mean responses across all 41 five-point ordinal items between the used sample (N = 1,692) and the sample with missing data (Ns for the 41 items between 1,991 and 2,026) was 0.02.

2.3.4.3 Sample characteristics

The mean age of the sample was 32.9 (*SD* = 10.4), and the mean age of obtaining a license for a car or motorcycle was 20.1 years (*SD* = 3.8; *N* = 1,686, after excluding 6 participants who reported an unrealistic licensing age below 14 years). Of the 1,692 respondents, 1,220 respondents were male and 472 were female. 1,127 respondents reported that a private vehicle was their primary transportation mode. 632 respondents reported driving every day, and 469 respondents reported driving 4 to 6 days a week. 444 respondents reported that they had driven between 5,001 and 15,000 km in the last 12 months, and 329 respondents reported that they had driven between 1,001 and 5,000 km. The respondents were located in 91 countries. The countries with the largest number of respondents were India (*n* = 169), Spain (*n* = 83), the United States (*n* = 82), Venezuela (*n* = 75), Canada (*n* = 71), and Italy (*n* = 68). Of the 1,692 respondents, 565 were located in middle-income countries (GNI per capita: $1,026–$12,475), and 1,127 were located in high-income countries (GNI per capita: $ 12,476 or higher; The World Bank, 2016). In the middle-income countries, 80% of the respondents were male, whereas the corresponding value was 68% for the high-income countries. In addition to this country-gender association, there was also a strong positive correlation between the mean age of the respondents and the GNI (Spearman ρ = 0.75 across the 23 countries with 25 or more respondents).

2.3.4.4 Respondents' opinion on auditory, visual, and vibrotactile TORs in highly automated driving

The respondents provided their opinion on whether auditory, visual, and vibrotactile messages are a good way to get their attention back to the road (Q46, Q52, and Q57, respectively), all questions on a five-point Likert scale from *Disagree strongly* to *Agree strongly*. The respondents appeared to be equally positive for all three modalities, with means of 4.06, 3.95, and 4.07 for auditory, visual, and vibrotactile messages, respectively.

Furthermore, the respondents indicated what type of messages should be used for TORs (Q30–Q34 for auditory messages, Q47–Q50 for visual messages, and Q53– Q56 for vibrotactile messages), all questions on a five-point Likert scale from *Disagree strongly* to *Agree strongly*. The results in Figure 4 show that among visual displays, the head-up display with a green icon received the highest ratings (*M* = 3.76), followed by the green icon on the dashboard (*M* = 3.69). The respondents were less favourable towards the brighter dashboard (*M* = 3.06) and the strip of lights at the bottom of the windshield (*M* = 2.84). The respondents expressed relatively similar acceptance of all four types of vibrotactile messages presented in the survey: vibration of the steering wheel (*M* = 3.66), vibration of the seat back (*M* = 3.34), vibration of the driver's seat pan (*M* = 3.19), and vibration of the seat belt (*M* = 3.13). Large differences were observed among the provided auditory messages, with spoken messages being more accepted than abstract sounds. Specifically, the female and male voices were the most preferred auditory options (*M* = 3.60 and *M* = 3.34, respectively), whereas a horn sound was the least preferred auditory option (*M* = 2.27).

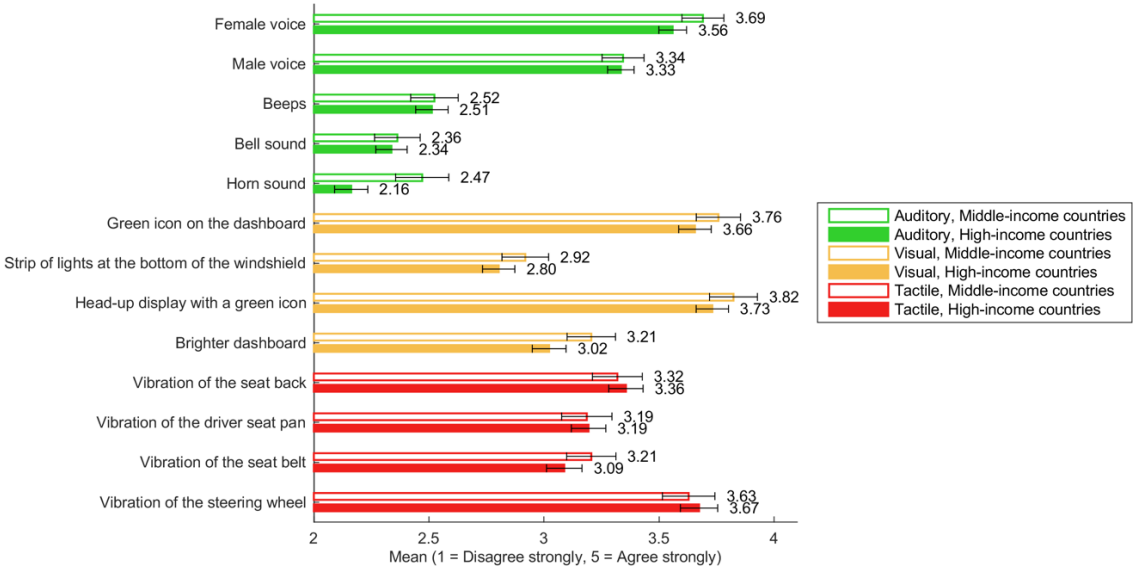


Figure 4. Respondents' agreement regarding whether a take-over request should be provided by means of specific auditory, visual, or vibrotactile messages (Q30–Q34; Q47–Q50; Q53–Q56). A distinction is made between respondents from middle- and high-income countries (n = 565 and 1,127, respectively). The number next to each bar is the mean on the scale from 1 to 5. Error bars denote 95% confidence intervals.

In Q35, the respondents were asked to select which of the proposed auditory messages they considered the most urgent, whereas in Q36 they were asked to report which auditory message they considered the most annoying. Figure 5 shows the results for both questions. The male voice was considered the most urgent auditory message (N = 518; 31% of the male respondents and 30% of the female respondents, p = 1), followed by the female voice (N = 495; 29% of the male respondents and 29% of the female respondents, p = 1) and the horn sound (N = 435). The horn sound was considered the most annoying of the auditory messages (N = 893).

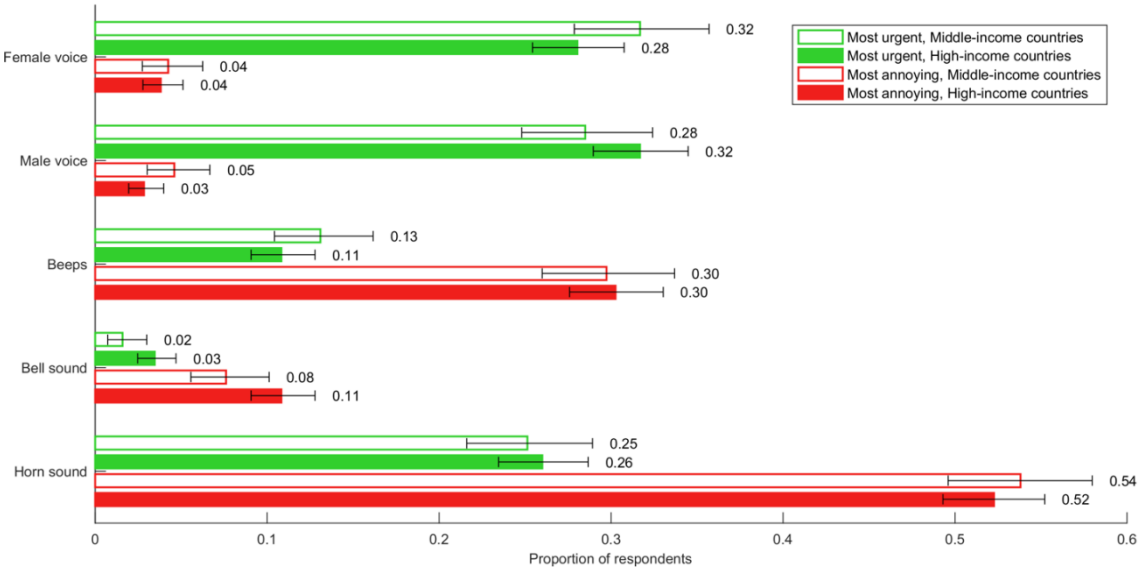


Figure 5. Proportion of respondents who selected a particular auditory message as the most urgent (Q35) and the most annoying (Q36) after listening to these auditory messages. A distinction is made between respondents from middle- and high-income countries (n = 565 and 1,127, respectively). Error bars denote 95% confidence intervals.

In Q39–Q42 the respondents were asked to rate on a five-point Likert scale from *Disagree strongly* to *Agree strongly* whether they considered each of the provided beeps as urgent. A monotonic relationship between interpulse interval and perceived urgency was observed, with the beeps with long interpulse intervals seen as the least urgent (Q39; $M = 2.28$) and the beeps with very short interpulse intervals as the most urgent (Q42; $M = 4.20$). Figure 6 shows the mean perceived urgency rate as a function of the beep rate for the four provided beep messages. A strong linear relationship is observed in logarithmic space ($r = .999$; exponent = 0.66). Among the looming sounds, the sound of 1 s duration was considered the least urgent (Q43; $M = 2.52$), followed by the looming sound of 2.5 s (Q44; $M = 2.92$), and 5 s (Q45; $M = 3.17$).

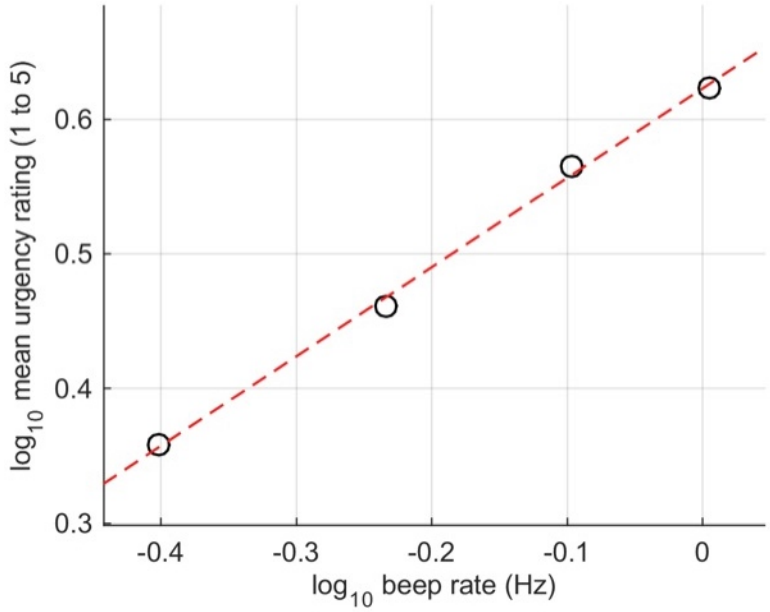


Figure 6. Mean perceived urgency as a function of the beep rate for the four provided beep messages (Q39–Q42).

The survey included a question (Q37) asking respondents to indicate their opinion on why a female voice is often seen as the most preferred type of auditory message to be used in cars. Most participants indicated that a female voice is more pleasant (51%; 854 out of 1,686)—significantly more male respondents (55%) than female respondents (40%; $p < 0.001$), followed by less commanding and easier to agree with ($N = 354$)—significantly more female respondents (29%) than male respondents (18%; $p < 0.001$). A female voice was seen as preferred in noisy environments by 285 respondents (16% of the male respondents vs. 18% of the female respondents; $p = 0.278$), and for driving in a car with males by 160 respondents (9% of the male respondents vs. 10% of the female respondents; $p = 0.645$). The respondents were given the opportunity to provide a textual comment, if they chose the option “Other”. One hundred fifty-two comments were collected; from these, 79 carried no meaningful information/were unclear, and 3 were not written in English. Nine respondents mentioned that the female voice had acoustic advantages over the male voice (as in being clearer, having higher pitch). Others repeated that the female voice is more pleasant ($N = 6$; an option already provided in Q37), calming or relaxing ($N = 4$), comforting, trustworthy, or familiar ($N = 3$), attracting attention ($N = 3$), soft ($N = 2$), authoritative ($N = 1$), believable ($N = 1$), or sensual ($N = 1$). Other

explanations included that most drivers are men (*N* = 3), that some drivers consider their cars feminine (*N* = 3), and that women play a protective role in people's life (*N* = 2).

2.3.4.5 Respondents' opinion on TORs as a function of the level of urgency

Figure 7 shows the opinion of the respondents on TORs in the five provided scenarios. A combination of all three types of messages (auditory, visual, and vibrotactile) was the preferred type of TOR for the scenarios of the highest levels of urgency, that is, Accident (Q64; *N* = 774) and Failure (Q63; *N* = 687). An auditory message was selected as the most preferred type of TOR for the scenarios of the lowest levels of urgency, that is, ExitHighway (Q60; *N* = 472) and ChangeLanes (Q61; *N* = 386). A combination of an auditory and a vibrotactile message was the most favourable TOR for the medium-urgency scenario ConstWorks (Q58; *N* = 446).

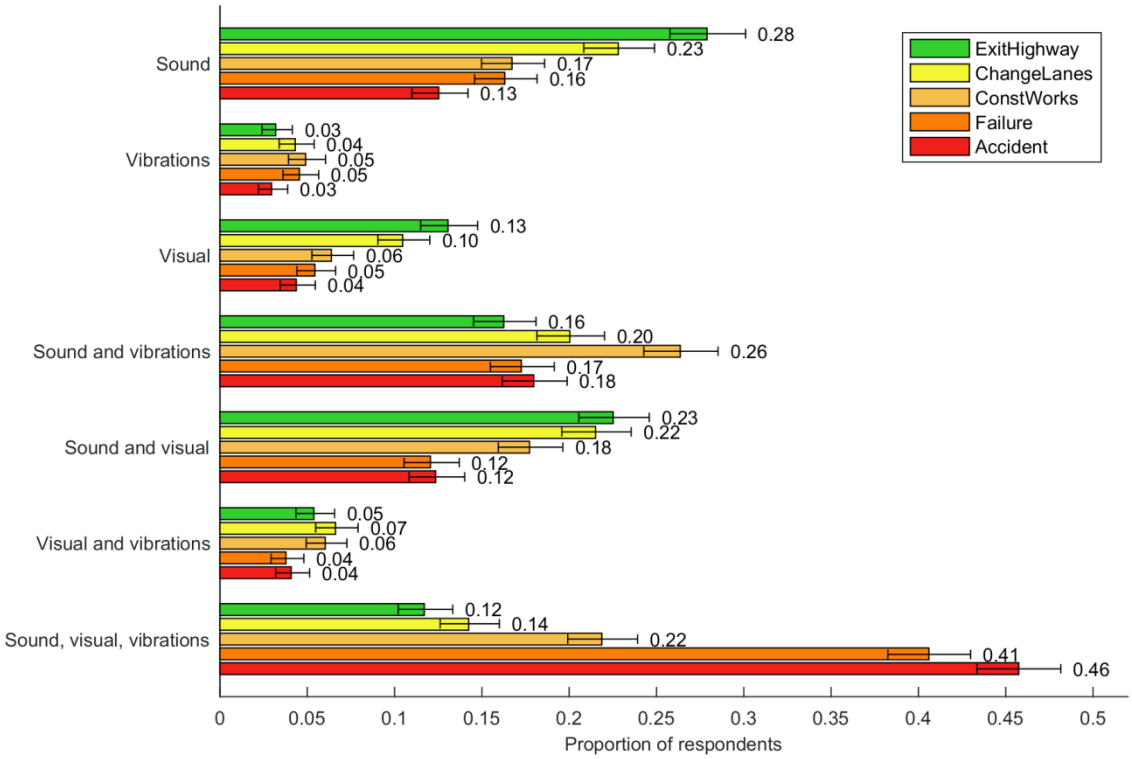


Figure 7. Proportion of respondents who selected particular take-over requests in the scenarios ExitHighway (Q60), ChangeLanes (Q61), ConstWorks (Q58), Failure (Q63), and Accident (Q64). The scenarios are presented in order of increasing urgency, from low to high. Error bars denote 95% confidence intervals.

Figure 8 shows the respondents' opinion on the types of messages they would like to receive for confirmation that the system is ready to switch back from manual to automated mode for the scenarios ConstWorks (Q59), ChangeLanes (Q62), and Accident (Q65). Auditory messages were the most preferred option for all three scenarios (ConstWorks: *N* = 634; ChangeLanes: *N* = 625; Accident: *N* = 604), followed by a combination of an auditory and a visual message (ConstWorks: *N* = 318; ChangeLanes: *N* = 329; Accident: *N* = 346). There were no distinct differences between the three scenarios (ChangeLanes, ConstWorks, Accident), which is logical, because the confirmation message indicates that the situation was over.

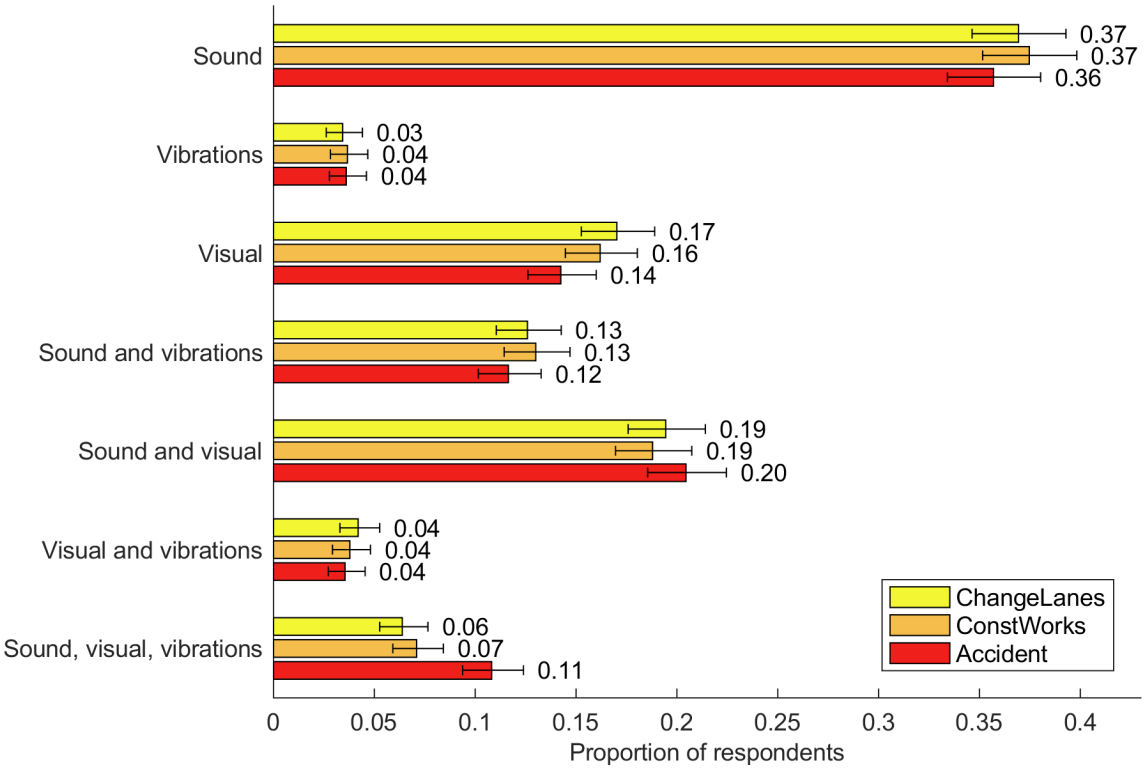


Figure 8. Proportion of respondents who selected particular messages for confirmation that the system is ready to switch back from manual to automated mode in the scenarios ChangeLanes (Q62), ConstWorks (Q59), and Accident (Q65). Error bars denote 95% confidence intervals.

2.3.4.6 Correlation analysis

Table S4 in the supplementary material contains a Spearman correlation matrix of selected predictor and outcome variables at the level of respondents. The correlations between the predictor variables and the outcome variables related to auditory (Audio6–10, Beeps1–4, Looming1–3, Audio15), visual (Visual1–6), and vibrotactile (Tactile1–5) TORs were overall small to moderate, ranging between −0.12 and 0.26. The mean of correlations between the trust score (Trust) and the variables that related to auditory, visual, and vibrotactile TORs were small: 0.04, 0.11, and 0.09, respectively. The highest correlation ($\rho = 0.24$) with Trust was found for Visual5, which was polling the people's acceptance of the automation taking over control after 5 s of a driver not looking at the road. The correlations of Trust with Audio15, Visual6, and Tactile5, which represent the opinion of people about whether auditory, visual, and vibrotactile messages, respectively, are a good way to get their attention back to the road, were 0.17, 0.18, and 0.16, respectively. The correlations between the DBQ violations score and the acceptance of the proposed messages were overly small, ranging between −0.06 and 0.06. A correlation of 0.27 was observed between the number of accidents over the last 3 years and the DBQ violations score, in line with the correlations found in previous crowdsourcing surveys ($\rho = 0.28$, $N = 1{,}862$ in De Winter et al., 2015; $\rho = 0.24$, $N = 1{,}205$ in Bazilinskyy & De Winter, 2015).

2.3.4.7 National comparisons

Figures 4 and 5 show that the respondents' preferences and ratings of urgency and annoyance were similar in middle- and high-income countries. Horn honking tended to be more preferred in middle- than in high-income countries, which is in line with De Winter & Dodou (2016) where it was found that lower-income countries exhibit more horn honking. Figure 9 shows the perceived urgency of long and very short beeps per country, as a function of GNI. It can be seen that in all countries long beeps were perceived as less urgent than very short beeps, with no interpretable trends with respect to the countries' GNI.

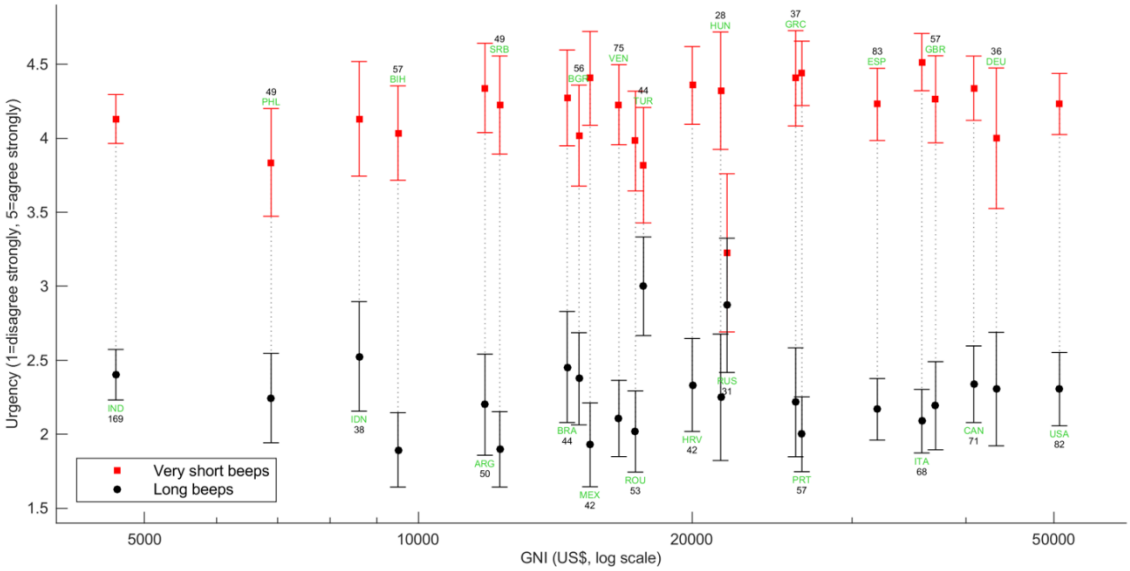


Figure 9. The reported urgency of long and very short beeps as a function of the gross national income (GNI) per capita. Error bars denote 95% confidence intervals. The numbers below or above each country abbreviation indicate the sample size per country. Country abbreviations are according to ISO 3166-1 alpha-3.

2.3.5 Discussion

In this study 3,000 respondents from 102 countries (1,692 respondents from 91 countries after filtering) were asked to state their opinion on various types of auditory, visual, and vibrotactile messages for TORs in general and in five scenarios of various levels of urgency during highly automated driving. In three of the scenarios, the respondents were also asked to indicate what type of messages they would prefer to receive to indicate that the system is ready to switch back from manual to automated mode.

2.3.5.1 Respondents' opinion on unimodal messages in highly automated driving

The respondents were moderately positive towards all three proposed (auditory, vibrotactile, and visual) types of messages for supporting TORs. Among the provided auditory messages, large differences in preference were observed, with verbal messages being more accepted than abstract sounds. Additionally, the female voice was ranked higher in preference than the male voice, in line with findings from a previous online survey (Bazilinskyy & De Winter, 2015). The respondents indicated their opinion regarding why the female voice may be seen as a preferred modality of feedback in cars. The most selected answer (by 51%

of the respondents) was “A female voice is more pleasant”. Our results further showed that male and female respondents preferred the female voice for different reasons: males in particular found a female voice pleasant, whereas females in particular found a female voice less commanding and easier to agree with than a male voice.

The difference in the perceived urgency of the male versus female voice in our study was small, with the male and female voices being seen as the most urgent type of auditory warning by 518 and 495 out of the 1,692 respondents, respectively (see Q35). Note that no background noise was present in the auditory recordings used in our survey. It has been shown that a male voice is more audible in environments with high noise levels (Nixon et al., 1998). In our survey, in the question asking respondents to indicate their opinion on why the female voice is often seen as the most preferred type of auditory message to be used in cars, 17% (N = 285) of the respondents considered that the female voice was easier to understand in a noisy environment.

The respondents perceived the horn signal as the most annoying type of auditory warning, which is in agreement with the survey results by Bazilinskyy and De Winter (2015). We observed a strong linear relationship between the perceived urgency ratings and the beep rate (r = .999; exponent = 0.66), which verifies previous experimental studies showing that the relationship between perceived urgency and interpulse duration follows Stevens’ power law (exponent = 0.61 in Hellier & Edworthy, 1989; between 0.47 and 0.77 in four experiments presented in Gonzalez, Lewis, & Baldwin, 2012; but see also Hellier et al., 1993, for a higher exponent of 1.35). The confirmation of this relationship by means of survey data is an indication that crowdsourcing offers an attractive solution for psychophysics research, which is traditionally carried out in the lab (Stevens & Boring, 1947). Note, however, that generalizations should be made with caution, as the respondents in our survey did not conduct a true magnitude estimation task on a ratio scale, but were asked to report their perceived urgency on a scale from *Disagree strongly* to *Agree strongly*. Further discussion on ratio scales versus ordinal scales in psychophysics research is provided by Stevens and Galanter (1957) and Walker (2002).

The looming warning of the longest duration (5 s) was perceived as the most urgent of the three looming warnings provided. This was unexpected, as we reasoned that a high rate of volume increase (i.e., a looming sound of short duration rapidly building up to a maximum volume) indicates a danger that approaches more rapidly. Note, however, that the looming warnings in our survey lacked context. That is, the respondents were not exposed to visual input from the environment (e.g., time-to-contact or other hazards). Moreover, it is possible that the long duration/slow volume increase in the 5 s looming was experienced as more arousing than the fast looming, in which the volume increase was difficult to perceive because the signal was of short duration (1 s). To shed more light on the effect of looming sounds on the perceived urgency, we recommend driving simulator research in which both (1) the duration of the looming sound and (2) the actual build-up of inter-vehicular conflict are systematically investigated in a two-factor design.

Among the visual messages provided in the part of the questionnaire investigating the respondents’ general opinion on feedback modalities, the head-up display with a green icon and the green icon on the dashboard were the most preferred. When presented within the context of specific TOR

scenarios, a combination of auditory and visual messages was moderately accepted for TORs of low and medium urgency (exiting the highway and changing lanes). For low-urgency scenarios, visual-only messages were more often selected than vibration-only messages, which could be explained by the fact that visual information can remain in view for a prolonged amount of time (e.g., as a state indicator), whereas auditory (and vibrotactile) warnings are preferred in cases in which the operator does not need to refer to the message later in time (Stanton & Edworthy, 1999; see also Wickens et al., 2012).

All four vibrotactile messages received similar ratings in the part of the questionnaire on general opinion (i.e., without offering the context of a specific scenario). Although in highly automated driving the human is not required to keep the hands on the steering wheel, providing vibrations via the steering wheel was rated as most preferred. It has to be noted that the survey did not mention whether the driver could take the hands off the steering wheel. Respondents did not prefer vibrotactile warnings as a single modality or in combination with visual warnings for any of the five TOR scenarios.

The correlations between the trust score and the acceptance of the proposed types of messages were small to moderate. The respondents with high trust scores indicated that they would prefer the automation to be able to take over control after 5 s of a driver not looking at the road ($\rho = 0.24$), which is an expected result, because adaptive automation ("automation design where tasks are dynamically allocated between the human operator and computer", Byrne & Parasuraman, 1996, p. 249) requires trust in automation (Parasuraman, Sheridan, & Wickens, 2008).

The preferences of the respondents for the use of messages for supporting take-over requests and the perception of auditory messages as the most urgent and the most annoying were similar in middle- and high-income countries, indicating that our findings are generalizable across different countries. These findings are in line with Klein et al. (2014) who found that effect sizes in typical psychological experiments are about the same regardless of whether the experiment had been executed in the United States or not.

2.3.5.2 Respondents' opinion on TORs as a function of the level of urgency

Five scenarios, with different levels of urgency, were presented in our survey (highway exit, changing lane, construction works, automation failure, and traffic accident ahead). For each of the five scenarios, the respondents were asked to state the modalities that they preferred for TORs, whereas for three of the scenarios (i.e., construction works, changing lane, and traffic accident), the respondents were also asked to choose their preferred modality for receiving a confirmation message that the system is ready to switch back from manual to automated mode. The combination of auditory, visual, and vibrotactile messages was the most preferred option for supporting TORs in high-urgency scenarios. This survey adds to the existing knowledge by showing that people have a preference for multimodal warnings when the situation becomes more critical, in line with the theories on the benefits of multimodal over unimodal displays described in the introduction (see also Haas & Casali, 1995; Selcon et al., 1995; Petermeijer, Abbink, Mulder, & De Winter, 2015). Future behavioural research should investigate whether tri-modal TORs are indeed more effective than audio-visual ones.

An auditory message was selected as the most preferred option for low-urgency scenarios. Also for confirmation that the system is ready to switch back from manual to automated mode, were auditory messages the most preferred option for all three provided scenarios, followed by a combination of auditory and visual messages. These results indicate that the respondents wanted different types of messages for urgent TORs and for low-urgency TORs/confirmation messages: multimodal displays were preferred for urgent TORs, whereas traditional audio-visual displays were preferred for warnings in low-urgency situations and as state indicators.

2.3.5.3 Limitations and future work

We applied a strict screening by excluding all participants who failed one or more test questions. We also excluded participants who skipped a question or who did not report having a driver's license. Accordingly, we expect that our data are of good quality and that our findings are robust. For example, we expect that the observation that mean perceived urgency increases as a function of the beep rate is replicable and generalizable. However, other findings may be contingent on context, as participants had to imagine concepts of displays in a highly automated vehicle based on textual, visual, and auditory descriptions. The results could be different if participants were physically interacting with such displays. This is particularly true for vibrotactile displays, which are not often used in the automotive domain (Meng & Spence, 2015), meaning that it is likely that the respondents had not experienced vibrotactile displays in the context of driving before (note, however, that humans nowadays are familiar with receiving vibrotactile notifications from their mobile devices).

A related limitation is that the survey merely contained descriptions of TOR scenarios and that the respondents had to imagine being distracted before judging which of the proposed messages would be the most preferred to draw their attention back to the driving scene. For example, participants had to imagine that visual warnings could be ineffective if the driver is engaged in a visually demanding non-driving task such as reading. Behavioural research in simulators or on actual roads should be conducted to acquire knowledge on the effectiveness of the warnings in scenarios that involve visual distractions or competing stimuli, such as a visually demanding outside environment, task-intrinsic vibrations in the driver seat (e.g., 'road rumble'), and sounds from in-vehicle devices and other road users (e.g., emergency vehicles). It is possible to extend the present crowdsourcing research towards an interactive environment in which participants are shown videos or animations of traffic scenes and have to respond to actual sounds and visual warnings. Although the present survey was of low perceptual fidelity (i.e., participants did not actually experience the interfaces in a driving context), it offered high scope and statistical power, with a large number of participants worldwide being surveyed with identical instructions, questions, sounds, and images. However, the participants' hardware and volume settings were not under our control.

Furthermore, we note that the TORs which we investigated are only a selection of all possible design solutions. For example, only one male voice and one female voice were provided, which means that results may be contingent on the voice generator, accent, pitch, etc. (for a similar discussion on TORs see Bazilinskyy & De Winter, 2017, in which participants rated a number of male and

female voices). Also, the five traffic situations are merely a selection of possible scenarios in actual automated driving. For example, transitions of control may be mandatory in case the driver is unable to take over (Lu et al., 2016), and there may be situations where the car has to bring itself to a minimum risk condition (Hoeger et al., 2008; Gasser & Westhoff, 2012).

A final point of attention is whether CrowdFlower respondents are representative of the (expectedly high-income) future stakeholders of highly automated driving technology. The representativeness of crowdsourced samples is the topic of ongoing research, and it has become clear that the active CrowdFlower and Amazon Mechanical Turk populations are relatively small (<10,000 persons), consisting of people who have evolved into specialized research participants and who may spend a large share of their time behind the computer (Chandler, Mueller, & Paolacci, 2014; Stewart et al., 2015). Of the 1,692 respondents, 67% were located in high-income countries and the remaining 33% were located in middle-income countries. Hence, a large number of the participants were from countries where highly automated vehicles will likely be initially introduced to public roads, and where warning and assistance systems such as Adaptive Cruise Control, Lane Keeping Assist, and parking sensors, are already common (Bishop, 2005; Bazilinskyy & De Winter, 2015). However, a previous study showed that only 8% of the CrowdFlower participants reported a gross annual income of EUR 44,000 or more (Kyriakidis et al., 2015). Thus, it is unlikely that the respondents themselves represent the typical early adopters of automated driving technology.

2.3.5.4 Conclusions

In conclusion, this study allowed us to access a large and diverse population to gain a first impression about display design for highly automated driving. The survey showed that people's preferences for the type of messages they would like to receive in highly automated driving depend on the urgency of the situation ahead and on whether the message concerns a TOR or a confirmation that the system is ready to switch back from manual to automated mode. Specifically: (1) For high-urgency situations, multimodal warnings were the most preferred option, (2) For low-urgency situations and for receiving confirmation that the system is ready to switch back from manual to automated mode, auditory messages were the most preferred option, (3) For low-urgency scenarios, visual-only TORs were more preferred than vibration-only TORs, (4) Among messages represented by beeps, beeps with shorter interpulse intervals were perceived as more urgent, consistent with Stevens' power law, (5) Among the five provided sounds, spoken messages were more accepted than abstract sounds, and the female voice was more preferred than the male voice.

With a large sample of 1,692 respondents, our survey validates previous experimental findings and theoretical considerations with respect to preferred modalities for issuing TORs during highly automated driving as a function of the urgency level of the TOR. The survey may be launched again in the future among high-income users who are likely to purchase automated cars. The present results are of preliminary and abstract nature, and do not immediately allow for specific design recommendations concerning TORs in future automated vehicles. Simulator-based and on-road studies need to be performed, and

designers need to be involved, before being able to transfer the results to TORs in real vehicles.

2.3.6 Acknowledgments

We would like to express our special gratitude to Daria Nikulina for designing the illustrations used in the survey.

2.3.7 Supplementary material

Supplementary material may be found at https://doi.org/10.4121/uuid:e3908ec5-d086-4737-8d4a-d4046dbbc53c

2.3.8 References

Arrabito, G. R. (2009). Effects of talker sex and voice style of verbal cockpit warnings on performance. *Human Factors, 51*, 3–20. https://doi.org/10.1177/0018720808333411

Bach, D. R., Schächinger, H., Neuhoff, J. G., Esposito, F., Di Salle, F., Lehmann, C., ... & Seifritz, E. (2008). Rising sound intensity: an intrinsic warning cue activating the amygdala. *Cerebral Cortex, 18*, 145–150. https://doi.org/10.1093/cercor/bhm040

Baldwin, C. L., Eisert, J. L., Garcia, A., Lewis, B., Pratt, S. M., & Gonzalez, C. (2012). Multimodal urgency coding: auditory, visual, and tactile parameters and their impact on perceived urgency. *Work-Journal of Prevention Assessment and Rehabilitation, 41*, 3586–3591. https://doi.org/10.3233/WOR-2012-0669-3586

Bate, A. J. (1969). *Cockpit warning systems comparative study* (Final Report AMRL-TR, 68-193). Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory, Aerospace Medical Division, Air Force Systems Command.

Bazilinskyy, P., & De Winter, J. (2015). Auditory interfaces in automated driving: an international survey. *PeerJ Computer Science*, *1*, e13. https://doi.org/10.7717/peerj-cs.13

Bazilinskyy, P., & De Winter, J. C. F. (2017). Analyzing crowdsourced ratings of speech-based take-over requests for automated driving. *Applied Ergonomics, 64*, 56–64.

Bazilinskyy, P., Kyriakidis, M., & De Winter, J. C. F. (2015). An international crowdsourcing study into people's statements on fully automated driving. Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics (AHFE),* Las Vegas, NV*,* 2534–2542.

Begg, D. (2014). A 2050 vision for London: What are the implications of driverless transport. Available at http://www.transporttimes.co.uk/Admin/uploads/64165-Transport-Times_A-2050-Vision-for-London_AWWEB-READY.pdf

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43*, 800–813. https://doi.org/10.3758/s13428-011-0081-0

Birrell, S. A., Young, M. S., & Weldon, A. M. (2013). Vibrotactile pedals: provision of haptic feedback to support economical driving. *Ergonomics, 56*, 282–292. https://doi.org/10.1080/00140139.2012.760750

Bishop, R. (2005). Intelligent vehicle R&D: A review and contrast of programs worldwide and emerging trends. *Annales Des Télécommunications, 60*, 228–263. https://doi.org/10.1007/BF03219820

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction, 4*, 11–44.

BMW (2013). The BMW 3 Series Sedan. Owner's manual. Available at https://carmanuals2.com/get/bmw-335i-2013-owner-s-manual-55036

Burke, J. L., Prewett, M. S., Gray, A. A., Yang, L., Stilson, F. R. B., Coovert, M. D., ... & Redden, E. (2006). Comparing the effects of visual-auditory and visual-tactile feedback on user performance: A meta-analysis. Proceedings of *the 8th International Conference on Multimodal Interfaces*,108–117, New York: ACM.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology, 42*, 249–268. https://doi.org/10.1016/0301-0511(95)05161-9

Cabrall, C. D. D., Lu, Z., Kyriakidis, M., Manca, L., Dijksterhuis, C. Happee, R., & De Winter, J. C. F. (2016). Validity and reliability of naturalistic driving scene categorization judgments from crowdsourcing. Proceedings of *the 17th International Conference Road Safety On Five Continents (RS5C)*. Rio de Janeiro, Brazil.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46*, 112–130. https://doi.org/10.3758/s13428-013-0365-7

CrowdFlower (2015). Job settings | Guide to: Basic job settings page. Available at https://success.crowdflower.com/hc/en-us/articles/201855719-Job-Settings-Guide-to-Basic-Job-Settings-Page

De Groot, S., De Winter, J. C. F., Mulder, M., Kuipers, J., Mulder, J. A., & Wieringa, P. A. (2013). The effects of route-instruction modality on driving performance in a simulator. In De Groot, S. *Exploiting the possibilities of simulators for driver training* (Doctoral dissertation). Delft University of Technology, the Netherlands.

De Groot, S., De Winter, J. C. F., García, J. M. L., Mulder, M., & Wieringa, P. A. (2011). The effect of concurrent bandwidth feedback on learning the lane-keeping task in a driving simulator. *Human Factors, 53*, 50–62. https://doi.org/10.1177/0018720810393241

De Winter, J. C. F. (2013). Predicting self-reported violations among novice license drivers using pre-license simulator measures. *Accident Analysis and Prevention, 52*, 71–79. https://doi.org/10.1016/j.aap.2012.12.018

De Winter, J. C. F., & Dodou, D. (2016). National correlates of self-reported traffic violations across 41 countries. *Personality and Individual Differences, 98*, 145–152. https://doi.org/10.1016/j.paid.2016.03.091

De Winter, J. C. F., Dodou, D., & Stanton, N. A. (2015). A quarter of a century of the DBQ: Some supplementary notes on its validity with regard to accidents. *Ergonomics, 58*, 1745–1769. https://doi.org/10.1080/00140139.2015.1030460

De Winter, J. C. F., & Hancock, P. A. (2015). Reflections on the 1951 Fitts list: Do humans believe now that machines surpass them? Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics*

(AHFE), Las Vegas, NV, 5334–5341.
De Winter, J. C. F., Kyriakidis, M., Dodou, D., & Happee, R. (2015). Using CrowdFlower to study the relationship between self-reported violations and traffic accidents. Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Las Vegas, NV, 2518–2525.
Dettmann, A., & Bullinger, A. C. (2017). Spatially distributed visual, auditory and multimodal warning signals–a comparison. In De Waard, D., Toffetti, A., Wiczorek, R., Sonderegger, A., Röttger, S., Bouchner, P., Franke, T., Fairclough, S., Noordzij, M., & Brookhuis, K.. (2016). Proceedings of *the Human Factors and Ergonomics Society Europe Chapter 2016 Annual Conference*, 185–199.
Diaconescu, A. O., Alain, C., & McIntosh, A. R. (2011). The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *Journal of Neurophysiology, 106*, 2896–2909. https://doi.org/10.1152/jn.00303.2011
Edworthy, J., & Hellier, E. J. (2003). Urgency in speech warnings. In R. G. De Jong, T. Houtgast, E. A. M. Fransten, & W. F. Hofman (Eds.), Proceedings of *the 8th International Congress on Noise as a Public Health Problem*, 65–69, Rotterdam, the Netherlands.
Eriksson, A., Solis Marcos, I., Kircher, K., Västfjäll, D., & Stanton, N. A. (2015). The development of a method to assess the effects of traffic situation and time pressure on driver information preferences. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics* (pp. 3–12). Springer International Publishing.
Fitch, G. M., Hankey, J. M., Kleiner, B. M., & Dingus, T. A. (2011). Driver comprehension of multiple haptic seat alerts intended for use in an integrated collision avoidance system. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*, 278–290. https://doi.org/10.1016/j.trf.2011.02.001
Fitts, P. M., & Jones, R. E. (1947). Psychological aspects of instrument display: 1. Analysis of 270 “pilot-error” experiences in reading and interpreting aircraft instruments (Memorandum Report TSEAA-694-12A). Wright-Patterson Air Force Base, OH: Air Materiel Command.
Flemisch, F., Kaussner, A., Petermann, I., Schieben, A., & Schöming, N. (2011). *HAVE-IT. Highly automated vehicles for intelligent transport. Validation of concept on optimum task repartition* (Deliverable D.33.6). Regensburg, Germany: Continental Automotive GmbH.
Gasser, T. & Westhoff, D. (2012). BASt-study: Definitions of automation and legal issues in Germany. *TRB Road Vehicle Automation Workshop*. Irvine, CA, USA.
Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction, 2*, 167–177.
Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). “Take over!” How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*, 1938–1942. https://doi.org/10.1177/1541931213571433
Gold, C., Lorenz, L., Damböck, D., & Bengler, K. (2013). Partially automated driving as a fallback level of high automation. 6. Tagung Fahrerassistenzsysteme. Der Weg zum automatischen Fahren. TÜV SÜD

Akademie GmbH.
Gonzalez, C., Lewis, B. A., & Baldwin, C. L. (2012). Revisiting pulse rate, frequency and perceived urgency: have relationships changed and why? Proceedings of *the 18th International Conference on Auditory Display,* Atlanta, GA, 44–51.
Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*, 93–104. https://doi.org/10.1037/0003-066X.59.2.93
Graham, R. (1999). Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. *Ergonomics, 42*, 1233–1248. https://doi.org/10.1080/001401399185108
Gray, R. (2011). Looming auditory collision warnings for driving. *Human Factors, 53*, 63–74. https://doi.org/10.1177/0018720810397833
Grah, T., Epp, F., Wuchse, M., Meschtscherjakov, A., Gabler, F., Steinmetz, A., & Tscheligi, M. (2015). Dorsal haptic display: a shape-changing car seat for sensory augmentation of rear obstacles. Proceedings of *the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 305–312, ACM.
Green, P. (1999). *Visual and task demands of driver information systems* (Technical Report UMTRI-98-16). Ann Arbor, MI: University of Michigan Transportation Research Institute.
Grether, W. F. (1949). Instrument reading. I. The design of long-scale indicators for speed and accuracy of quantitative readings. *Journal of Applied Psychology, 33*, 363–372. https://doi.org/10.1037/h0058374
Haas, E. C., & Casali, J. G. (1995). Perceived urgency of and response time to multi-tone and frequency-modulated warning signals in broadband noise. *Ergonomics, 38*, 2313–2326. https://doi.org/10.1080/00140139508925270
Haas, E. C., & Edworthy, J. (1999). The perceived urgency and detection time of multitone auditory signals. In N. A. Stanton & J. Edworthy (Eds.), *Human factors in auditory warnings* (pp. 129–149). Aldershot: Ashgate Publishing Limited.
Hancock, P. A., Lawson, B., Cholewiak, R., Elliott, L. R., Van Erp, J. B., Mortimer, B. J., ... & Redden, E. S. (2015). Tactile cuing to augment multisensory human-machine interaction. *Ergonomics in Design: The Quarterly of Human Factors Applications, 23*, 4–9. https://doi.org/10.1177/1064804615572623
Hellier, E., & Edworthy, J. (1989). Quantifying the perceived urgency of auditory warnings. *Canadian Acoustics, 17*, 3–11.
Hellier, E., & Edworthy, J. (1999). On using psychophysical techniques to achieve urgency mapping in auditory warnings. *Applied Ergonomics, 30*, 167–171. https://doi.org/10.1016/S0003-6870(97)00013-6
Hellier, E. J., Edworthy, J., & Dennis, I. (1993). Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency. *Human Factors, 35*, 693–706. https://doi.org/10.1177/001872089303500408
Hellier, E., Edworthy, J., Weedon, B., Walters, K., & Adams, A. (2002). The perceived urgency of speech warnings: Semantics versus acoustics. *Human Factors, 44*, 1–17. https://doi.org/10.1518/0018720024494810
Hoeger, R., Zeng, H., Hoess, A., Kranz, T., Boverie, S., Strauss, M., et al.

(2011). *The future of driving – HAVEit* (Final Report, Deliverable D61.1). Regensburg, Germany: Continental Automotive GmbH.

Honda (2014). 2014 ACCORD 4D online reference owner's manual. Available at https://carmanuals2.com/get/honda-accord-2014-owner-s-manual-80195

Jang, P. S. (2007). Designing acoustic and non-acoustic parameters of synthesized speech warnings to control perceived urgency. *International Journal of Industrial Ergonomics, 37*, 213–223. https://doi.org/10.1016/j.ergon.2006.10.018

Jones, L. A., & Sarter, N. B. (2008). Tactile displays: Guidance for their design and application. *Human Factors, 50*, 90–111. https://doi.org/10.1518/001872008X250638

Kelsch, J., & Dziennus, M. (2015). Joint driver-automation system design: Gradual action-oriented ambient stimuli. Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Las Vegas, NV.

Kim, N., Jeong, K., Yang, M., Oh, Y., & Kim, J. (2017). Are you ready to take-over? An exploratory study on visual assistance to enhance driver vigilance. Proceedings of *the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 1771–1778, New York, NY: ACM.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., ... & Cemalcilar, Z. (2014). Investigating variation in replicability. *Social Psychology, 45*, 142–152. https://doi.org/10.1027/1864-9335/a000178

Kyriakidis, M., Happee, R., & De Winter, J. C. F. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour, 32*, 127–140. https://doi.org/10.1016/j.trf.2015.04.014

Lamble, D., Laakso, M., & Summala, H. (1999). Detection thresholds in car following situations and peripheral vision: Implications for positioning of visually demanding in-car displays. *Ergonomics, 42*, 807–815. https://doi.org/10.1080/001401399185306

Langlois, S., & Soualmi, B. (2016). Augmented reality versus classical HUD to take over from automated driving: An aid to smooth reactions and to anticipate maneuvers. Proceedings of *the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC),* 1571–1578.

Large, D. R., & Burnett, G. E. (2013). Drivers' preferences and emotional responses to satellite navigation voices. *International Journal of Vehicle Noise and Vibration, 9*, 28–46. https://doi.org/10.1504/IJVNV.2013.053815

Lee, J., McGehee, D., Brown, T., & Marshall, D. (2006). Effects of adaptive cruise control and alert modality on driver performance. *Transportation Research Record: Journal of the Transportation Research Board, 1980*, 49–56. https://doi.org/10.3141/1980-09

Lee, J. H., & Spence, C. (2008). Assessing the benefits of multimodal feedback on dual-task performance under demanding conditions. Proceedings of *the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction, 1*,185–192. British Computer Society.

Liu, Y.-C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics, 44*, 425–442. https://doi.org/10.1080/00140130010011369

Löcken, A., Heuten, W., & Boll, S. (2015). Supporting lane change decisions with ambient light. Proceedings of *the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 204–211, New York, NY: ACM.

Lu, S. A., Wickens, C. D., Sarter, N. B., & Sebok, A. (2011). Informing the design of multimodal displays: a meta-analysis of empirical studies comparing auditory and tactile interruptions. Proceedings of *the Human Factors and Ergonomics Society Annual Meeting, 55*, 1170–1174. https://doi.org/10.1177/1071181311551244

Lu, Z., Happee, R., Cabrall, C. D. D., Kyriakidis, M., & De Winter, J. C. F. (2016). Human factors of transitions in automated driving: A general framework and literature survey. *Transportation Research Part F: Traffic Psychology and Behaviour, 43*, 183–198.

Manca, L., De Winter, J. C. F., & Happee, R. (2015). Visual displays for automated driving: A survey. Paper presented at *the Workshop on Adaptive Ambient In-Vehicle Displays and Interactions*, Nottingham, UK.

Meng, F., Ho, C., Gray, R., & Spence, C. (2015). Dynamic vibrotactile warning signals for frontal collision avoidance: towards the torso versus towards the head. *Ergonomics, 58*, 411–425. https://doi.org/10.1080/00140139.2014.976278

Meng, F., & Spence, C. (2015). Tactile warning signals for in-vehicle systems. *Accident Analysis and Prevention, 75*, 333–346. https://doi.org/10.1016/j.aap.2014.12.013

Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2012). I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human Factors, 55*, 520–534. https://doi.org/10.1177/0018720812465081

Meschtscherjakov, A., Döttlinger, C., Rödel, C., & Tscheligi, M. (2015). ChaseLight: ambient LED stripes to control driving speed. Proceedings of *the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 212–219, New York, NY: ACM.

Mok, B., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., & Ju, W. (2015). Emergency, automation off: unstructured transition timing for distracted drivers of automated vehicles. *Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems*. Canary Islands, Spain.

Naujoks, F., Forster, Y., Wiedemann, K., & Neukum, A. (2016). Speech improves human-automation cooperation in automated driving. *Workshop Automotive HMI*, Aachen. https://doi.org/10.18420/muc2016-ws08-0007

Naujoks, F., Mai, C., & Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. In N. Stanton, S. Landry, G. Di Bucchianico, & A. Vallicelli (Eds.), *Advances in Human Aspects of Transportation: Part I* (pp. 431–438).

Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. *Reviews of Human Factors and Ergonomics, 7*, 58–99. https://doi.org/10.1177/1557234X11410396

Nixon, C. W., Morris, L. J., McCavitt, A. R., McKinley, R. L., Anderson, T. R., McDaniel, M. P., & Yeager, D. G. (1998). Female voice communications in high levels of aircraft cockpit noises–part I: spectra, levels, and microphones. *Aviation, Space, and Environmental Medicine, 69*, 675–683.

Noyes, J. M., Hellier, E., & Edworthy, J. (2006). Speech warnings: A review. *Theoretical Issues in Ergonomics Science, 7*, 551–571. https://doi.org/10.1080/14639220600731123

Oviatt, S. (2003). Multimodal interfaces. In J. Jacko & A. Sears (Eds.), *Handbook of human-computer interaction* (pp. 286–304). Hillsdale, NJ: Erlbaum.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230–253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*, 140–160. https://doi.org/10.1518/155534308X284417

Park, K. S., & Jang, P. S. (1999). Effects of synthesized voice warning parameters on perceived urgency. *International Journal of Occupational Safety and Ergonomics, 5*, 73–95. https://doi.org/10.1080/10803548.1999.11076412

Patterson, R. D. (1982). *Guidelines for auditory warning systems on civil aircraft* (CAA paper 82017). London: Civil Aviation Authority.

Patterson, R. D., & Mayfield, T. F. (1990). Auditory warning sounds in the work environment [and Discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 327*, 485–492. https://doi.org/10.1098/rstb.1990.0091

Petermeijer, S. M., Abbink, D. A., Mulder, M., & De Winter, J. C. F. (2015). The effect of haptic support systems on driver performance: a literature survey. *IEEE Transactions on Haptics, 8*, 467–479. https://doi.org/10.1109/TOH.2015.2437871

Petermeijer, S. M., De Winter, J. C. F., & Bengler, K. J. (2016). Vibrotactile displays: a survey with a view on highly automated driving. *IEEE Transactions on Intelligent Transporation Systems, 17*, 897–907. https://doi.org/10.1109/TITS.2015.2494873

Pfromm, M., Cieler, S., & Bruder, R. (2013). Driver assistance via optical information with spatial reference. Proceedings of *the 16th International IEEE Conference on Intelligent Transportation Systems*, The Hague, the Netherlands, 2006–2011.

Politis, I., Brewster, S. A., & Pollick, F. (2014). Evaluating multimodal driver displays under varying situational urgency. Proceedings of *the SIGCHI Conference on Human Factors in Computing Systems*, 4067–4076.

Politis, I., Brewster, S., & Pollick, F. (2015a). To beep or not to beep?: Comparing abstract versus language-based multimodal driver displays. Proceedings of *the 33rd Annual ACM Conference on Human Factors in Computing Systems,* 3971–3980.

Politis, I., Brewster, S., & Pollick, F. (2015b). Language-based multimodal displays for the handover of control in autonomous cars. Proceedings of *the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications,* 3–10.

Prewett, M. S., Elliott, L. R., Walvoord, A. G., & Coovert, M. D. (2012). A meta-analysis of vibrotactile and visual information displays for improving task performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part*

C: Applications and Reviews, 42, 123–132. https://doi.org10.1109/TSMCC.2010.2103057

Ramkissoon, R. N. (2001). *The use of auditory cues to provide dynamic alerts to drivers* (Doctoral dissertation). Massachusetts Institute of Technology.

Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S., Balaji, T. S., Buisine, S., ... & Wang, Q. Y. (2004). Guidelines for multimodal user interface design. *Communications of the ACM, 47*, 57–59. https://doi.org/10.1145/962081.962106

Sanders, M. S., & McCormick, E. J. (1987). *Human Factors in Engineering and Design.* New York: McGraw-Hill.

Schwalk, M., Kalogerakis, N., & Maier, T. (2015). Driver support by a vibrotactile seat matrix – Recognition, adequacy and workload of tactile patterns in take-over scenarios during automated driving. *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Las Vegas, NV, 1427–1434.

Scott, J. J., & Gray, R. (2008). A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. *Human Factors*, *50*, 264–275.

Selcon, S. J., Taylor, R. M., & McKenna, F. P. (1995). Integrating multiple information sources: Using redundancy in the design of warnings. *Ergonomics, 38*, 2362–2370. https://doi.org/10.1080/00140139508925273

Shaw, B. K., McGowan, R. S., & Turvey, M. T. (1991). An acoustic variable specifying time to contact. *Ecological Psychology, 3*, 253–261. https://doi.org/10.1207/s15326969eco0303_4

Society of Automotive Engineers [SAE] (2014). Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems (Standard No. J3016). SAE International. Available at http://standards.sae.org/j3016_201401

Silva, R. M., Lamas, J., Silva, C. C., Coello, Y., Mouta, S., & Santos, J. A. (2017). Judging time-to-passage of looming sounds: Evidence for the use of distance-based information. *PLOS ONE*, *12*, e0177734.

Sivak, M., & Owens, D. A. (1996). The information that drivers use: is it indeed 90% visual? *Perception, 25*, 1081–1090. https://doi.org/10.1068/p251081

Spence, C., & Ho, C. (2008). Tactile and multisensory spatial warning signals for drivers. *IEEE Transactions on Haptics, 1*, 121–129. https://doi.org/10.1109/TOH.2008.14

Stanton, N. A., & Edworthy, J. (1999). Auditory warnings and displays: An overview. In N. A. Stanton & J. Edworthy (Eds.), *Human factors in auditory warnings* (pp. 3–30). Aldershot: Ashgate Publishing Limited.

Stevens, S. S., & Boring, E. G. (1947). The new Harvard psychological laboratories. *American Psychologist, 2*, 239–243. https://doi.org/10.1037/h0060913

Stevens, S. S. (1957). On the psychophysical law. *The Psychological Review, 64*, 153–181. https://doi.org/10.1037/h0046162

Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology, 54*, 377–411. https://doi.org/10.1037/h0043680

Stewart, N., Ungemach, C., Harris, A. J., Bartels, D. M., Newell, B. R., Paolacci, G., & Chandler, J. (2015). The average laboratory samples a population of

7,300 Amazon Mechanical Turk workers. *Judgment and Decision Making, 10*, 479–491.
Stokes, A., Wickens, C., & Kite, K. (1990). *Display technology-human factors concepts* (Technical Report SAE R-102). Warrendale, PA: Society of Automotive Engineers, Inc.
Swain, A. D., & Guttmann, H. E. (1983). *Handbook of human reliability analysis with emphasis on nuclear power plant applications* (Final Report NUREG/CR-1278). Albuquerque, NM: Sandia National Laboratories.
Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14*, 400–410. https://doi.org/10.1016/j.tics.2010.06.008
Telpaz, A., Rhindress, B., Zelman, I., & Tsimhoni, O. (2015). Haptic seat for automated driving: preparing the driver to take control effectively. Proceedings of *the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 23–30, New York, NY: ACM.
The World Bank (2015). Indicators [data]. Available at http://data.worldbank.org/indicator/
The World Bank (2016). Country and lending groups. Available at http://data.worldbank.org/about/country-and-lending-groups
Toyota (2014). 4RUNNER 2014. Owner's manual. Available at https://carmanuals2.com/get/toyota-4runner-2014-owner-s-manual-89510
Underwood, S. E. (2014). Automated vehicles forecast vehicle symposium opinion survey. Paper presented at *the Automated Vehicles Symposium 2014*. San Francisco, CA.
Van Erp, J. B. F., Toet, A., & Janssen, J. B. (2015). Uni-, bi-and tri-modal warning signals: Effects of temporal parameters and sensory modality on perceived urgency. *Safety Science, 72*, 1–8. https://doi.org/10.1016/j.ssci.2014.07.022
Volkswagen (2014). Park assist. Available at http://www.volkswagen.co.uk/technology/parking-and-manoeuvring/park-assist
Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied, 8,* 211–221. https://doi.org/10.1037/1076-898X.8.4.211
Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2012). *Engineering psychology and human performance (Fourth edition)*. Psychology Press.
Wickens, C. D., & Seppelt, B. (2002). *Interference with driving or in-vehicle task information: The effects of auditory versus visual delivery* (Technical Report AHFD-02-18/GM-02-3). Urbana-Champaign: University of Illinois at Urbana-Champaign, Aviation Human Factors Division.
Winkler, S., Powelleit, M., Kazazi, J., Vollrath, M., Krautter, W., Korthauer, A., ... & Bendewald, L. (2018). HMI strategy–warnings and interventions. In K. Bengler, J. Drüke, S. Hoffmann, D. Manstetten, & A. Neukum (Eds.), *UR:BAN Human factors in traffic. Approaches for safe, efficient and stress-free urban traffic* (pp. 75–103). Wiesbaden: Springer Vieweg.
World Health Organization (2015). WHO global status report on road safety 2015. WHO.

You, F., Wang, Y., Wang, J., Zhu, X., & Hansen, P. (2017). Take-over requests analysis in conditional automated driving and driver visual research under encountering road hazard of highway. In I. Nunes (Eds.), *Advances in human factors and systems interaction.* Applied Human Factors and Ergonomics 2017. Advances in Intelligent Systems and Computing (pp. 230–240). Cham: Springer.

2.4 An international crowdsourcing study into people's statements on fully automated driving

Corresponding article: Bazilinskyy, P., Kyriakidis, M., & De Winter, J. C. F. (2015). An international crowdsourcing study into people's statements on fully automated driving. Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE), Las Vegas, NV, 2534–2542.

2.4.1 Abstract

Fully automated driving can potentially provide enormous benefits to society. However, it has been unclear whether people will appreciate such far-reaching technology. This study investigated anonymous textual comments regarding fully automated driving, based on data extracted from three online surveys with 8,862 respondents from 112 countries. Initial filtering of comments with fewer than 15 characters resulted in 1,952 comments. The sample consisted primarily of males (74%) and had a mean age of 32.6 years. Next, we launched a crowdsourcing job and asked 69 workers to assign each of the 1,952 comments to at least one of 12 predefined categories, which included positive and negative attitude to automated driving, enjoyment in manual driving, concerns about trust, reliability of software, and readiness of road infrastructure. 46% of the comments were classified into the category 'no meaningful information about automated driving', leaving 792 comments for further analysis. 39% of the comments were classified as 'positive attitude towards automated driving' and 23% were classified as 'negative attitude towards automated driving'. In conclusion, the public opinion appears to be split, with a substantial number of respondents being positive and a significant number of respondents being negative towards fully automated driving.

2.4.2 Introduction

It is generally believed that fully automated driving (FAD), or 'level 5 automation' according to the SAE levels of driving automation (SAE, 2013), will be a common mode of transportation in the (far) future. Automated driving could have large positive influences on society in terms of safety and efficiency of road transport.

Automated driving is currently a much discussed topic in academic institutions (De Winter, Happee, Martens, & Stanton, 2014; Hoeger et al., 2008; Kato et al., 2002; Jamson, Merat, Carsten & Lai, 2011; Begg, 2014; Casley Jardim & Quartulli, 2013; KPMG, 2013), governmental bodies (European Commission, 2011; NHTSA, 2013), and industries (Volvo, 2014a; Volvo, 2014b; Davies, 2014; Walker, 2014; Daimler, 2014; Sommer, 2013). Recently, automated driving has also become a topic of great interest to the public (Casley Jardim & Quartulli, 2013; KPMG, 2013; Schoettle & Sivak, 2013; Howard & Dai,

2014). For example, one blog article on the lane changing capabilities of the Tesla S was read 27,842 times and received a relatively large number of 96 comments (as recorded on 12 December 2014) (Lavrinc, 2014). A particular comment on this blog illustrates that people have legitimate questions regarding the robustness of automated driving technology in demanding environmental conditions: “Does an inch-thick crust of mud and salt screw up the sensors’ ability to accurately measure the environment around the car?”

Although the topic of automated driving is widely discussed in public fora, little scientific knowledge is available regarding the international perspective on the foreseen radical change in society and the level of acceptance of this technology. The present study aimed to investigate the public opinion on FAD.

2.4.2.1 Collected comments on automated driving

During 2014, in our research group, three surveys were launched on the CrowdFlower online platform (www.crowdflower.com) to poll the public opinion on fully automated driving. In this paper, we analyse the textual comments obtained from these three surveys

Survey 1 (S1) “Research study about driving behavior” (De Winter, Kyriakidis, Dodou & Happee, 2015) is an innovative study that explored the use of the crowdsourcing service CrowdFlower in academia. The 15-item survey focused on respondents’ knowledge of automated driving systems and cross-national differences in traffic violations as measured with the Manchester Driver Behaviour Questionnaire (DBQ). In total 1,862 responses were obtained within 20 hours at a cost of $247. The 16th question in the survey “*Any comments?”* invited the respondents to give any comments related to the survey itself and to the topic of the questionnaire – automated driving.

A larger survey (Survey 2; S2) “Opinion on automated driving systems” (Kyriakidis, Happee & De Winter, 2014) investigated user acceptance, worries, and willingness to buy partially, highly, and fully automated vehicles. In total 5,000 responses from 109 countries (40 countries with at least 25 respondents) were collected. This study further investigated cross-national differences and assessed correlations with personal variables, such as age, gender, and personality traits as measured with a short version of the Big Five Inventory. The 63rd question “*Please provide any additional comments you may have about the survey*” asked respondents to provide any comments, including their thoughts on automated driving.

Finally, Survey 3 (S3) (Bazilinskyy & De Winter, 2015) examined user acceptance of auditory interfaces in modern cars and their willingness to be exposed to auditory feedback in highly and fully automated driving. This survey obtained 2,000 responses from 96 countries. The 31st question “*Please provide any suggestions which could help engineers to build safe and enjoyable automated cars*” was targeted specifically at receiving feedback on automated driving.

Comments received in these three surveys formed a large amount of text. Text is often analyzed manually by the researchers themselves, a process that can be time consuming and prone to investigator bias. Text mining is a more efficient approach for analyzing large quantities of lexical structures (Tan, 1999; Hotho, Nürnberger & Paaß, 2005; Berry, 2004). Statistical text mining techniques allow researchers to tag and annotate texts, establish distributions of word

frequencies, and extract underlying patterns. There are numerous examples of efficient and fast analyses of text with such approaches. For example, Twitter messages that have no more than 140 characters were processed to receive real-time information on distracted driving messages (Roberts & Lee, 2014). Text mining is commonly employed in the field of biomedical research (Tanabe et al., 1999; Cohen & Hersh, 2005) and is a widely used technique for analyzing web content (Cooley, Mobasher & Srivastava, 1997; Pal, Talwar & Mitra, 2002). Analysis of text with text mining techniques is often faster than manual analyses. However, it requires a deep understanding of the underlying tools. A novel approach of crowdsourcing the task of text analysis was employed in this study: we delegated the classification of comments to dozens of workers from all over the world.

2.4.2.2 Using CrowdFlower for classifying comments

A preliminary inspection of the comments led us to conclude that comments consisting of less than 15 characters contained no meaningful information. These comments (N = 5,884, 75%), including empty comments, were therefore removed. Accordingly, a total of 1,952 comments were left for further analysis. The sample consisted primarily of males (1,429 males, 513 females, 10 gender unknown) and had a mean age of 32.6 years (SD = 11.4, N = 1914 with available age data).

The 1,952 comments represented text of considerable size: 175,378 characters (M = 90, SD =101, N = 1,952), or, assuming that the average length of a word in the English language is 5.1 characters (WolframAlpha, 2014), about 34,388 words. We reasoned that a manual analysis of such a large number of comments would not be reasonable, since it would take a significant amount of person-hours of work. The option of outsourcing the task among colleagues was discarded, as we suspected it could have led to biased results. Thus, crowdsourcing the job appeared to be a good solution.

Table 1. Twelve categories used for classifying the comments.

Code name	Category	Description
NEGATV	Negative attitude towards automated driving	Statements that express general negativity towards automated driving.
MANUAL	Preference to manual driving (i.e. ability to choose manual driving)	Statements saying that manual driving would be preferred over automated driving. By manual driving we mean a present day situation where cars are controlled by humans.
SEMAUT	Preference to semi-automated driving (i.e. ability to choose manual driving)	Statements saying that manual driving would be preferred over automated driving because a driver wants to be in control of his own vehicle.
ENJOYM	Enjoyment in manual driving	Statements saying that manual driving would be preferred over automated driving because of the “joy of driving”.
CCOSTS	Concerns about costs	Statements that express concerns about the cost of automated driving.
CTRUST	Concerns about trust	Statements that refer to lack of trust for a vehicle that can drive on its own.

CSOFTW	Concerns about security of software (i.e. threats from hackers)	Statements that express concerns about misuse of the software of automated cars (such as threats from computer hackers).
CINFST	Concerns about readiness of infrastructure (i.e. unprepared roads)	Statements that express concerns that current modern roads are not prepared to support automated driving.
POSITV	Positive attitude towards automated driving	Statements that express general positivity towards automated driving.
VISION	Vision of a highly-automated vehicle	Statements regarding the vision of highly-automated driving.
NOMEAN	No meaningful information about automated driving	Statements that carry no meaningful information about automated driving.
OTHER	Other	All other statements.

The comments were categorized by means of a survey project launched on CrowdFlower. We outlined 12 categories for the classification of the comments. These 12 categories were created through a manual analysis of a random selection of 200 comments. Specifically, the categories were defined based on the frequencies of comments that could be assigned to particular categories. The categories encompassed positive and negative opinions towards automated driving, concerns about the different aspects of automated driving, and the public's vision of automated driving. Table 1 shows the established categories and provides short descriptions.

In our crowdsourcing project, we only allowed workers from English speaking countries. CrowdFlower provides the option to select up to 15 countries per project. Hence, our workers were from Antigua and Barbuda, Australia, Bahamas, Barbados, Belize, Canada, Dominica, Grenada, Ireland, Jamaica, New Zealand, Saint Lucia, Trinidad and Tobago, United Kingdom, and the United States. To assure sufficient quality of the categorization, the highest (third) level of performance of contributors was selected. That is, only the most highly ranked workers were invited to perform the categorization. A maximum of 200 randomly selected judgments per contributor and IP address were permitted. In total, 69 workers classified the comments. The total amount to be paid for the crowdsourced categorization of 1,952 comments was $120. Each comment was processed by at least five workers, while the workers were not allowed to review the same comment more than once. The workers received the comments in random order, and they were allowed to classify comments in more than one category.

To control the quality of data, we adopted a *threshold* when analyzing the data. The threshold defines the minimum number of workers that assigned a comment to a particular category for the categorization to be accepted as valid. The cases where *threshold* was equal to 1, 2, 3, 4 and 5 judgments were handled.

Furthermore, countrywide differences were analyzed at the national level by comparing the opinion of people on automated driving as a function of their country's income. Information on the Gross Domestic Product (GDP) per capita of countries involved in the surveys was extracted from the records of the World Bank (The World Bank, 2014). The values of GPD per capita of The Bahamas, Barbados, Taiwan, and United Arab Emirates were retrieved from the

International Monetary Fund (International Monetary Fund, 2013). One comment originated from the Palestinian Territories, and no information on the GDP per capita of that country could be found. Hence, that comment was excluded, leaving 1,951 comments for the cross-national analysis.

Finally, the authors selected three comments from each category as representative examples of the opinions of the respondents. Only comments that were written in English language, were clearly stated and easily interpretable, and were at least 50 characters long.

2.4.2.3 Results

In total, 11,760 reviews (or 'judgments' according to the terminology used by CrowdFlower) of comments were received. That is, on average, each comment was reviewed 6.02 times (*SD* = 1.52, *N* = 1,952). The responses were gathered between 19 November 2014 17:05 and 20 November 2014 00:48 (CET). The categorization job received an overall satisfaction rating of 4.5 out of 5.0. The respondents ranked the clarity of the instructions as 4.4 / 5.0, fairness of the questions as 4.2 / 5.0, easiness of the survey as 3.9 / 5.0, and the offered payment ($0.75 for categorization of 100 comments) as 4.2 / 5.0.

We first explored the effect of different values of the *threshold* parameter. If *threshold* equaled 1, no judgments were ignored; if *threshold* equaled 2, then 3,438 judgments were ignored; if *threshold* equaled 3, 4,792 judgments were ignored; if *threshold* equaled 4, 5,581 judgments were ignored; and finally, if *threshold* equaled 5, 6,186 judgments were ignored.

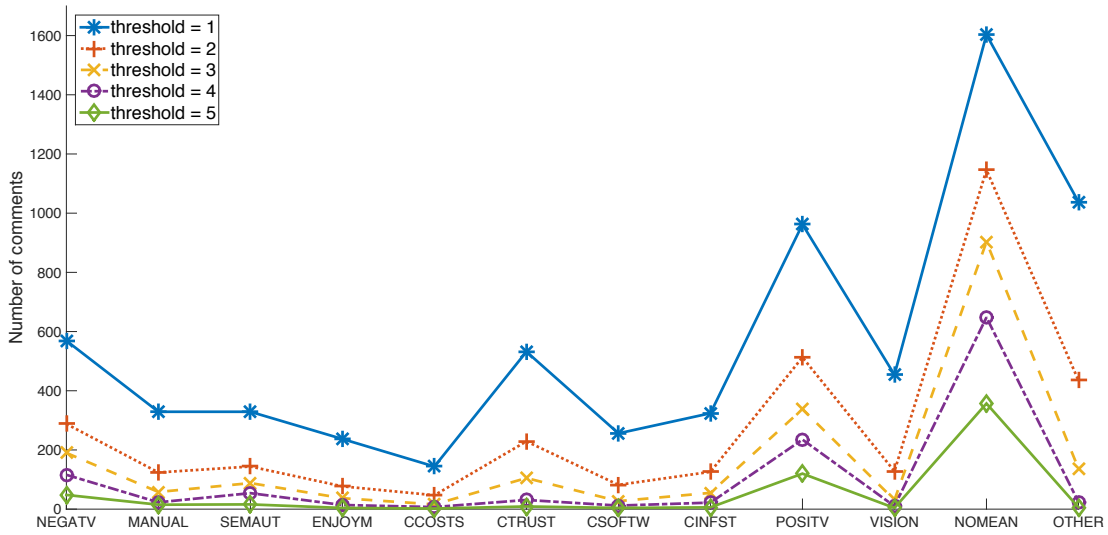


Figure 1. Classification of comments for the values of *threshold*. N = 1,952.

Figure 1 shows the numbers of accepted comments per category. According to our interpretation, the most valid and robust outcome was achieved when *threshold* equaled 3. When *threshold* equaled 3, 16% (*N* = 309) of the comments were classified as POSITV, while 9% (*N* = 185) were classified as NEGATV. In addition, 5% (*N* = 98) of the comments were marked as CTRUST, whereas 3% (*N* = 52) expressed CINFST. Furthermore, 4% (*N* = 83) of the comments expressed a preference for semi-automated driving (i.e., SEMAUT). Finally, 3% (*N* = 56) of the comments indicated that people would prefer manual

driving, that is, they were marked as MANUAL, and 2% (*N* = 36) of the comments were classified as ENJOYM.

The dominant category in the classification of comments was NOMEAN, with 46% (*N* = 903) of the comments classified into this category. These comments were seen as statements that carried no meaningful information about FAD and were excluded from the data set. Subsequently, 792 meaningful comments were left for further analysis.

Table 2. Numbers of comments classified into a category (diagonal) and into two categories (off-diagonal), with *threshold* = 3. (*N* = 1,952)

	NEGATV	MANUAL	SEMAUT	ENJOYM	CCOSTS	CTRUST	CSOFTW	CINFST	POSITV	VISION	OTHER
NEGATV	185										
MANUAL	32	56									
SEMAUT	6	0	83								
ENJOYM	11	14	7	36							
CCOSTS	0	0	1	0	14						
CTRUST	22	3	7	1	1	98					
CSOFTW	6	0	1	0	0	8	25				
CINFST	7	1	0	1	0	4	1	52			
POSITV	2	2	6	6	2	9	2	7	309		
VISION	0	0	1	0	0	1	0	2	15	28	
OTHER	1	0	0	0	0	1	0	2	1	1	72

Table 2 shows the numbers of the comments assigned to combinations of categories. Numbers on the diagonal indicate the total numbers of comments that were classified into each category. The off-diagonal values are subsets of the corresponding numbers on the diagonal of the table; they indicate how many comments were tagged into each pair of categories. A total of 185 of 792 comments (23%) were classified as negative towards automated driving (NEGATV), while 309 of 792 (39%) were classified as positive (POSITV).

A comparatively small number of comments were assigned to more than one category. Specifically, 32 comments were categorized as both MANUAL and NEGATV, which is an expected result, as respondents who have a negative attitude to automated driving also prefer to have manual control of a car. Furthermore, 22 respondents expressed a low level of trust towards automated driving and indicated a negative attitude towards automated driving (CTRUST & NEGATV). Seven people indicated that they would rather use a semi-automated vehicle while also expressing a low level of trust in automated driving (CTRUST & SEMAUT). However, 9 respondents had a positive attitude to automated driving while they also mentioned a low level of trust in automated driving (CTRUST & POSITV). Furthermore, 11 comments were categorized as ENJOYM and NEGATV, 14 comments as ENJOYM and MANUAL, and 7 comments as ENJOYM and SEMAUT.

Next, the comments were analyzed at the national level. Figure 2 shows the numbers of comments for three groups created based on the GDP per capita of corresponding countries: low-income countries (with GDP per capita between $694 and $3,900, N = 265), medium income countries (with GDP per capita between $4,403 and $21,035, N = 263), and high income countries (with GDP

per capita between $21,910 and $111,162, N = 264). The categories were created automatically by sorting the comments and splitting them into three equally sized groups. Fisher's exact test showed that people from high-income countries were more likely to be negative ($p < .001$) and less likely to be positive ($p = .001$) about automated driving than people from low income countries. People from high-income countries were also more concerned about software issues ($p = 0.048$). The other differences between respondents from low versus high income countries were not statistically significant ($p > .05$ for each of the other 8 categories).

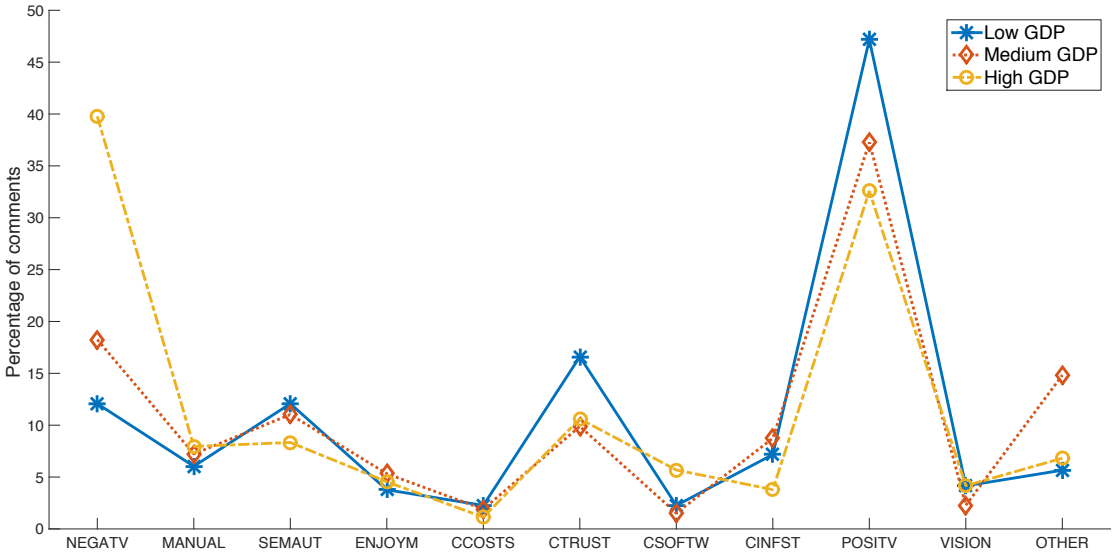


Figure 2. Percentage of comments assigned to the categories based on the GDP per capita of a country of origin of the comment: low, medium, and high. Groups of GDP per capita were created by sorting the comments based on the GDP per capita and splitting them into three equal groups. *N* = 792.

Finally, Table 3 introduces examples of comments for all categories. Code names indicate assigned categories.

Table 3. Examples of respondents' comments per category. The comments are not edited for grammar and spelling.

Code name	Comment
NEGATV1	The idea of fully automated cars scares me even more than other drivers.
NEGATV2	I will never set foot in a fully automated vehicle.
NEGATV3	I think this idea is unsafe and bizarre, actually.
MANUAL1	I will prefer to use manual driving because fully automated will make you lazy mentally.
MANUAL2	I prefer a manually driving 100%.
MANUAL3	I like manual driving.
SEMAUT1	I don't like the fully automated vehicles because i cant control it, a highly automated vehicle sounds much better.
SEMAUT2	I can't think of any as I don't like the idea of automated cars, as I prefer to have control.
SEMAUT3	Totally automated is giving up total control and some people may not like it.
ENJOYM1	A fully automated car will eliminate driving pleasure. There should be an option for manual driving.
ENJOYM2	I enjoy the manually driving too. Cause I feel I'm the driver :)

ENJOYM3	It is not a wise idea at all. If it really happens than there will be no race driver and no one will enjoy driving. No will ever say "Let's go for a long drive".
CCOSTS1	I hope the automated cars will be sold at a price that is not too expensive.
CCOSTS2	I would buy it for a good price and use it once it is on the market for a while and I'm sure that the system is safe. I would not be a pioneer on that, since safety is evolved.
CCOSTS3	Both price and quality accessible to everyone.
CTRUST1	I think this technology will take a long time to be really reliable and trusted.
CTRUST2	I trust my driving much more than I trust a computer system to do it for me. With the fully automated system, I would not want it because it would not allow for driver control if something major happened that a computer couldn't respond to.
CTRUST3	The cars should never be fully automated, at least the cars should maintain a certain degree of manual system as technology sometimes can fail.
CSOFTW1	I think manual control as a must. If anything malfunctioned or something like virus attack will be really dangerous without manual control. It will be like knight rider huh.
CSOFTW2	Well, my main concern with automated cars is the possibility of someone hacking the car and being able to take over the car. So I would think that security would be extensively tested to prevent such cases.
CSOFTW3	It’s a good survey to take but in my opinion fully automated will be a sure shot risk because computers are also justified as devil at bad hacking times.
CINFST1	The driving conditions in our country (country name) needs a lot of improvement.
CINFST2	In my country, the road infrastructure is very bad, and i think cars will have a tough time becoming automated.
POSITV1	I think the concept of Fully Automated Driving is very fascinating and it could be possible in the near future as technology develops and human beings advance.
POSITV2	Hopefully this wonderful technology will occur in my lifetime.
POSITV3	I hope this becomes available in the very near future.
VISION1	Since I cannot drive manually due to my bad vision, a fully automated car would be great for me, as long as I can see the instructions in the car and program it to get where I want to go, I should be able to get a licence for it.
VISION2	Opportunities to disabled.
VISION3	I am a disabled person and I have really bad eyes and limited field of view. I would be really happy if we had fully automated vehicles here in (name of country) so I wouldn't need second person to drive a car for me. It is sad though because it will take a lot of years for our country to introduce such vehicles on large scale.
NOMEAN1	I have no additional comments.
NOMEAN2	Don't judge me.
NOMEAN3	Comfy chair, spacious.
OTHER1	Correct sensors should be installed.
OTHER2	Don’t you have any other thing to do? Like finding a solution for global warming.
OTHER3	Correct sensors should be installed.

2.4.3 Discussion

In this study, free-response comments from three crowdsourced surveys involving 1,952 respondents were categorized by means of another project submitted to CrowdFlower. The decision to involve external people in the categorization was taken after determining that the classification of all comments would be cumbersome and time-consuming to do ourselves, and it could be biased. Our approach to the categorization of a large amount of text proved to be efficient and successful. Moreover, a threshold for accepting

categorization of comments was developed, and robust results were obtained when this variable was set to 3.

The main finding was that the public opinion appears to be split, with a significant number of respondents being positive (POSITV) and a significant number of respondents being negative (NEGATV) towards FAD. This result is consistent with a previous survey study (Kyriakidis, Happee & De Winter, 2014), which analyzed the public opinion on automated driving using five-point Likert items. A portion of the population does not appear to trust automated vehicles (CTRUST) and prefers to drive manually (MANUAL). A small number of comments were categorized into multiple categories. A dual categorization indicates obvious connections between categories. For example, 32 of 1,952 comments were categorized with both a negative attitude towards automated driving and a preference for manual control of cars in the future.

The comments were also analyzed at the national level, where they were grouped by GDP per capita of the respondents' country. The results revealed an association between income level and the number of comments per category. People from high-income countries were more likely to express a negative comment and less likely to express a positive comment about automated driving. In one of our previous surveys using five-point Likert items, we found that people from countries with a higher GDP were more concerned about automated vehicles transmitting data than people from low income countries (Kyriakidis, Happee & De Winter, 2014).

One of the categories presented to the workers of CrowdFlower for the categorization was "*preference to manual driving*". This category was created to indicate that the preference is given to manual driving over automated driving. However, it could also be understood as "I prefer automated driving to manual driving", which indicates that the preference is given to automated driving. Another issue was that the category "*vision*" could have been understood by some crowdworkers literally as 'eyesight' instead of 'imagination' or 'prospect' (as we intended). The examples in Table 3 (e.g., VISION1 & VISION3) reveal that FAD is preferred among people having bad eyesight or physical disabilities. Nonetheless, the workers indicated that instructions were well defined by ranking their clearness as 4.4 / 5.0.

A limitation of the present results is that CrowdFlower respondents are not representative of the entire population of stakeholders of future FAD vehicles. It is expected that such vehicles will initially be purchased by wealthy people, whereas jobs on CrowdFlower are performed mostly by people with relatively low income (Kyriakidis, Happee & De Winter, 2014). One recommendation is to launch more questionnaires, possibly using the 'classic' approach of asking people in person, in an attempt to target more diverse layers in the society. The questions that were asked in the surveys may be re-launched on CrowdFlower a few years in the future. This would create the opportunity to elucidate temporal trends in the public opinion.

2.4.4 References

Bazilinskyy, P., & De Winter, J. (2015). Auditory interfaces in automated driving: an international survey. *PeerJ Computer Science*, *1*, e13. https://doi.org/10.7717/peerj-cs.13

Begg, D. (2014). A 2050 vision for London: What are the implications of

driverless transport.
Berry, M. W. (2004). Survey of Text Mining: Clustering, Classification, and Retrieval. https://doi.org/10.1007/978-1-84800-046-9
Casley, S., Jardim, A., & Quartulli, A. (2013). A Study of Public Acceptance of Autonomous Cars. Worcester Polytechnic Institute.
Cohen, A. M., & Hersh, W.R. (2005). A survey of current work in biomedical text mining. *Briefings in bioinformatics*, *6*(1), 57–71.
Daimler. (2014). Mercedes-Benz sends autonomous automobiles onto the USA's most extensive testing ground. Available at http://www.daimler.com/dccom/0-5-658451-1-1745411-1-0-0-0-0-0-0-0-0-0-0-0-0-0-0.html
Davies, A. (2014). A system that any automaker can use to build self-driving cars. Available at http://www.wired.com/2014/11/delphi-automated-driving-system
De Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation research part F: traffic psychology and behaviour*, *27*, 196–217. https://doi.org/10.1016/j.trf.2014.06.016
De Winter, J. C. F., Kyriakidis, M., Dodou, D., & Happee, R. (2015). Using CrowdFlower to study the relationship between self-reported violations and traffic accidents. *Procedia Manufacturing*, *3*, 2518–2525.
European Commission. (2011). Roadmap to a single European transport area – towards a competitive and resource efficient transport system.
Hoeger, R., Amditis, A., Kunert, M., Hoess, A., Flemisch, F., Krueger, H.-P., Bartels, A., & Beutner, A. (2008). Highly automated vehicles for intelligent transport: HAVEit approach. Proceedings of *ITS World Congress,* NY, USA.
Hotho, A., Nürnberger, A., & Paaß, G. (2005). A Brief Survey of Text Mining In *Ldv Forum* (Vol. 20, No. 1, pp. 19–62).
Howard, D., & Dai, D. (2014). Corridors to the Future: Transportation and Technology - Opportunities and challenges presented by connected–automated vehicles. Proceedings of *Transportation Research Board 93rd Annual Meeting*.
International Monetary Fund. (2013). World Economic Outlook Database. Available at http://www.imf.org/external/pubs/ft/weo/2013/01/weodata/weoselgr.aspx
Jamson, A. H., Merat, N., Carsten, O., & Lai, F. C. H. (2011). Fully-automated driving: The road to future vehicles. Proceedings of *the Sixth International Driving Symposium*, 2–9.
Kato, S., Tsugawa, S., Tokuda, K., Matsui, T., & Fujii, H. (2002), Vehicle control algorithms for cooperative driving with automated vehicles and intervehicle communications. *IEEE Transactions on Intelligent Transportation Systems*, *3*(3), 155–161. https://doi.org/10.1109/TITS.2002.802929
KPMG. (2013). Self-driving cars: Are we ready?
Kyriakidis, M., Happee, R., & De Winter, J. C. F. (2014). Public opinion on automated driving: Results of an international questionnaire among 5,000 respondents. *Transportation research part F: traffic psychology and behaviour*, *32*, 127–140. https://doi.org/10.1016/j.trf.2015.04.014
Lavrinc, D. (2014). Tesla Model S Auto Pilot Does Lane Changes For You. Available at http://jalopnik.com/tesla-model-s-auto-pilot-does-lane-

changes-for-you-1644644652
NHTSA. (2013), Preliminary Statement of Policy Concerning Automated Vehicles.
Pal, S. K., Talwar, V., & Mitra, P. (2002). Web mining in soft computing framework: relevance, state of the art and future directions. *IEEE Transactions on Neural Networks*, *13*(5), 1163–1177.
Roberts, S.C., & Lee, J. D. (2014). Deciphering 140 Characters: Text Mining Tweets On #DriverDistraction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 2195–2199). Sage CA: Los Angeles, CA: SAGE Publications. doi:10.1177/1541931214581461.
SAE. (2013). SAE Levels of Driving Automation. Available at http://www.sae.org/misc/pdfs/automated_driving.pdf
Schoettle, B., & Sivak, M. (2013). A survey of public opinion about autonomous and self-driving vehicles in the U.S., the U.K., and Australia, Michigan, USA.
Sommer, K. (2013). Mobility Study 2013.
Tan, A.-H. (1999). Text mining: The state of the art and the challenges. Proceedings of *the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, 8, 65–70.
Tanabe, L., Scherf, U., Smith, L. H., Lee, J. K., Hunter, L., & Weinstein, J. N. (1999). MedMiner: an Internet text-mining tool for biomedical information, with application to gene expression profiling. *Biotechniques*, *27*,1210–1214.
The World Bank. (2014). GDP per capita. Available at http://data.worldbank.org/indicator/NY.GDP.PCAP.CD
Volvo. (2014a). Volvo Car Group's first self-driving Autopilot cars test on public roads around Gothenburg. Available at https://www.media.volvocars.com/global/en-gb/media/pressreleases/145619/volvo-car-groups-first-self-driving-autopilot-cars-test-on-public-roads-around-gothenburg
Volvo. (2014b). Volvo Cars conducts research into driver sensors in order to create cars that get to know their drivers. Available at https://www.media.volvocars.com/global/en-gb/media/pressreleases/140898/volvo-cars-conducts-research-into-driver-sensors-in-order-to-create-cars-that-get-to-know-their-driv
Walker, J. (2014). Jaguar Land Rover pledges to produce a self-drive car within ten years. Available at http://www.birminghampost.co.uk/business/manufacturing/jaguar-land-rover-pledges-produce-8148485
WolframAlpha. (2014). Average word length in English. Available at http://www.wolframalpha.com/input/?i=average+word+length+in+English

3 AUDITORY FEEDBACK FOR SUPPORTING TAKEOVER REQUESTS DURING HIGHLY AUTOMATED DRIVING

3.1 Crowdsourced measurement of reaction times to audiovisual stimuli with various degrees of asynchrony

Corresponding article: Bazilinskyy, P., & De Winter, J. C. F. (2018). Crowdsourced measurement of reaction times to audiovisual stimuli with various degrees of asynchrony. Human Factors.

3.1.1 Abstract

Objective: This study was designed to replicate past research concerning reaction times to audiovisual stimuli with different stimulus onset asynchrony (SOA) using a large sample of crowdsourcing respondents.

Background: Research has shown that reaction times are fastest when an auditory and a visual stimulus are presented simultaneously and that SOA causes an increase in reaction time, this increase being dependent on stimulus intensity. Research on audiovisual SOA has been conducted with small numbers of participants.

Method: Participants (N = 1,823) each performed 176 reaction time trials consisting of 29 SOA levels and three visual intensity levels, using CrowdFlower, with a compensation of US$0.20 per participant. Results were verified with a local Web-in-lab study (N = 34).

Results: The results replicated past research, with a *V* shape of mean reaction time as a function of SOA, the *V* shape being stronger for lower-intensity visual stimuli. The level of SOA affected mainly the right side of the reaction time distribution, whereas the fastest 5% was hardly affected. The variability of reaction times was higher for the crowdsourcing study than for the Web-in-lab study.

Conclusion: Crowdsourcing is a promising medium for reaction time research that involves small temporal differences in stimulus presentation. The observed effects of SOA can be explained by an independent-channels mechanism and also by some participants not perceiving the auditory or visual stimulus, hardware variability, misinterpretation of the task instructions, or lapses in attention.

Application: The obtained knowledge on the distribution of reaction times may benefit the design of warning systems.

3.1.2 Introduction

Reaction times are widely used to examine human information-processing mechanisms, such as in studies of cognitive ability (Der & Deary, 2006; Jensen,

2006), visual search (Wolfe, 1998), and memory (Baddeley & Ecob, 1973). In human factors science, reaction times are typically measured for applied purposes, for example, to quantify stimulus-response compatibility of human–machine interfaces (Chapanis & Lindenbaum, 1959; Fitts & Seeger, 1953) and the effectiveness of warning systems (Abe & Richardson, 2006). In the design of any warning system, it should be decided whether the warning signal is auditory, visual, vibrotactile, or multimodal. For example, in automated driving, a takeover warning can be a vibrotactile stimulus in the seat (Petermeijer, De Winter, & Bengler, 2016), an auditory signal (Merat & Jamson, 2009), a visual notification on the dashboard (Larsson, Johansson, Söderman, & Thompson, 2015), or a multimodal signal, such as an audiovisual alarm (e.g., Gold, Damböck, Lorenz, & Bengler, 2013) or a vibrotactile-auditory alarm (e.g., Petermeijer, Bazilinskyy, Bengler, & De Winter, 2017). The present study is concerned with a new method for large-scale research on reaction times to multimodal stimuli.

3.1.2.1 Previous research on the effect of Stimulus onset Asynchrony (SoA) on reaction times

It is well established that in simple reaction time tasks, multimodal feedback yields faster reaction times than unimodal feedback (Diederich & Colonius, 2004; Todd, 1912). However, the timing and intensity of the stimuli have an important effect on reaction times. Literature shows that average reaction times to bimodal stimuli are fastest when the onsets of the stimuli occur at the same moment, with the mean reaction time as a function of SOA exhibiting a V shape (e.g., Miller, 1986). This V shape is illustrated in Figure 1, showing results from our literature survey on reaction times to audiovisual stimuli as a function of SOA. Only studies that used equivalent task conditions were included in this figure (for additional relevant research on SOA, see Harrar, Harris, & Spence, 2017; Leone & McCourt, 2015; Van der Stoep, Spence, Nijboer, & Van der Stigchel, 2015). Each subfigure shows mean reaction times as a function of SOA, where a negative SOA value means that the onset of the visual stimulus occurred after the onset of the auditory stimulus. The middle and right subfigures concerned studies that focused on manipulating the intensity of the visual and auditory stimuli, as indicated with lowercase (v, a) and uppercase letters (V, A).

It can also be seen in Figure 1 that the degree with which reaction times increase as a function of SOA depends on the intensity of the stimuli (see also Miller & Ulrich, 2003). More specifically, if the visual stimulus is difficult to see, then participants are likely to respond to the auditory stimulus, and so the onset of the auditory stimulus will have a dominant effect on the mean reaction time. Conversely, if the stimulus is poorly audible, then the onset of the visual stimulus will determine the reaction time. These interactions between SOA and stimulus intensity were illustrated by Gondan, Götze, and Greenlee (2010; see Fig. 1 middle) and Leone and McCourt (2013; see Fig. 1 right). Thus, the relationship between mean reaction time and SOA is asymmetric (i.e., one side of the *V* shape is flatter than the other) when the auditory stimulus is weak and the visual stimulus is intense (i.e., aV in Fig. 1) or when the visual stimulus is weak and the auditory stimulus is intense (Av in Fig. 1).

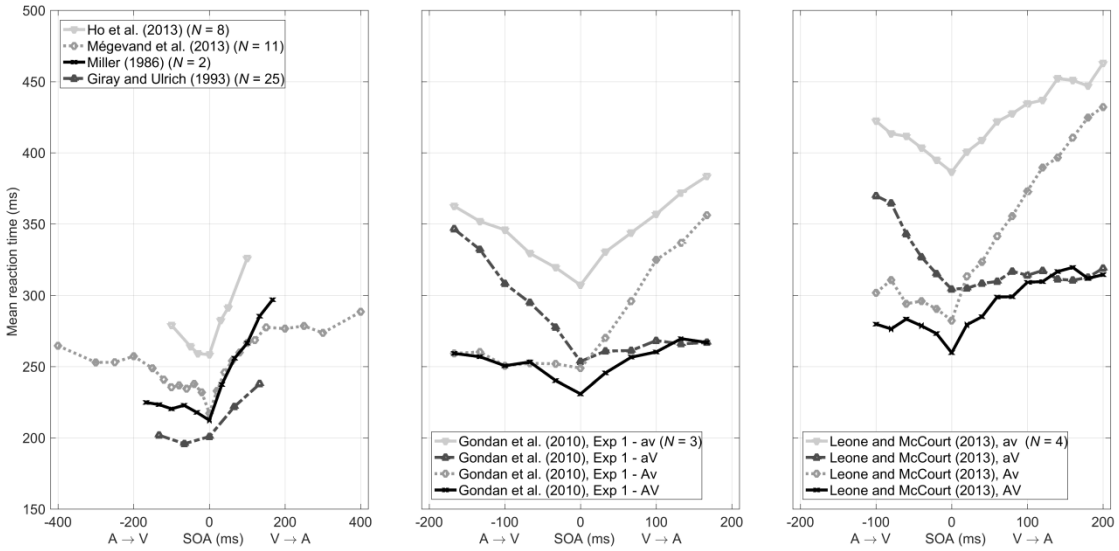


Figure 1. Mean reaction times from a selection of literature on stimulus onset asynchrony in audiovisual reaction time tasks. Left = four independent studies (Giray & Ulrich, 1993; Ho, Gray, & Spence, 2013; Mégevand, Molholm, Nayak, & Foxe, 2013; Miller, 1986); middle = study that manipulated stimulus intensity (Gondan, Götze, & Greenlee, 2010); right = study that also manipulated stimulus intensity (Leone & McCourt, 2013); a (A) = low- (high-) intensity auditory stimulus; v (V) = low- (high-) intensity visual stimulus.

Differences in the overall mean reaction time between the experiments shown in Figure 1 are of lesser interest, as these depend on factors such as the participants' level of experience, outlier removal, overall stimulus intensity, and hardware used during the experiment (e.g., Dodonova & Dodonov, 2013; Gondan & Minakata, 2016). For example, in Diederich and Colonius (2004) and Hershenson (1962), the mean reaction times to audiovisual stimuli were in the range of 135 to 155 ms (SOA 0–50 ms) and 98 to 144 ms (SOA 0–85 ms), respectively. These phenomenally fast reaction times may be explained by the fact that participants were highly trained, the use of intensive stimuli, and specialized hardware that records reaction times with little delay.

The research on the effect of audiovisual SOA has been conducted with small sample sizes (see the legends in Fig. 1) but typically with dozens of trials per stimulus condition. Accordingly, investigations of the distributions of reaction times have been performed within subjects rather than between subjects. For example, in Miller (1986), there were two participants who each completed 40 test blocks over a period of about 1 month, each block consisting of 130 test trials. It would be relevant to examine whether there exist individual differences in susceptibility to SOA effects. Within the human factors community, it has been emphasized that the design of warning systems should not be based on the mean reaction time but that slowly responding participants should be considered, too (Eriksson & Stanton, 2017; Wickens, 2001).

3.1.2.2 The potential of crowdsourcing for performing reaction time research

The Internet is a now well-established medium for experimental psychological research with large sample sizes (Fortenbaugh et al., 2015). Various studies have replicated classical psychological effects using online crowdsourcing methods (Crump, McDonnell, & Gureckis, 2013; Hilbig, 2016). For example, Barnhoorn, Haasnoot, Bocanegra, and Van Steenbergen (2015), using a JavaScript engine,

replicated three reaction time paradigms (Stroop task, attentional blink task, masked priming task) via crowdsourcing.

A number of studies suggest that online software and hardware can cause small delays compared to regular psychophysics methods. For example, De Leeuw and Motz (2016) found an additive reaction time delay of 25 ms, and no difference in variance, when using jsPsych (a library for creating behavioral experiments using JavaScript) running in Google Chrome as compared with MATLAB's Psychophysics Toolbox on the same laptop hardware. Reimers and Stewart (2016) described a limitation of

JavaScript, in that audio and visual stimuli scheduled to appear on a Web page at the same time are presented with a small temporal offset that can vary up to 40 ms, depending on the type of browser. Schubert, Murteira, Collins, and Lopes (2013) replicated the Stroop effect online and noted that the online software contributed to additional reaction time variance compared with a controlled lab study. According to simulations by Brand and Bradley (2012), the effect of technical variance (due to e.g., keyboards, CPU load, operating systems) is negligible compared with individual differences in reaction time, and they argued that "researchers' preconceptions concerning the unsuitability of web experiments for conducting research using response time as a dependent measure are misguided" (p. 350).

However, concerns have also been raised about the validity of online research, especially when small stimulus durations are involved. Semmelmann and Weigelt (2017) replicated well-known paradigms (e.g., Stroop test, flanker test) in three settings (classical lab, Web-in-lab, Web), with a total of 147 participants. Although the replication was successful, the mean reaction times in a simple reaction time task were 253 ms, 280 ms, and 318 ms, respectively, for the three settings. That is, the Web-in-lab method caused an additive delay, presumably due to the browser engine and JavaScript, whereas the Web method might be further affected by differences in participants' hardware and testing environments. Woods, Velasco, Levitan, Wan, and Spence (2015) provided a review of 10 online research platforms that can be used for measuring reaction times and concluded that the quality of online data is usually high. However, these authors also discussed sources of technical variability in online reaction time research, such as variability in screen brightness, screen color, and volume of auditory stimuli, and they argued that studies that require short stimulus presentation are not well suitable to online research. Similarly, Schubert et al. (2013) argued that "the smaller the effect, the more problematic the noise introduced by . . . online experimentation" (p. 10).

In summary, although the Internet can be used to replicate psychological phenomena concerning reaction times, online research is associated with additive bias and extra sources of variance compared to lab-based research, and it is unknown whether reaction times to small temporal manipulations can be replicated online.

3.1.2.3 Aim of this research

Given the knowledge gap, this study was designed to replicate previous research on the effect of SOA and stimulus intensity on audio-visual reaction times using a large sample of participants via crowdsourcing. A replication study of well-established previous findings may contribute to the understanding of the

validity of crowdsourcing and yield new knowledge on the relationship between SOA and reaction times.

Our analysis was concerned with investigating whether a *V* shape of mean reaction times (Fig. 1) replicates and whether a lower intensity of the visual stimulus causes the slope of the *V* shape to be steeper. As pointed out earlier, crowdsourcing research can yield a high variance in reaction times. Therefore, in addition to investigating mean reaction times, we examined individual differences in reaction times (percentiles and trial-to-trial correlations). Furthermore, we assessed the sources of variability in reaction times by examining learning curves, by comparing the results with a Web-in-lab study using the same software, and by studying the effects of experimental conditions, such as whether participants were using a keyboard or mobile phone or whether they were indoors or outdoors.

3.1.3 Method

This research complied with the American Psychological Association Code of Ethics and was approved by the Human Research Ethics Committee (HREC) at the Delft University of Technology. Informed consent was obtained from each participant.

3.1.3.1 Stimuli

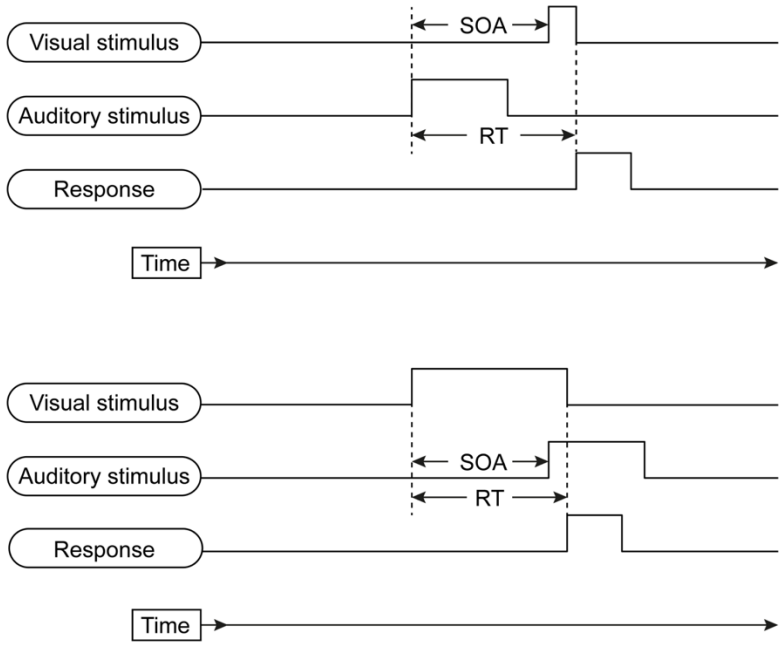


Figure 2. Timelines for events of a reaction time trial. The top figure concerns stimulus onset asynchrony (SOA) < 0 ms; the bottom figure concerns SOA > 0 ms. The auditory stimulus always had a duration of 210 ms, whereas the visual stimulus disappeared when the participant pressed the response key.

Participants were presented with audiovisual stimuli. The auditory stimuli were single 210-ms-long beeps of 1,840 Hz. The visual stimuli were red circles on a white background. A total of 29 SOA values were used: –1,000, –500, –300, –200, –100, –90, –80, –70, –60, –50, –40, –30, –20, –10, 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 500, and 1,000 ms. These 29 SOA values have a

range that is higher than the ranges of SOA values used in previous research (Fig. 1) while offering a higher temporal resolution (10 ms for SOA values between −100 and 100 ms). A negative SOA value means that the onset of the auditory stimulus occurred before the onset of the visual stimulus, and a positive SOA value means that the onset of the auditory stimulus occurred after the visual stimulus (as in Fig. 1). Figure 2 shows example timelines of reaction time trials with negative and positive SOA.

If the auditory stimulus was presented at the same moment or after the visual stimulus (SOA ≥ 0), then a .png file was presented together with a .wav file, with the time delay (SOA) encoded in the .wav file. If the visual stimulus was presented after the auditory stimulus (SOA < 0), then an animated .gif file was presented together with a .wav file. The animated .gif (via its graphics control extension) was a practical solution to encode a time delay of the onset of the visual stimulus. The rendering of the stimuli was powered by the jsPsych JavaScript library for running behavioral experiments online (De Leeuw, 2015).

Figure 3. Visual stimulus in the browser's full-screen mode (RGB 246-166-174, screen resolution: 1,920 × 1,080 pixels).

The red circles were uniform, had a diameter of 195 pixels, and had three levels of intensity: low, medium, or high (see Fig. 3 for an example). These three intensity levels (i.e., shades of red) were selected to be notably different but in such a way that the low-intensity stimulus was still clearly distinguishable from the white background, as we did not want that participants would fail to detect the visual stimuli. High-intensity stimuli were rendered on the screen as RGB 233-33-53. Low- and medium-intensity .png files were created using 40% and 70% transparency setting, respectively, which translates into rendered stimuli of RGB 246-166-174 and RGB 240-99-113, respectively. Low- and medium-intensity .gif files were RGB 251-211-215 and RGB 242-122-134, respectively. Because of the different RGB rendering of .png and .gif files, the reaction times to low- and medium-intensity stimuli between SOA < 0 and SOA ≥ 0 should not be directly compared. The auditory stimuli were always 210-ms beeps; they were not varied in intensity to keep the total number of conditions manageable.

3.1.3.2 Crowdsourcing experiment

Participants in the online experiment participated via the crowdsourcing platform CrowdFlower (https://www.crowdflower.com). Participants became aware of this research by logging into one of many channel websites (e.g., https://www.clixsense.com), where they would see our study in the list of other projects available for completion. We allowed contributors from all countries to participate. It was not permitted to complete the study more than once from the same worker ID. A payment of US$0.20 was offered for the completion of the experiment. A total of 2,000 participants completed the experiment, at a total cost of US$480. Our payment was assumed to be high enough to incentivize participants. Litman, Robinson, and Rosenzweig (2015) investigated the effect of payment for a 6-min task among crowdworkers from India and found that a payment of US$0.10 ("above-minimum-wage condition") yielded higher data quality than a payment of US$0.02 ("below-minimum-wage condition"), whereas a payment of US$1.00 ("far above the minimum wage") did not improve data quality compared with a payment of US$0.10.

Participants first answered a number of questionnaire items. At the beginning of the questionnaire, contact information of the researchers was provided, and the purpose of the upcoming study was described as "to determine reaction times for different types of visual and auditory signals." Participants were informed that the study would take approximately 8 min. The participants were also informed that they could contact the investigators to ask questions about the study and that they had to be at least 18 years old. Information about anonymity and voluntary participation was provided as well. The questionnaire started with the following questions:

- "Have you read and understood the above instructions?" ("Yes," "No")
- "What is your gender?" ("Male," "Female," "I prefer not to respond")
- "What is your age?" (positive integer)
- "In which type of place are you located now?" ("Indoor, dark"; "Indoor, dim light"; "Indoor, bright light"; "Outdoor, dark"; "Outdoor, dim light"; "Outdoor, bright light"; "Other"; "I prefer not to respond")
- "Which input device are you using now?" ("Laptop keyboard," "Desktop keyboard," "Tablet on-screen keyboard," "Mobile phone on-screen keyboard," "Other," "I prefer not to respond")

Several additional questions were asked about driving habits, which were not used in this study. The participants were then asked to leave the questionnaire by clicking on a link that opened a Web page with the reaction time task. Participants were presented with instructions on how to complete the given task:

In this experiment, you will hear sounds and see red circles. Please make sure that your audio is on and set your screen to bright. You need to press "F" after hearing a sound OR seeing a red circle (whichever comes first) as fast as possible. Your reaction times will be recorded. After each group of 25 stimuli you will be able to take a small break. Please press any key to start with the first stimulus.

The participants had to respond to 88 different stimuli in random order. Each stimulus was repeated twice, yielding 176 stimuli for each participant (i.e.,

29 SOA values × 3 visual intensity levels × 2 repetitions + 2 repetitions of an audio-only stimulus). There was no upper limit to the reaction times; the next stimulus trial was loaded after the participants pressed the *F* button. The stimuli were presented in six batches of 25 and one last batch of 26. After a batch, participants were shown the following text: "You have now completed 25 [50, 75, 100, 125, 150] stimuli out of 176. When ready press 'C' to proceed to the next batch." An analysis of the elapsed times showed that participants took a median time of 9 s to press *C* after the first batch and a median time of 4 s to press *C* after the sixth batch.

After pressing the *F* key, a new stimulus was presented after a uniform random delay between 1,000 and 3,299 ms, in agreement with Diederich and Colonius (2004). The images and sounds were preloaded to eliminate unwanted delays between the stimuli. Data for each participant were saved in a database after the 176th stimulus. Analyses of the distribution of reaction times per participant showed that the temporal resolution of the reaction time measurements (i.e., the minimum difference that could be detected) differed between participants: For the majority of participants (88%), the temporal resolution was between 2.6 and 3.0 ms. For 6% of the participants, the temporal resolution was between 3 ms and 12 ms, whereas 4% of participants had a temporal resolution of 42.7 ms.

These differences in temporal resolution may be due to different platforms and browsers used by the participants.

At the end of the experiment, participants were shown a unique code. Participants were asked to note down this code and return to the Web page of the questionnaire. They were required to enter the code on the questionnaire as proof that they completed the experiment and to receive their remuneration.

3.1.3.3 Web-in-lab experiment

To verify the results of the crowdsourcing experiment in controlled experimental conditions, we launched the same task in a laboratory setting. We collected responses from 42 participants from the university community. All participants completed the task on the same MacBook Air (13-in. screen, 8 GB memory, Intel Core I7 processor, two cores) laptop behind a table in a standard office room of about 3 × 3 m. The blinds were closed to control the lighting conditions; the ceiling lights (fluorescent lamps) were always on. The volume of the laptop was set to 60% (corresponding to a measured sound intensity of 60–65 dBA), and the brightness of the display was 100%. The experimenter started up the task and left the room so that the participant completed the task while being alone in the room. The temporal resolution of the reaction times of the Web-in-lab experiment was 5.8 ms. Participants of the Web-in-lab experiment did not receive remuneration because it is common practice at our institution to not pay participants for a short-lasting experiment.

3.1.3.4 Handling of reaction time outliers and statistical testing

Reaction times less than 0 ms were removed from the analysis, whereas reaction times greater than 1,500 ms were set equal to 1,500 ms. Using this so-called winsorization method, extremely slow reaction times (>1,500 ms) were retained in the analysis (as recommended by Gondan & Minakata, 2016), while limiting the skewness and kurtosis of the reaction time distribution (Ratcliff, 1993).

Differences between participants' conditions (e.g., input device) were compared using an unequal-variance *t* test (Welch, 1947) after performing an inverse transformation of the reaction times (Ratcliff, 1993). Effect sizes were assessed using Cohen's *d* of the inverse reaction times.

3.1.4 Results

The responses were collected between March 3, 2017, 12:42 and March 4, 2017, 17:30 (GMT). Two hundred twenty-four participants completed an optional user satisfaction survey offered by CrowdFlower. The study received an overall satisfaction score of 4.4 out of 5.0 on a scale from 1 (*very dissatisfied*) to 5 (*very satisfied*). The mean response to the question "How clear were the task instructions and interface?" was 4.6 on a scale from 1 (*very unclear*) to 5 (*very clear*), and the mean response to "How would you rate the pay for this task relative to other tasks you've completed?" was 4.2 on a scale from 1 (*much worse*) to 5 (*much better*).

3.1.4.1 Participant filtering and participant characteristics

Out of 2,000 participants, 177 were removed during data filtering. These were participants for whom no reaction time data were available due to a server/recording error (n = 119), participants with more than 20% negative reaction times (due to pressing the response key before the stimulus was presented; n = 55), or participants who answered "no" to the question whether they had read and understood the task instructions (n = 9). The 20% threshold was assumed to discriminate between participants with genuine anticipatory reaction times (i.e., accidentally pressing the *F* key too early) and participants cheating the system by repeatedly pressing *F*.

In the group of the remaining 1,823 participants, 1,283 were male, 533 were female, and 7 did not specify their gender. Three participants reported an unrealistic age or an age that was not in agreement with the task instructions (3, 5, and 17 years). Because these ages could be the result of a basic typographical error, and because these three participants did complete the task, they were retained in the analysis. The participants' mean age for the 1,820 participants of 18 years and older was 33.9 years (SD = 10.1, min = 18, max = 71).

The participants were from 83 different countries, with 22 countries having 25 or more respondents and four countries (Spain, Russia, Serbia, Venezuela) having more than 100 respondents.

3.1.4.2 Learning curve

Figure 4 shows that the mean reaction times decreased with trial number, that is, the participants showed faster reaction times as the experiment progressed. The spikes in the graph represent the trials that directly followed the breaks after each 25th stimulus. We removed Trials 1 through 5, 26, 51, 76, 101, 126, and 151 from the remaining analysis (except the correlations among trials), because these trials may be invalid as it is likely that some participants were still learning the basics of the task or pressed an incorrect key during these trials.

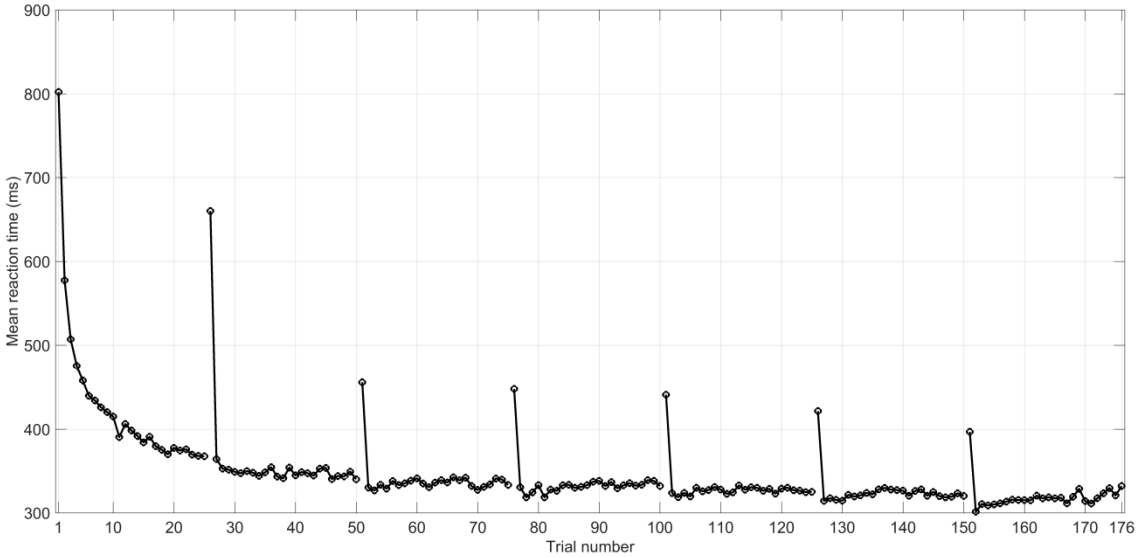


Figure 4. Mean reaction times versus trial number in the crowdsourcing study (N = 1,823). Each data point represents the mean across approximately 1,815 trials (i.e., 1,823 participants minus excluded responses).

3.1.4.3 Effects of SoA and stimulus intensity on reaction time

Figure 5 shows the mean reaction times as a function of SOA. Note that the results for SOA = −10 ms are not shown in the figures because the animated .gif files showed a delay of 100 ms when programmed with a delay of 10 ms. It can be seen that the lowest reaction times were obtained when the SOA was 0 ms. Furthermore, the visually delayed stimuli (i.e., SOA < 0 ms) yielded a mean reaction time that was about 43 ms higher than the auditorily delayed stimuli (i.e., SOA > 0 ms). It can also be seen that the low-intensity visual stimuli were associated with a stronger increase of the mean reaction time for increasing SOA than the high-intensity visual stimuli, which is consistent with the literature presented in Figure 1.

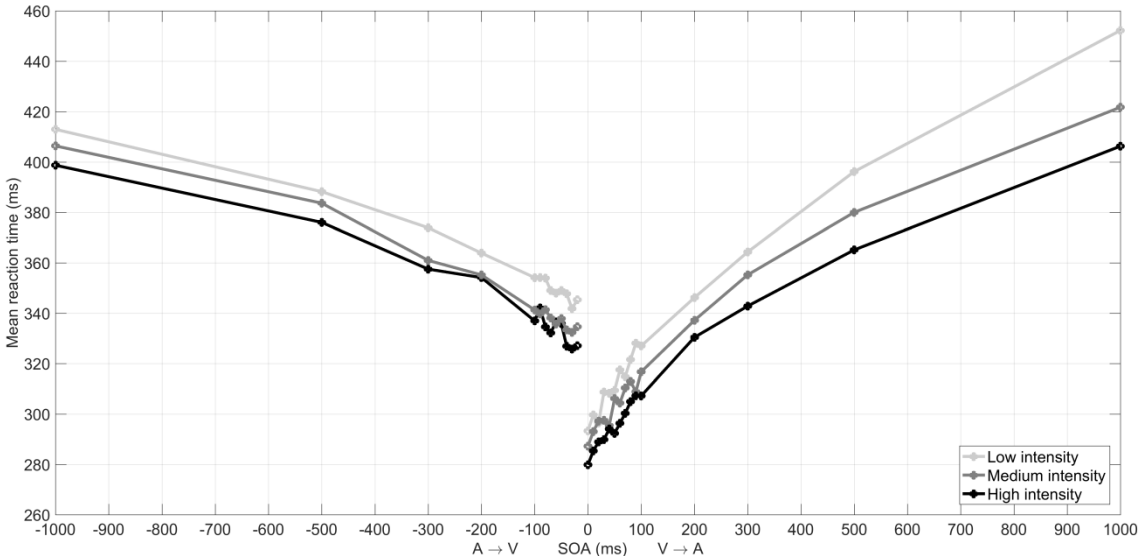


Figure 5. Mean reaction times for 28 levels of stimulus onset asynchrony (SOA) and three levels of visual intensity in the crowdsourcing study. Each data point represents the mean across approximately 3,400 trials (i.e., 1,823 participants × 2 trials per participant minus excluded responses).

Individual differences were assessed using percentiles of the observed reaction times, from low (i.e., fast reactions) to high (i.e., slow reactions). Figure 6 shows that the lowest reaction times were hardly affected by SOA, whereas the 95th percentile is strongly sensitive to SOA. In other words, the changes in mean reaction time observed in Figure 5 can be largely attributed to differences in the right tail of the reaction time distribution.

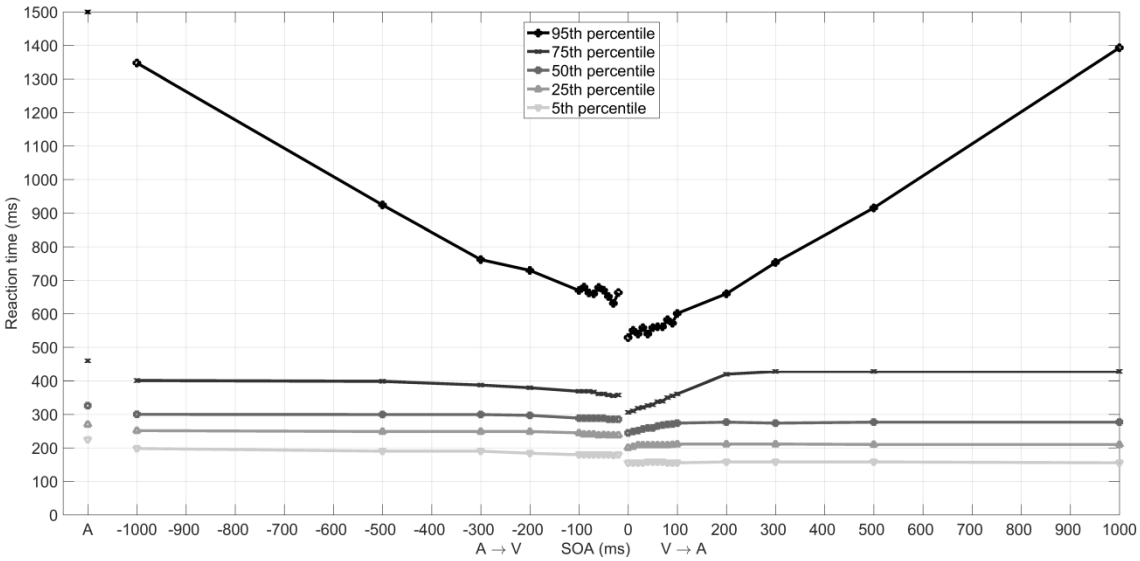


Figure 6. Percentiles of the reaction times for 28 levels of stimulus onset asynchrony (SOA) in the crowdsourcing study. Each data point is based on approximately 10,200 trials (i.e., 1,823 participants × 6 trials per participant minus excluded responses). The auditory-only trial (A) is based on 3,397 trials (i.e., 1,823 participants × 2 trials per participant minus 249 excluded responses).

3.1.4.4 Effects of experimental conditions on reaction time

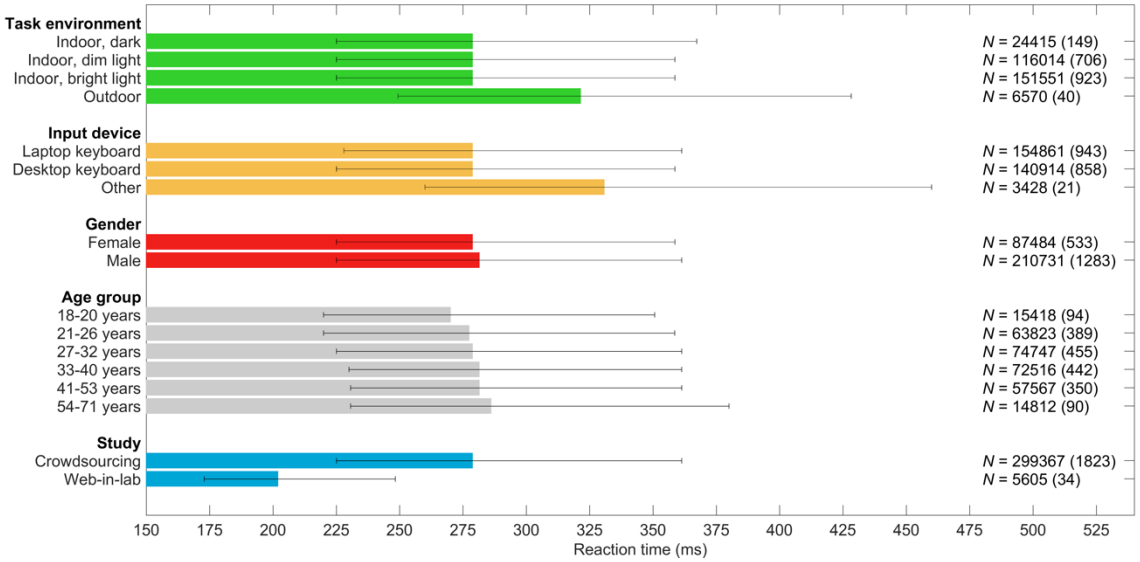


Figure 7. Median reaction time at the level of trials, per task environment, input device, gender, and age group for the crowdsourcing study and for the crowdsourcing versus the Web-in-lab study. The error bars denote the 25th and 75th percentiles. Also listed is the number of trials with the number of participants in parentheses. *Outdoor* refers to "Outdoor, dark"; "Outdoor, dim light"; and "Outdoor, bright light" combined. *Other* refers to "Tablet on-screen keyboard," "Mobile phone on-screen keyboard," and

"Other" combined. Ages of 20, 26, 32, 40, 53, and 71 years are the 5th, 25th, 50th, 75th, 95th, and 100th percentiles of participants' ages, respectively.

Figure 7 shows that indoor lighting condition did not have a large impact on the mean reaction times. A Welch's test showed no significant difference between dark and bright indoor light, $t(189.3) = 0.51$, $p = .608$, $d = 0.05$. However, completing the task outdoors was associated with significantly higher reaction time than completing the task indoors, $t(40.7) = 3.32$, $p = .002$, $d = 0.55$.

Figure 7 also shows that participants who completed the task with a laptop or desktop keyboard had faster reaction times than participants who used other input devices (e.g., tablets or mobile phones), $t(20.4) = 2.73$, $p = .013$, $d = 0.62$. There were no statistically significant gender differences, $t(1020.9) = 0.39$, $p = .697$, $d = 0.02$, nor age differences in reaction time (Spearman's correlation between age and mean reaction time: $\rho = 0.03$, $N = 1{,}820$).

3.1.4.5 Trial-to-trial correlations (Stability) of reaction times

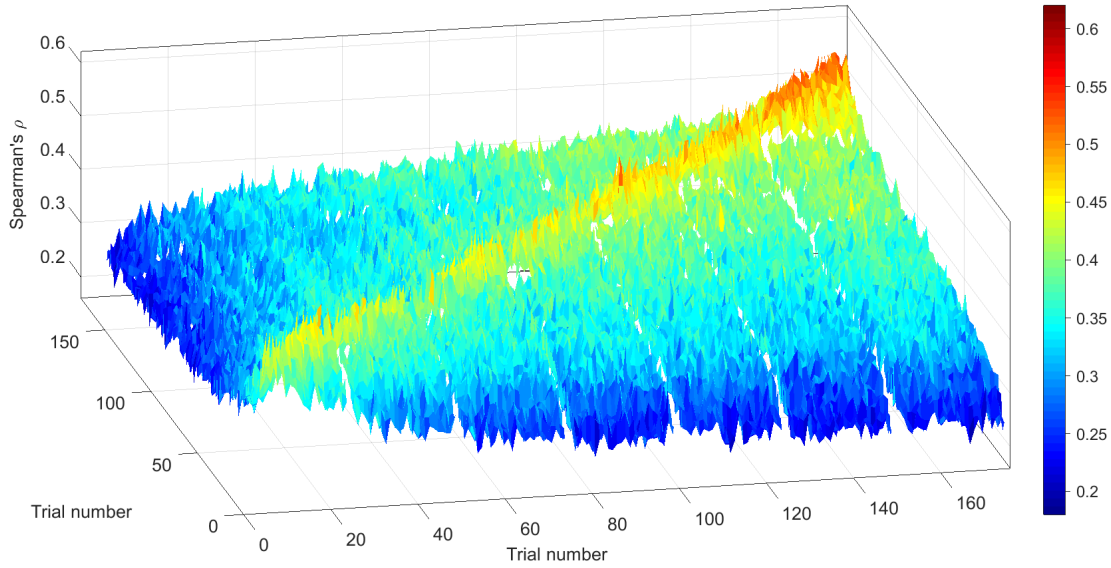


Figure 8. Heat map of Spearman rank-order correlations of crowdsourced participants' reaction times between Trials 1 and 176.

Finally, we calculated trial-to-trial correlations to obtain an indication of the stability of participants' reaction times. Figure 8 shows a Spearman correlation matrix among the reaction times per trial number for the crowdsourced participants. A high correlation between a pair of trials means that participants' reaction times had a similar rank ordering in these two trials, whereas a correlation of zero would be expected if participants were not consistent at all. A clear simplex pattern can be seen, with temporally adjacent trials showing higher correlations than temporally disparate trials (see also Ackerman, 1987). It can also be seen that reaction times stabilized (i.e., higher correlations) in later trials.

3.1.4.6 Reaction times from the web-in-lab experiment compared with the crowdsourcing experiment

In the laboratory setting, we retained responses from 34 participants, obtained between March 7, 2017, 10:57, and March 10, 2017, 13:13 (GMT). We removed

four participants with incomplete data and four participants who were involved in pilot tests conducted during the design of the study. The participants were six females and 28 males, having a mean age of 27.1 years (*SD* = 6.6 years, min = 18, max = 56). The reaction times were processed identically to the crowdsourcing experiments.

The results in Figure 7 show substantial differences between the international crowdsourcing method and the local lab method, $t(34.8) = 10.68$, $p < .001$, $d = 1.57$. The Web-in-lab method featured a lower mean reaction time and lower variability of reaction time (Fig. 9).

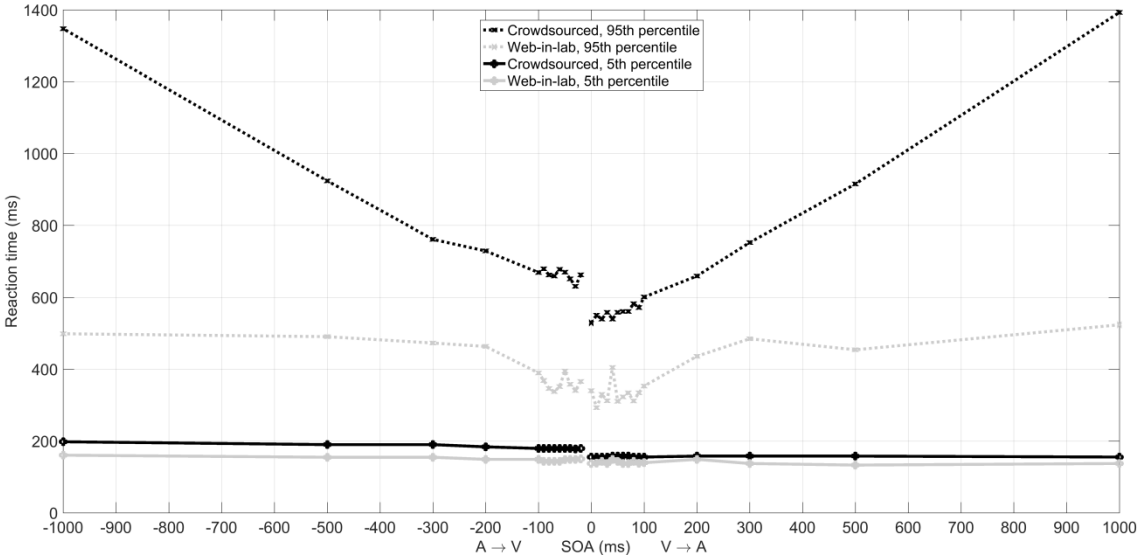


Figure 9. The 5th and 95th percentiles for 28 levels of stimulus onset asynchrony (SOA) in the crowdsourcing and Web-in-lab studies. Each data point of the crowdsourcing study is based on approximately 10,200 trials (i.e., 1,823 participants × 6 trials per participant minus excluded responses). Each data point of the Web-in-lab study is based on approximately 191 trials (34 participants × 6 trials per participant minus excluded responses).

3.1.5 Discussion

3.1.5.1 Replicated effects

In this study we aimed to replicate published research regarding the effects of audiovisual SOA and visual stimulus intensity on reaction times with a large sample of crowdsourced participants and to examine sources of variability of mean reaction times (e.g., learning curves, task conditions, comparison with Web-in-lab study).

Our findings replicated the *V* shape as observed in past research, with the mean reaction time being fastest when SOA = 0 ms and increasing monotonically both with increasing and decreasing SOA. The effect of stimulus intensity was also replicated, as evidenced by the higher reaction times for visual stimuli of lower intensity, as well as by the relatively steep slope of mean reaction times for low-intensity visual stimuli when the auditory stimulus was presented after the visual one (SOA > 0). This steep slope could also be seen for low-intensity visual stimuli in Figure 1 (Av condition).

Crowdsourcing allows researchers to access a large pool of participants, thereby yielding high statistical power. This can be illustrated with a post hoc

power analysis: For a false-positive rate of 1%, a sample size of 1,823, and an effect size for a pair of conditions (*dz*) of 0.109 (calculated from a mean difference of 20 ms, an observed *SD* across participants of 179 ms, and an observed correlation between the two groups of 0.47), the achieved statistical power is 98.0% (Faul, Erdfelder, Lang, & Buchner, 2007). The results in the figures allowed for a reliable assessment of experimental effects and individual differences results (effect of SOA, stimulus intensity, learning curves, percentiles).

3.1.5.2 Effects of experimental conditions and comparison between the crowdsourcing experiment and the web-in-lab experiment

Although the expected effects were clearly replicated, there were substantial differences in reaction times between the crowdsourcing study and the Web-in-lab study. The differences between the two methods may be because the Web-in-lab participants used the same high-quality laptop, which displayed the stimuli with the same intensity, whereas it is plausible that at least some crowdsourcing participants had poor or malfunctioning hardware or did not have their audio turned on despite the task instructions. Some of the crowdsourcing participants completed the task outdoors, which was associated with slower reaction times, possibly due to poor lighting conditions or distractive elements in the environment. Also, crowdsourcing participants who used a handheld device had a higher mean reaction time than participants who used a laptop or PC, which may be because the former involves hardware delays or may be hard to use if one's task is to provide input as quickly as possible. Furthermore, it is possible that the lab participants were concentrated and motivated to perform well, whereas the crowdsourced participants may have taken the task less seriously because they were anonymous. Previous research shows that IQ and reaction time share a negative correlation (Jensen, 2006; Madison, Mosing, Verweij, Pedersen, & Ullén, 2016). The lab participants, who were mostly students at a technical university, may have faster reaction times than the typical international crowdsourcing participant.

We did not find a statistically significant correlation between the mean reaction time and the mean age of the crowdsourced participants. This lack of correlation may be because the oldest participant in our study was 71 years old, whereas simple reaction times increase with age especially for people above 70 years old (Der & Deary, 2006). It is also possible that the relationship between age and reaction time is confounded because CrowdFlower participants from lower-income countries tend to be younger (De Winter & Dodou, 2016).

Another source of difference between the crowdsourcing and Web-in-lab study may concern differences in understanding of the task instructions. In previous CrowdFlower research, we found that participants from English-speaking countries (De Winter, Kyriakidis, Dodou, & Happee, 2015) and participants from countries with a higher gross domestic product (GDP) per capita (De Winter & Dodou, 2016) took less time to complete a questionnaire, which may have been due to difficulty in processing English-language text. Similarly, in a supplementary analysis of the present study, we found that participants from countries with a higher GDP per capita had a lower median time to complete the experiment, including the questionnaire (Spearman's $\rho = -0.67$, $p < .001$, based on 22 countries with 25 or more respondents), and had a faster mean reaction time as well (Spearman's $\rho = -0.36$, $p = .101$). In

summary, national differences may be a source of heterogeneity in the crowdsourcing study.

In our analysis, 55 of 2,000 participants were excluded due to negative reaction times. We aimed to show the variability of reaction times and therefore did not exclude slow-responding participants. However, others who use crowdsourcing and aim for clean data could opt for applying stricter screening criteria.

3.1.5.3 Learning curve and trial-to-trial correlations

The first trials were associated with slower reaction times as the participants needed time to get used to the system. Also, the participants showed increased reaction times after the breaks, which is presumably because some participants did not have their finger on the keyboard yet or initially pressed an incorrect key. That is, participants had to press *C* to proceed to the next batch of trials but had to press *F* after each trial, which may result in initial confusion. We also found that performance became more stable (i.e., higher between-trial correlations) as the experiment progressed. This increase of stability may be caused by the fact that participants learned the nature of the task and entered the autonomous phase of skill learning, in which performance is less susceptible to task-irrelevant distractions (Fitts & Posner, 1967).

3.1.5.4 Individual differences and reinterpretation of the Effects of SoA on reaction times

We found that SOA hardly affected the fastest reaction times, but it did have a substantial effect on the slowest (e.g., 95th percentile) of the reactions. That is, the hypothesized *V* shape of mean reaction time as a function of SOA was evident only in the mean reaction time and the higher percentiles of reaction time but was hardly evident from the 5th, 25th, and 50th percentiles of reaction time. These observations suggest that reductions in mean reaction times caused by simultaneous multimodal feedback are not necessarily due to multisensory neural integration, in which auditory and visual information is summed or combined in the central nervous system or at the level of individual neurons (Stein & Stanford, 2008). Our findings can be explained using an independent-channels mechanism where the visual and auditory channels operate in parallel (see Nickerson, 1973, for a review). That is, our results can be explained by the notion that participants sometimes do not attend to the auditory or visual stimulus. For example, a participant may be temporarily blinded due to an eye blink (typically lasting 150 ms; Wang, Toor, Gautam, & Henson, 2011), as a result of which he or she is more likely to react to the auditory stimulus. Similarly, a participant may have a lapse in hearing (e.g., due to an internal distraction or masking due to external noise), as a result of which he or she is more likely to react to the visual stimulus. Future research could use eye tracking and neurophysiological measures to investigate how reaction times depend on eye blinks and lapses in attention.

Participants were not prescreened based on their hearing or visual disabilities or other criteria that could affect their ability to complete this task. The variability in the right tail of the reaction time distribution may also be caused by individual differences in the understanding of the task instructions (i.e., to respond to the first of the two stimuli), in sensory ability, and in computer

hardware. People with a hearing disability or with malfunctioning speakers, for example, by definition have to react to the visual stimuli. The auditory-only stimulus caused a relatively high proportion of delayed responses (≥1,500 ms), which suggests that a portion of participants were "waiting" for the visual stimulus to arrive or did not have their sound enabled. More generally, our findings suggest that warning signals should be audiovisual rather audio only or visual only and that the visual and auditory warning should be presented simultaneously (SOA = 0 ms), as was done in an automated driving study by Petermeijer et al. (2017), for example.

3.1.5.5 Limitations of crowdsourcing regarding temporal resolution and timing of audiovisual stimuli

Based on their review, Woods et al. (2015) argued that "only subset of studies, specifically those requiring short stimulus presentation, are not so well suited to online research" (p. 15). We indeed did have some technical problems in the presentation of the stimuli. First, we observed a limited temporal resolution of the reaction time measurements in the crowdsourcing experiment, being 2.6 to 3 ms for 88% of the participants but 42.7 ms for 4% of the participants. Second, the animated .gifs do not render properly for delays of 10 ms (i.e., SOA = −10 ms), a known issue in computer graphics (Karonen, 2012). Also, the .gif files were associated with an average additive delay of about 43 ms. This additive delay was not observed in the lab study and was hardly present among the faster responses. Thus, it is possible that the 43-ms delay was caused by certain browsers not displaying the animated .gif files properly. Despite the problems observed with animated .gif files, differences in reaction times could be detected even for auditory and visual delays of 10-ms increments, which is noteworthy when considering that a typical screen refresh rate is only 17 ms (see Woods et al., 2015, for further discussion). In summary, we obtained credible experimental effects despite imperfect control of the SOA and despite a limited temporal resolution of the measurements. The robustness of reaction times to noise and temporal resolution is in agreement with simulations by Ulrich and Giray (1989) and Reimers and Stewart (2015). Authors of future research could extend our approach by varying not only the intensity of visual stimuli but also the intensity of the auditory stimuli.

3.1.6 Conclusions

We conclude that crowdsourcing may allow for large-scale reaction time research, at the expense of a lack of control of the test environment. For example, screen brightness and rendering problems may affect the perception of visual stimuli, whereas hardware volume level can affect the perception of auditory stimuli. The expected effects of SOA were replicated despite the variable test environment, which indicates that crowdsourcing is a powerful tool in reaction time research.

3.1.7 Supplemental material

Supplemental material is available at https://doi.org/10.4121/uuid:673c9bbc-bf17-42fa-a23a-3d716e141b1f

3.1.8 References

Abe, G., & Richardson, J. (2006). Alarm timing, trust and driver expectation for forward collision warning systems. *Applied Ergonomics*, *37*, 577–586. https://doi.org/10.1016/j.apergo.2005.11.001

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, *102*, 3–27. https://doi.org/10.1037/0033-2909.102.1.3

Baddeley, A. D., & Ecob, R. J. (1973). Reaction time and short-term memory: Implications of repetition effects for the high-speed exhaustive scan hypothesis. *Quarterly Journal of Experimental Psychology*, *25*, 229–240. https://doi.org/10.1080/14640747308400342

Barnhoorn, J. S., Haasnoot, E., Bocanegra, B. R., & Van Steenbergen, H. (2015). QRTEngine: An easy solution for running online reaction time experiments using Qualtrics. *Behavior Research Methods*, *47*, 918–929. https://doi.org/10.3758/s13428-014-0530-7

Brand, A., & Bradley, M. T. (2012). Assessing the effects of technical variance on the statistical outcomes of web experiments measuring response times. *Social Science Computer Review*, *30*, 350–357. https://doi.org/10.1177/0894439311415604

Chapanis, A., & Lindenbaum, L. E. (1959). A reaction time study of four control-display linkages. *Human Factors*, *1*, 1–7. https://doi.org/10.1177/001872085900100401

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE*, *8*, e57410. https://doi.org/10.1371/journal.pone.0057410

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*, 1–12. https://doi.org/doi:10.3758/s13428-014-0458-y

De Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, *48*, 1–12. https://doi.org/10.3758/s13428-015-0567-2

Der, G., & Deary, I. J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey. *Psychology and Aging*, *21*, 62–73. https://doi.org/10.1037/0882-7974.21.1.62

De Winter, J. C. F., & Dodou, D. (2016). National correlates of self-reported traffic violations across 41 countries. *Personality and Individual Differences*, *98*, 145–152. https://doi.org/10.1016/j.paid.2016.03.091

De Winter, J. C. F., Kyriakidis, M., Dodou, D., & Happee, R. (2015). Using CrowdFlower to study the relationship between self-reported violations and traffic accidents. Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics (AHFE), Las Vegas, NV*, 2518–2525. https://doi.org/10.1016/j.promfg.2015.07.514

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*, 1388–1404. https://doi.org/10.3758/BF03195006

Dodonova, Y. A., & Dodonov, Y. S. (2013). Is there any evidence of historical

slowing of reaction time? No, unless we compare apples and oranges. *Intelligence*, *41*, 674–687. https://doi.org/10.1016/j.intell.2013.09.001

Eriksson, A., & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors*, *59*, 689–705. https://doi.org/10.1177/0018720816685832

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, CA: Brooks/Cole.

Fitts, P. M., & Seeger, C. M. (1953). SR compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, *46*, 199–210. https://doi.org/10.1037/h0062827

Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman, M. (2015). Sustained attention across the life span in a sample of 10,000: Dissociating ability and strategy. *Psychological Science*, *26*, 1497–1510. https://doi.org/10.1177/0956797615594896

Giray, M., & Ulrich, R. (1993). Motor coactivation revealed by response force in divided and focused attention. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 1278–1291. http://dx.doi.org/10.1037/0096-1523.19.6.1278

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). “Take over!” How long does it take to get the driver back into the loop? Proceedings of *the Human Factors and Ergonomics Society 57th Annual Meeting* (pp. 1938–1942). Santa Monica, CA: Human Factors and Ergonomics Society. https://doi.org/10.1177/1541931213571433

Gondan, M., Götze, C., & Greenlee, M. W. (2010). Redundancy gains in simple responses and go/no-go tasks. *Attention, Perception, & Psychophysics*, *72*, 1692–1709. https://doi.org/10.3758/APP.72.6.1692

Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention, Perception, & Psychophysics*, *78*, 723–735. https://doi.org/10.3758/s13414-015-1018-y

Harrar, V., Harris, L. R., & Spence, C. (2017). Multisensory integration is independent of perceived simultaneity. *Experimental Brain Research*, *235*, 763–775. https://doi.org/10.1007/s00221-016-4822-2

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, *63*, 289–293. https://doi.org/10.1037/h0039516

Hilbig, B. E. (2016). Reaction time effects in lab-versus Web-based research: Experimental evidence. *Behavior Research Methods*, *48*, 1718–1724. https://doi.org/10.3758/s13428-015-0678-9

Ho, C., Gray, R., & Spence, C. (2013). Role of audiovisual synchrony in driving head orienting responses. *Experimental Brain Research*, *227*, 467–476. https://doi.org/10.1007/s00221-013-3522-4

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Amsterdam, Netherlands: Elsevier. Karonen, I. (2012, March 8). Why is this GIF’s animation speed different in Firefox vs. IE? [Answer]. Available at https://webmasters.stackexchange.com/questions/26994/why-is-this-gifs-

animation-speed-different-in-firefox-vs-ie
Larsson, P., Johansson, E., Söderman, M., & Thompson, D. (2015). Interaction design for communicating system state and capabilities during automated highway driving. *Procedia Manufacturing*, *3*, 2784–2791. https://doi.org/10.1016/j.promfg.2015.07.735
Leone, L. M., & McCourt, M. E. (2013). The roles of physical and physiological simultaneity in audiovisual multisensory facilitation. *i-Perception*, *4*, 213–228. https://doi.org/10.1068/i0532
Leone, L. M., & McCourt, M. E. (2015). Dissociation of perception and action in audiovisual multisensory integration. *European Journal of Neuroscience*, *42*, 2915–2922. https://doi.org/10.1111/ejn.13087
Litman, L., Robinson, J., & Rosenzweig, C. (2015). The relationship between motivation, monetary compensation, and data quality among US- and India-based workers on Mechanical Turk. *Behavior Research Methods*, *47*, 519–528. https://doi.org/10.3758/s13428-014-0483-x
Madison, G., Mosing, M. A., Verweij, K. J., Pedersen, N. L., &
Ullén, F. (2016). Common genetic influences on intelligence and auditory simple reaction time in a large Swedish sample. *Intelligence*, *59*, 157–162. https://doi.org/10.1016/j.intell.2016.10.001
Mégevand, P., Molholm, S., Nayak, A., & Foxe, J. J. (2013). Recalibration of the multisensory temporal window of integration results from changing task demands. *PLOS ONE*, *8*, e71608. https://doi.org/10.1371/journal.pone.0071608
Merat, N., & Jamson, A. H. (2009). How do drivers behave in a highly automated car. Proceedings of *the 5th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Iowa City: University of Iowa, 514–521.
Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Attention, Perception, & Psychophysics*, *40*, 331–343. https://doi.org/10.3758/BF03203025
Miller, J., & Ulrich, R. (2003). Simple reaction time and statistical facilitation: A parallel grains model. *Cognitive Psychology*, *46*, 101–151. https://doi.org/10.1016/S0010-0285(02)00517-0
Nickerson, R. S. (1973). Intersensory facilitation of reaction time: Energy summation or preparation enhancement? *Psychological Review*, *80*, 489–509. https://doi.org/10.1037/h0035437
Petermeijer, S. M., Bazilinskyy, P., Bengler, K. J., & De Winter, J. C. F. (2017). Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics*, *62*, 204–215. https://doi.org/10.1016/j.apergo.2017.02.023
Petermeijer, S., De Winter, J. C. F., & Bengler, K. (2016). Vibrotactile displays: A survey with a view on highly automated driving. *IEEE Transactions on Intelligent Transportation Systems*, *17*, 897–907. https://doi.org/10.1109/TITS.2015.2494873
Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510–532.
Reimers, S., & Stewart, N. (2015). Presentation and response time accuracy in Adobe Flash and HTML5/Javascript Web experiments. *Behavior Research Methods*, *47*, 309–327. https://doi.org/10.3758/s13428-014-0471-1
Reimers, S., & Stewart, N. (2016). Auditory presentation and synchronization in

Adobe Flash and HTML5/JavaScript Web experiments. *Behavior Research Methods*, *48*, 897–908. https://doi.org/10.3758/s13428-016-0758-5

Schubert, T. W., Murteira, C., Collins, E. C., & Lopes, D. (2013). ScriptingRT: A software library for collecting response latencies in online studies of cognition. *PLOS ONE*, *8*, e67769. https://doi.org/10.1371/journal.pone.0067769

Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*, *49*, 1241–1260. https://doi.org/10.3758/s13428-016-0783-4

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*, 255–266. https://doi.org/10.1038/nrn2331

Todd, J. W. (1912). Reaction to multiple stimuli. In R. S. Woodworth (Ed.), *Archives of psychology, XXI*. Columbia Contributions to Philosophy and Psychology. New York: Science Press.

Ulrich, R., & Giray, M. (1989). Time resolution of clocks: Effects on reaction time measurement-Good news for bad clocks. *British Journal of Mathematical and Statistical Psychology*, *42*, 1–12. https://doi.org/10.1111/j.2044-8317.1989.tb01111.x

Van der Stoep, N., Spence, C., Nijboer, T. C. W., & Van der Stigchel, S. (2015). On the relative contributions of multisensory integration and crossmodal exogenous spatial attention to multisensory response enhancement. *Acta Psychologica*, *162*, 20–28. https://doi.org/10.1016/j.actpsy.2015.09.010

Wang, Y., Toor, S. S., Gautam, R., & Henson, D. B. (2011). Blink frequency and duration during perimetry and their relationship to test-retest threshold variability. *Investigative Ophthalmology & Visual Science*, *52*, 4546–4550. https://doi.org/10.1167/iovs.10-6553

Welch, B. L. (1947). The generalization of “Student’s” problem when several different population variances are involved. *Biometrika*, *34*, 28–35. https://doi.org/10.1093/biomet/34.1-2.28

Wickens, C. D. (2001, March). *Attention to safety and the psychology of surprise*. Paper presented at the 2001 International Symposium on Aviation Psychology, Columbus, OH.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*, 33–39. https://doi.org/10.1111/1467-9280.00006

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the Internet: A tutorial review. *PeerJ*, *3*, e1058. https://doi.org/10.7717/peerj.1058

3.2 Analyzing crowdsourced ratings of speech-based take-over requests for automated driving

Corresponding article: Bazilinskyy, P., & De Winter, J. C. F. (2017). Analyzing crowdsourced ratings of speech-based take-over requests for automated driving. Applied Ergonomics, 64, 56–64.

3.2.1 Abstract

Take-over requests in automated driving should fit the urgency of the traffic situation. The robustness of various published research findings on the

valuations of speech-based warning messages is unclear. This research aimed to establish how people value speech-based take-over requests as a function of speech rate, background noise, spoken phrase, and speaker's gender and emotional tone. By means of crowdsourcing, 2,669 participants from 95 countries listened to a random 10 out of 140 take-over requests, and rated each take-over request on urgency, commandingness, pleasantness, and ease of understanding. Our results replicate several published findings, in particular that an increase in speech rate results in a monotonic increase of perceived urgency. The female voice was easier to understand than a male voice when there was a high level of background noise, a finding that contradicts the literature. Moreover, a take-over request spoken with Indian accent was found to be easier to understand by participants from India than by participants from other countries. Our results replicate effects in the literature regarding speech-based warnings, and shed new light on effects of background noise, gender, and nationality. The results may have implications for the selection of appropriate take-over requests in automated driving. Additionally, our study demonstrates the promise of crowdsourcing for testing human factors and ergonomics theories with large sample sizes.

3.2.2 Introduction

3.2.2.1 Take-over requests

Until cars can drive autonomously, there will be situations where the driver has to resume manual control. Prior to such control transition, the automation may issue a take-over request to the driver (SAE International, 2016; Zeeb, Buchner, & Schrauf, 2015). How to provide a take-over request is a widely studied topic in human factors and ergonomics (Hergeth, Lorenz, Krems, & Toenert, 2015; Naujoks, Mai, & Neukum, 2014; Petermeijer, De Winter, & Bengler, 2016; Pfromm, Khan, Oppelt, Abendroth, & Brudera, 2015).

A take-over request can be provided through pre-recorded voice (Gold, Berisha, & Bengler, 2015; Mok et al., 2015; Politis, Brewster, & Pollick, 2015), which may be an effective approach because humans are able to perceive sounds irrespective of head or eye orientation (Bazilinskyy & De Winter, 2015; Meng & Spence, 2015). In aviation, a similar approach is used: traffic alert and collision avoidance systems (TCAS), which are mandatory in today's aircraft, apply voice commands (Kuchar & Yang, 2000).

Take-over situations may be of different urgency. Several studies have measured driver behavior in highly urgent situations, such as Mok et al. (2015), who found that 50% of the drivers veered off the road when a critical lane-closure event followed only 2 seconds after a take-over request ("Emergency, Automation off"). Other studies have been concerned with larger lead times of 5 or 7 seconds (Gold, Damböck, Lorenz, Bengler, 2013, see Eriksson & Stanton, in press, for an overview) or with discretionary transitions having a low urgency (Damböck, Weißgerber, Kienle, & Bengler, 2013; Merat & Jamson, 2009; Nilsson, Strand, Falcone, & Vinter, 2013). Politis et al. (2015) found that participants reacted 1.3 seconds faster to urgent take-over requests ("Danger! Collision imminent; You have control!") than to non-urgent ones (e.g., "Warning! GPS signal weak; Want to take over?"). In sum, how to convey the right sense of urgency is regarded as an important topic in automated driving research.

3.2.2.2 Speech warnings

Previous research has shown that semantics have an effect on urgency, in that a word such as 'danger' is perceived as more urgent than 'attention' (Arrabito, 2009; Baldwin, 2011; Wogalter & Silver, 1995; Wogalter, Conzola, & Smith-Jackson, 2002). Second, emotional tone has important effects: phrases are considered more urgent if spoken in an urgently intoned style (Edworthy, Hellier, Walters, Clift-Mathews, & Crowther, 2003; Ljungberg, Parmentier, Hughes, Macken, & Jones, 2012). Third, it has been found that the greater the speech rate, the higher the perceived urgency (Hollander & Wogalter, 2000; Jang, 2007; Park & Jang, 1999). No clear gender effects seem to exist: words spoken by a female typically yield similar urgency ratings as the same words spoken by a male (e.g., Hellier, Edworthy, Weedon, Walters, & Adams, 2002; Wogalter et al., 2002). However, Jang (2007) and Park and Jang (1999) found that a male voice yielded higher urgency ratings than a female voice. Furthermore, interaction effects have been observed where the word "Note" received a higher urgency rating when spoken by a male instead of a female (Hellier et al., 2002). Differences in the degree of smoothness, pitch, and timbre, may explain these gender differences (Edworthy et al., 2003; Edworthy, Hellier, & Rivers, 2003; Jang, 2007).

In addition to urgency, it is important to consider whether the message is comprehensible and pleasant. If people become displeased with a warning, they may ignore or disable the warning system, potentially causing unsafe situations (Eichelberger & McCartt, 2014; Parasuraman & Riley, 1997). A female voice has been regarded as more pleasant (Bazilinskyy & De Winter, 2015; Machado, Duarte, Teles, Reis, & Rebelo, 2012) and is more often used in route navigation devices (Large & Burnett, 2013) than a male voice. The female and male voice are supposedly equal in terms in intelligibility, but it has been reported that the male voice is easier to understand in a noisy environment such as an aircraft cockpit (Nixon et al., 1998, Noyes, Hellier, & Edworthy, 2006). However, it is unknown whether this effect is replicable. Arrabito (2009) stated that "further research is required to study the effects of speech parameters and word semantics across multiple talkers of each sex for variations of urgency under different background noise sources" (p. 18).

There is currently an irony in automated driving, because the technologies are deployed in the highest-income countries, which already have commendable road safety statistics, while low-income countries account for the vast majority of fatal road traffic accidents (Gururaj, 2008; World Health Organization, 2015). At present, car manufacturers are exploring cross-national perceptions of warnings (Langlois, Suied, Lageat, & Charbonneau, 2008), but it is unknown whether speech-based take-over requests should be differentially developed per country. Research has shown that there are national differences in how people perform at basic visual perception tasks (Henrich, Heine, & Norenzayan, 2010). Regarding the appraisal of sounds, similar differences may exist. For example, it has been found that the sound of a bell was rated as pleasant among German listeners (possibly because it yielded connotations to a church bell), whereas this sound was rated as dangerous and unpleasant among Japanese listeners (Fastl, 2006). One specific question is whether a speech-based warning should be tailored to the language and accent of the host country. For example, it is possible that drivers from the UK prefer a British accent, and drivers from

the US prefer an American accident. It has been found that a foreign English accent does not reduce the intelligibility and comprehensibility of speech (Munro & Derwing, 1995; Munro, 2008; Smith & Rafiqzad, 1979), but these findings deserve further investigation.

3.2.2.3 Aim of the study

This paper assesses how different speech-based take-over requests are perceived. Specifically, in line with the above research gaps, we assessed (1) the effects of speech rate on perceived urgency, commandingness, pleasantness, and ease of understanding, for speakers that differ in gender and emotional tone. Additionally, we investigated (2) the effects of spoken phrase (semantic content) on perceived urgency for a male and female speaker, (3) the effects of noise on the ease of understanding, for a male and female speaker, and (4) the effect of participants' (i.e., listeners') gender on pleasantness. Finally, we explored (5) the relationship between the participants' country and the ease of understanding of the messages. To acquire a large sample, we used crowdsourcing, an approach that is gaining popularity (Bazilinskyy & De Winter, 2015; Behrend, Sharek, Meade, & Wiebe, 2011; Buhrmester, Kwang, Gosling, 2011; Crump, McDonnell, & Gureckis, 2013; Kyriakidis, Happee, & De Winter, 2015; Rand, 2012).

3.2.3 Methods

This research was approved by the Human Research Ethics Committee at the TU Delft under the ethics approval application titled "Rating audio messages by means of crowdsourcing" on May 24, 2016. Informed consent was obtained from each participant via a dedicated survey item.

3.2.3.1 Speech-based messages

Speech-based messages "Take over, please" were created using the online tool Acapela-Box (https://acapela-box.com). Acapela-Box reproduces the natural sound of language based on voice of human speakers, and was selected because it offers high-quality speech and adjustability of speech rate. Two male voices (Will: US English accent; Graham: UK English accent) and two female voices (Karen: US English accent; Deepa: Indian English accent) were used. These three English accents represent highly populated countries with a strong automotive industry where English is either the first language (US and UK) or one of the official languages (India). The tool offered the option for speech to be generated with an emotional tone. We created recordings for two emotional tones by selecting speakers Will Happy and Will FromAfar. We expected that Will FromAfar, in which the speaker shouts the words from a distance, would be interpreted as urgent. Will Happy was expected to sound pleasant among listeners. Note that Acapela-Box offered a limited number of speakers and emotional tones: there was no male voice with Indian English accent, and among the US English speakers, the Happy and Afar emotional tones were only available for Will. Furthermore, different voices exhibited different speech rates (e.g., Deepa spoke relatively fast).

Table 1. Overview of the sound samples for the phrase "Take over, please" at the nominal speech rate. Shown in parentheses is the sound volume when background noise was added to the original sample, for noise levels 1, 2, 3, and 4.

Speaker	Duration (s)	Maximum volume (0–1)	Mean volume (0–1)
Will	1.67	0.332 (0.384, 0.541, 0.816, 1.000)	0.038 (0.053, 0.108, 0.180, 0.311)
Karen	1.62	0.335 (0.435, 0.660, 0.919, 1.000)	0.051 (0.062, 0.113, 0.183, 0.312)
Graham	1.59	0.333 (0.357)	0.031 (0.047)
Deepa	1.28	0.266 (0.326)	0.029 (0.048)
Will Happy	1.78	0.337 (0.405)	0.044 (0.058)
Will FromAfar	1.96	0.046 (0.160)	0.009 (0.033)

Using Acapela-Box, each of the six speakers was recorded at eight additional settings of speech rate: −60, −45, −30, −15, +15, +30, +45, and +60, which altered the duration of the sample to approximately 151%, 131%, 119%, 109%, 90%, 85%, 79%, and 76% of its nominal value, respectively. In addition, for each speaker and speech rate, background noise was added, extracted from a YouTube video showing a Tesla Model S in Autopilot mode (Oedegaarde, 2015). For Will and Karen, noise with three extra levels of volume was added (Table 1).

Moreover, 13 phrases were recorded using Will and Karen at a nominal speech rate and without added noise: (1) “Take over please?”, (2) “Take over”, (3) “Please take over”, (4) “Could you please take over”, (5) “Could you please take over?”, (6) “Take over now”, (7) “Take over immediately”, (8) “Hazard: take over”, (9) “Danger: take over”, (10) “Warning: take over”, (11) “Caution: take over”, (12) “Attention: take over”, and (13) “Note: take over”.

In summary, the number of recordings was 140, consisting of 108 recordings where speech rate and noise were varied for each of the six speakers (6 speakers x 9 speech rate levels x 2 noise levels) plus 32 recordings (3 noise levels and 13 additional phrases, for Will and Karen).

3.2.3.2 Survey

A survey was developed using CrowdFlower (http://www.crowdflower.com). At the beginning of the survey, contact information of the researchers was provided, and the purpose of the survey was described as “to determine the public opinion on auditory messages that may be used in automated driving”. Participants were informed that the survey would take 5 minutes of their time. The participants were also informed that they had to be at least 18 years old. Information about anonymity and voluntarily participation was provided as well.

The survey started with a question about whether the participant had read and understood the instructions, and contained questions on the participant’s age, gender, driving experience, and opinion on automated driving. The main part of the survey focused on the voice recordings. Each participant was given a randomized selection of 10 out of the 140 voice recordings. The participants were asked to click on the recordings to listen to them. The filenames of the recordings were masked as voiceXXX.mp3 (with XXX being a number between 1 and 140).

Below each recording, five questions were provided: (1) “Did you listen to the recording of a female or male voice in recording XX?” (this was a test question), (2) “The message in recording XX is urgent.”, (3) “The message in recording XX is pleasant.”, (4) “The message in recording XX is commanding.”, (5) “The message in recording X is easy to understand.”, where XX denotes a number between 1 and 10. For questions 2–5, the response options were

"Disagree strongly", "Disagree a little", "Neither agree nor disagree", "Agree a little", "Agree strongly", and "I prefer not to respond". The participants had to answer all questions in order to complete the survey. The survey did not explain the notions of urgency, pleasantness, commandingness, and ease of understanding to the participants.

3.2.3.3 CrowdFlower configuration

We allowed contributors from all countries to participate in the survey. Completing the survey more than once from the same CrowdFlower worker ID was not permitted. A payment of $0.14 was offered for the completion of the survey. We collected responses from 3,061 participants, at a total cost of $524.

3.2.3.4 Analyses

Five analyses were conducted. First, we determined the effect of speech rate on the degree to which the message was regarded as urgent, pleasant, commanding, and easy to understand. Second, we evaluated the effect of the 14 different phrases on perceived urgency, and whether there were differences between the speakers of different gender (Will vs. Karen). Third, we assessed the effect of noise level on whether the message was easy to understand for each of the six speakers, and whether the male voice (Will) was easier to understand than the female voice (Karen) as a function of noise level. Fourth, we determined the effect of the participants' (i.e., listeners') gender on pleasantness, for each of the six speakers. Finally, we assessed whether the six speakers had different levels of comprehensibility for participants from different countries, with the participants' country being automatically identified by CrowdFlower. In order to arrive at statistically reliable conclusions, we included only those countries with 100 or more participants in the cross-national analyses.

All analyses were conducted at the level of participants. If multiple responses per condition were available per participant (e.g., responses to recordings with and without background noise for the same speaker and speech rate), then these responses were averaged per participant. The mean scores on a scale from 1 (*Disagree strongly*) to 5 (*Agree strongly*) were calculated and visualized in bar graphs. For each depicted mean, the 95% confidence interval was provided, defined as the mean ± 1.96 times the standard deviation divided by the square root of the sample size. Comparisons between selected pairs of conditions were conducted by means of independent-samples *t* tests. A previous simulation study showed that for five-point Likert data, the *t* test provides appropriate statistical power and protection against false positives (De Winter & Dodou, 2010). In principle our experiment has elements of a within-subject design, because each participant rated multiple auditory samples. However, because each participant rated only 7.1% (=10/140) of random auditory samples, the probability was low that a participant rated a reference sample that could be used in a paired comparison. Therefore, we conducted between-subjects statistical analyses.

No corrections for multiple comparisons were applied, because our interest was not only in detecting whether effects are statistically significant, but also in showing whether effects are not statistically significant despite large sample sizes. In other words, if we had reduced the significance level to a value

smaller than the nominal 0.05, then a finding of 'no statistically significant differences' would not be compelling.

3.2.4 Results

The responses were collected between 29 May 2016, 13:30 and 5 June 2016, 21:35 (GMT). Each of the 3,061 participants answered four queries (urgency, pleasantness, commandingness, ease of understanding) regarding 10 voice recordings. 337 participants completed the optional user satisfaction survey. The satisfaction survey received an overall satisfaction score of 4.4 out of 5.0 (1 = *very dissatisfied*, 5 = *very satisfied*), with "instructions clear", "test questions fair", "ease of job", and "pay" receiving ratings of 4.6, 4.3, 4.3, and 4.1, respectively.

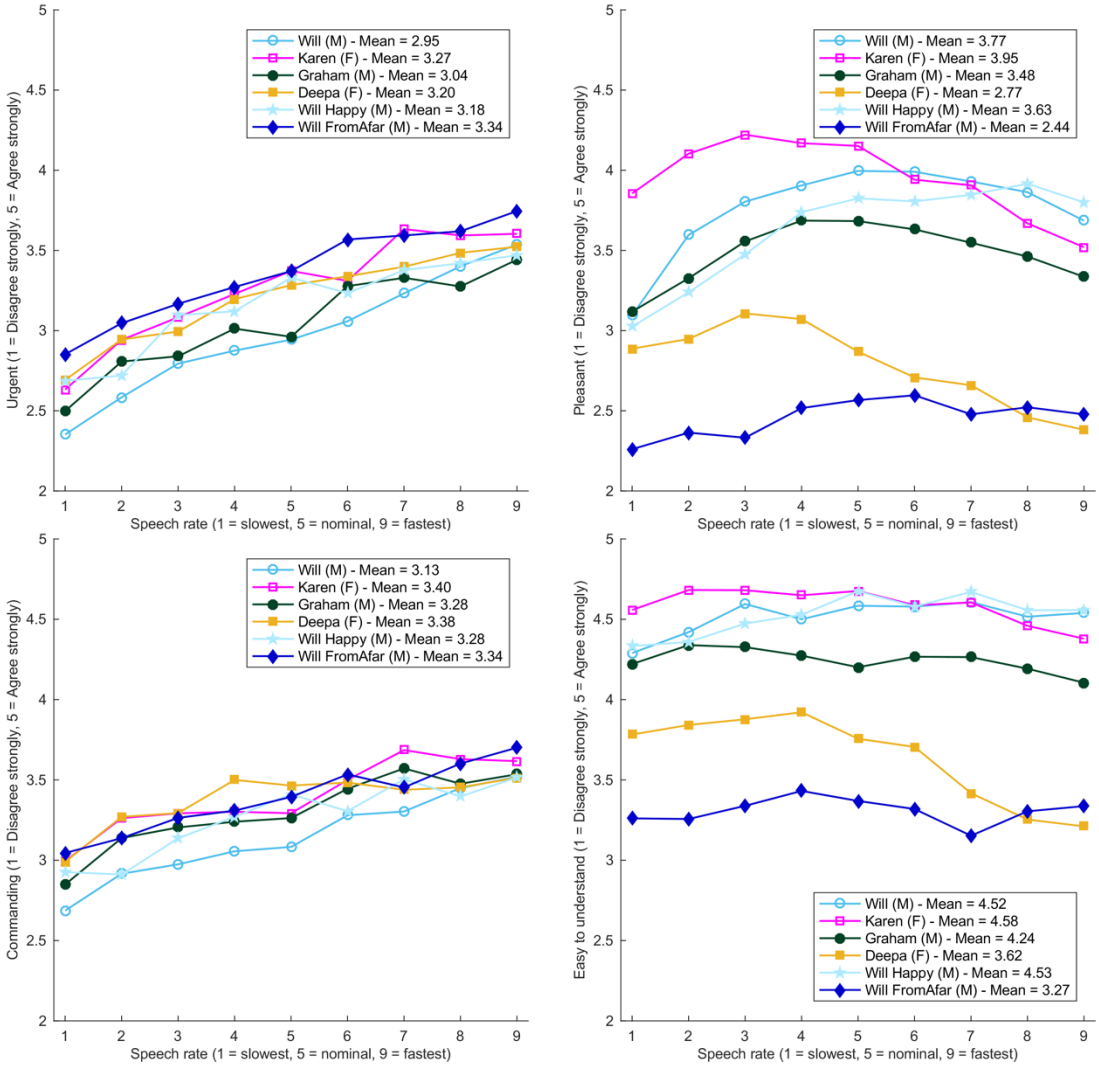


Figure 1. Participants' ratings as a function of speech rate and speaker. Left top = mean urgency, Right top = mean pleasure, Left bottom = mean commandingness, Right bottom = mean ease-of-understanding. Each individual point in these four graphs represents the average across 2 of 140 sound recordings (i.e., noise trials and no-noise trials averaged), from an average of 366 participants (min = 320, max = 402; the corresponding 95% confidence intervals per point ranges between 0.12 and 0.31). The overall mean across the nine speech rates per speaker is indicated in the legend box. The corresponding sample size per speaker (all nine speech rates aggregated) in the

four graphs is on average 2019 (min = 1992, max = 2046) and the 95% confidence interval per speaker ranges between 0.060 and 0.119.

Participants who indicated they had not read the instructions (N = 25), were 17 or younger (N = 3), or whose country was not identified (N = 3) were excluded. As a data quality filter, participants who made one or more mistakes in the question 'did you listen to the recording of a female or male voice?' were excluded (N = 375). Regarding this latter exclusion criterion, Deepa was not taken into consideration because we ourselves had difficulty identifying whether Deepa was male or female. A sizeable portion of participants also seemed to have difficulty distinguishing whether Deepa was male or female (there were 10% errors for Deepa versus 4% error for the other speakers). The results were hardly affected by the decision not to include Deepa in this filtering process. In total 392 participants were excluded, leaving 2,669 participants from 95 countries. The mean survey completion time was 580 s (SD = 285 s). The mean age was 33.7 years (SD = 10.6) and the sample consisted of 1,777 males, 884 females, and 8 participants with unknown gender. These 8 participants selected 'I prefer not to respond' and were retained in the analysis.

The effects of speaker and speech rate are shown in Figure 1. The higher the speech rate, the higher the ratings of urgency and commandingness. Averaged across the nine speech rates, Will FromAfar received the highest urgency ratings (M = 3.34) and Will the lowest (M = 2.95). Speech rate had non-monotonic effects on pleasure, showing different inverted U-shapes per speaker. The female speaker Karen was rated as most pleasant at low speech rates. Although Will FromAfar was rated as urgent, he was rated as least pleasant (M = 2.44) and least well understood (M = 3.27). It is possible that the low intelligibility of Will FromAfar was caused by its low volume (Table 1). The speaker with Indian accent (Deepa) did not receive high pleasure ratings either (M = 2.77); Deepa had a high speech rate in its nominal condition (Table 1), and higher speech rates were considered to be unpleasant.

Figure 2 confirms that the spoken phrase has an effect on urgency, with "Take over immediately" and "Danger: take over" yielding the highest urgency and "Could you please take over?" the lowest. There were statistically significant gender differences, with "Take over, please", "Take over please?", and "Please take over" being perceived as more urgent when spoken by Karen than when spoken by Will, while the opposite was observed for "Take over now". It is worth noting that for the 14 spoken phrases, there was a positive correlation between the mean urgency and the mean commandingness (r = 0.92, n = 14), but a negative correlation between the mean urgency and the mean pleasantness (r = –0.72, n = 14). The highest commandingness was found for "Take over now" (M = 4.40) and the highest pleasantness was found for "Take over, please" (M = 4.06).

The results regarding noise are shown in Figure 3. The t tests show that a mild noise level (Level 1 noise) has only minor effects on ease of understanding, except for Will FromAfar, which had a low volume without noise (Table 1). The ease of understanding dropped with increasing noise level (see Will and Karen in Fig. 3). In contrast to Nixon et al. (1998), the female voice (Karen) was easier to understand than the male voice (Will), especially at higher noise levels (Fig. 3). These gender differences were statistically significant; No noise: $t(2690) = -1.84$, $p = 0.065$; Level 1 noise: $t(2597) = -3.22$, $p = 0.001$; Level 2 noise: $t(378) = -2.79$,

$p = 0.005$; Level 3 noise: $t(367) = -4.20$, $p < 0.001$; Level 4 noise: $t(359) = -2.90$, $p = 0.004$.

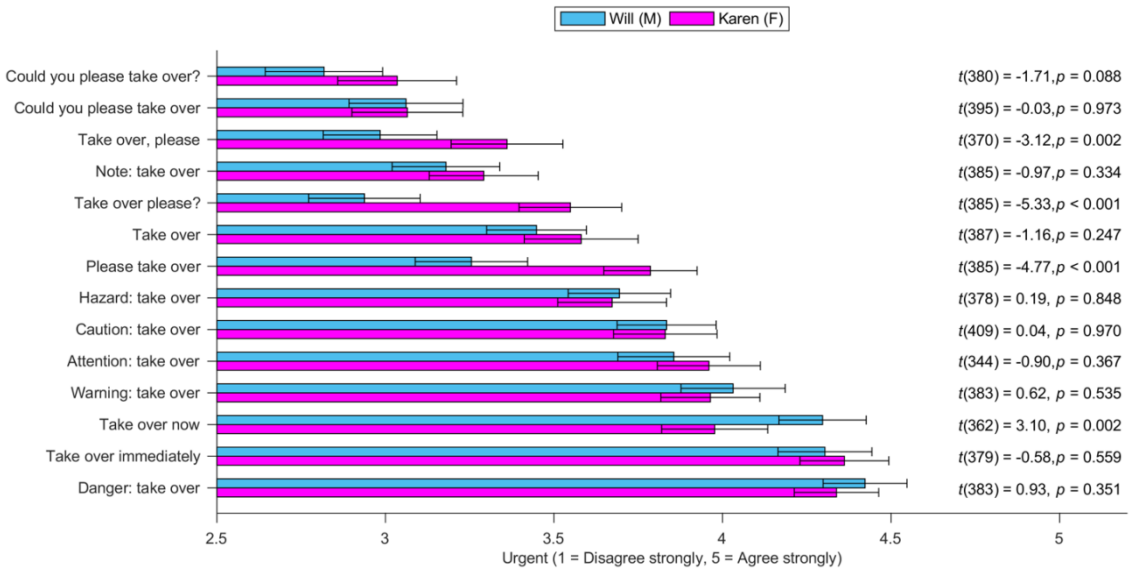


Figure 2. Mean urgency scores for different phrases expressed by Will and Karen. The figure also depicts the results of a comparison between Will and Karen by means of independent-samples *t* tests. Each individual bar in the graph represents the average for 1 of 140 sound recordings. The error bars represent 95% confidence intervals. The 14 phrases are sorted on the mean urgency level.

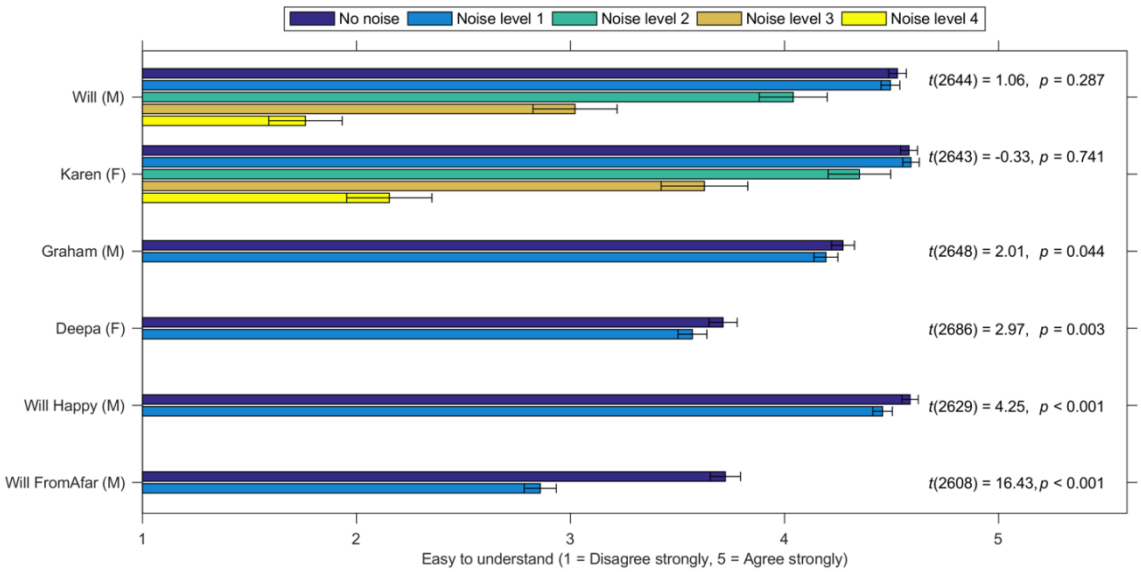


Figure 3. Mean ease-of-understanding scores for the six speakers as a function of background noise level. The figure also depicts the results of a comparison between noise level 0 and noise level 1 by means of independent-samples *t* tests. Noise and volume levels are described in Table 1. The bars for No noise and Noise level 1 represent the average across 9 of 140 sound recordings (i.e., across the 9 levels of speech rate). The bars for Noise levels 2, 3, and 4 are the average for 1 of 140 recordings. The error bars represent 95% confidence intervals.

The results regarding the participants' gender revealed no statistically significant differences on pleasantness, for four of the six speakers, despite the fact that statistical power was high, with about 2,000 degrees of freedom (Fig.

4). Deepa and Will FromAfar were rated as slightly more pleasant by male participants than by female participants.

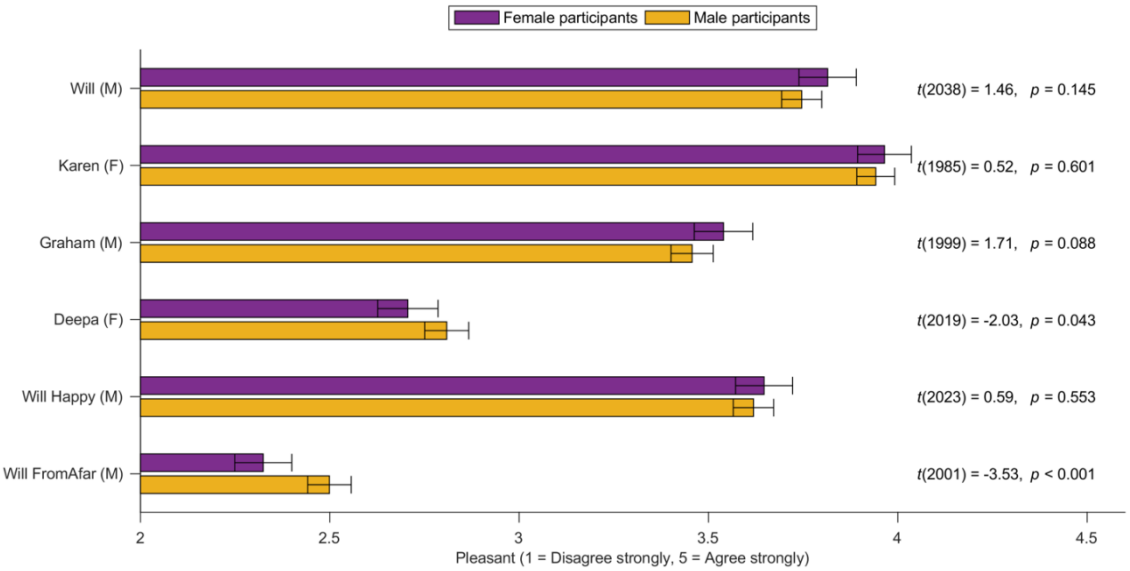


Figure 4. Mean pleasure scores for the six speakers as a function of participant's gender. The figure also depicts the results of a comparison between female and male participants by means of independent-samples *t* tests. The bars represent the average across 18 of 140 sound recordings (i.e., across the 9 levels of speech rate, and for both noise levels). The error bars represent 95% confidence intervals.

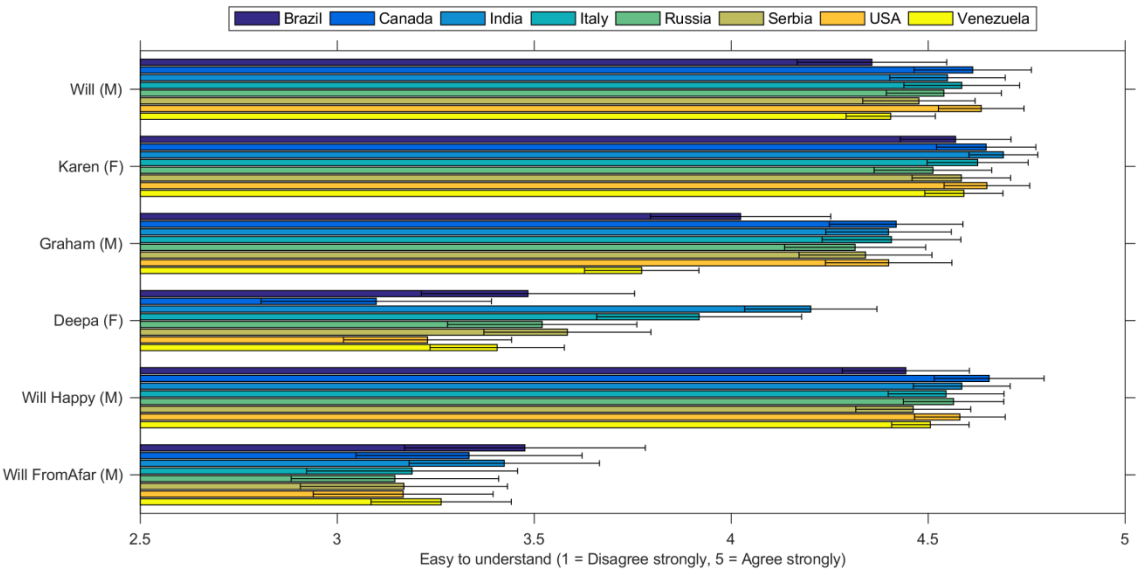


Figure 5. Mean 'easy to understand' scores per participant's country and speaker. Only countries with 100 or more participants were shown. The bars represent the average across 18 of 140 sound recordings (i.e., across the 9 levels of speech rate, and for both noise levels). The sample size per bar is 70–79 for Brazil, 80–84 for Canada, 114–131 for India, 84–97 for Italy, 83–92 for Russia, 104–113 for Serbia, 131–142 for USA, and 232–254 for Venezuela. The error bars represent 95% confidence intervals.

Finally, we assessed national differences. The mean ease-of-understanding scores for the eight countries with 100 or more participants are shown in Figure 5. The results in Figure 5 are consistent across these geographically diverse countries, with Will FromAfar being rated as difficult to understand, and Will, Karen, and Will Happy receiving high scores. Deepa, who

had an Indian accent, received higher ratings from Indian participants than from participants from other countries. To illustrate, the mean ease-of-understanding rating for Deepa was significantly greater for participants from India (M = 4.20, N = 123) than for participants from the USA (M = 3.23, N = 138; $t(259) = 6.88$, $p < 0.001$) and Venezuela (M = 3.41, N = 239; $t(360) = 5.84$, $p < 0.001$).

3.2.5 Discussion

This study determined how people value speech-based take-over requests as a function of speech rate, background noise, speaker (gender and emotional tone), and spoken phrase, by means of a crowdsourcing study with a large sample size. A total of 2,669 participants completed the task over the course of 7 days.

There are several advantages to using a large sample size. First, a larger sample size increases statistical power, which means that if a research finding is true, it is more likely to be detected. Second, a larger sample size increases the probability that a research finding is in fact true. Third, if the sample size is larger, the results are less susceptible to bias (Gadbury & Allison, 2012; Ioannidis, 2005; Wagenmakers et al., 2015). In recent years, psychology has been said to be in a replication crisis (Maxwell, Lau, & Howard, 2015). This concern was recently confirmed by the Open Science Collaboration (2015), showing that from 97 published significant effects, only 35 replicated. Small samples are a prime cause of poor replicability, a message that has now transpired to many fields, including medicine (Arrowsmith, 2011; Begley & Ellis, 2012; Freedman, Cockburn, & Simcoe, 2015), economics (Ioannidis & Doucouliagos, 2013), and neuroscience (Button et al., 2013). Asendorpf et al. (2013) argued that “it cannot be stressed enough that researchers should collect bigger sample sizes, and editors, reviewers and readers should insist on them” (p. 110).

Our research replicated several published effects. In agreement with Park and Jang (1999), an increase of speech rate yielded an increase of self-reported urgency, an effect that held regardless of the gender or emotional tone of the speaker. In agreement with Hellier et al. (2002), amongst others, the spoken phrase (e.g., “Danger” versus “Note”) had an important impact on perceived urgency as well. Overall, our results point to the robustness of published human factors and ergonomics research, and are in line with the idea that psychological effects generalize well across different research settings (Klein et al., 2014).

Several of our findings are in disagreement with the literature. First, Hellier et al. (2002) found that the word “Note” received higher urgency ratings when spoken by a male than when spoken by a female, whereas we found no statistically significant gender effect for “Note: take over”. This discrepancy may be a consequence of the specific phrase and intonation. Perhaps our findings represent a social-psychological phenomenon in which direct utterances (“now”) are deemed urgent when spoken by a male, whereas suggestive utterances (“please”) are deemed urgent when spoken by a female (cf. Fig. 2). Second, Nixon et al. (1998) found that the intelligibility of female speech was lower than that of male speech, especially for strong cockpit noise, whereas our results showed the opposite, with the female voice being easier to understand under strong background noise. Third, the fact that a speaker with Indian accent was relatively easy to understand by listeners from India is in line with the ’interlanguage speech intelligibility benefit’ (Bent & Bradlow, 2003; Podlipský,

Šimáčková, & Petráž, 2016), but appears to contradict published literature stating that "listeners did not consistently exhibit an intelligibility benefit for speech produced in their own accent" (Munro, Derwing, & Morton, 2006, p. 111). It is noted that we did not perform a direct replication of past research, but rather a conceptual replication (Stroebe, 2016). Our findings therefore do not refute the original findings, but rather suggest that there may be various unknown moderators at play. Possibly, specific features in the speaker's voice relative to the background noise (e.g., vehicle vs. aircraft noise) may have made the female voice stand out more (see also Cooke, Mayo, Valentini-Botinhao, Stylianou, Sauert, & Tang, 2013; Lerner, Singer, Kellman, & Traube, 2015).

One limitation of crowdsourcing is that the participant pool is limited in size, encompassing several thousands of people (Chandler, Mueller, & Paolacci, 2014; Stewart et al., 2015). Participants in our study were from 95 different countries, and previous research has found that there are national income-related differences in driving culture, traffic violations, and opinion about automated cars (De Winter & Dodou, 2016; Kyriakidis et al., 2015; Özkan & Lajunen, 2007). Furthermore, it has been found that people from non-English speaking countries take longer to complete CrowdFlower surveys than people from English speaking countries, which may signal difficulties with reading and interpreting the questions (De Winter et al., 2015). Considering the heterogeneity of our participant pool, it remains to be seen how well the observed effects apply to a specific target population of prospective users of automated cars. However, a similar limitation applies to lab-based research often conducted at universities with students as research participants

Second, our study was not performed in a realistic driving context. The participants were not shown any automated driving scenarios, and we did not measure the response times of participants (see Arrabito, 2009; Ljungberg & Parmentier, 2012, in which participants' responses to speech were measured). It remains to be investigated how actual drivers would respond to speech-based take-over requests. It is possible that in demanding real-life traffic scenarios, a driver may be confused by the message "Take over please", especially if other warning sounds can be heard simultaneously. In addition, we learned through a discussion with fellow researchers that the phrase "Take over" (i.e., to reclaim manual control) may be suboptimal because it can be confused with "Overtake" (i.e., to pass a vehicle in front). Driving simulator studies in which drivers are exposed to different driving contexts are recommended in order to resolve these uncertainties.

Third, even though we used as many as 140 different auditory samples, our results may still be limited because the auditory samples reflect only a snapshot of the types of male and female voices and their emotional tones. To further explore whether the female voice is easier to understand than the male voice under background noise, multiple male and female voices and different types of noise spectra could be tested. Additionally, to better understand the interaction between listeners' gender and the speaker's accent, different accents could be included (other than the present US, UK, and Indian accents). The text-to-speech tool that we used offered a limited number of English accents and emotional tones. Recent developments in artificial intelligence give rise to increasingly flexible text-to-speech systems (e.g., Arik et al., 2017). The development of new software that offers a high range of choices for the customization of synthesized voice will be beneficial for future research on the

valuation of speech. The number of auditory samples that can be tested depends on the researchers' financial resources and on the size of the participant pool on the crowdsourcing platform.

Because we used computerized speech, automotive researchers can readily reproduce the same speech warnings as used in this research. The results in Figure 1 can be used to select a take-over request by considering each of the four dimensions. Our study also has implications for human factors research in general. We showed that robust knowledge can be generated via the Internet, confirming earlier claims that crowdsourcing is a viable research tool (Crump et al., 2013). An important strength of this research is that it is effectively a between-subjects design, with participants listening to only 10 out of 140 take-over requests. In lab-based research, one usually has to resort to within-subjects designs to generate sufficient statistical power. Within-subject designs introduce carryover effects, which counterbalancing does not perfectly resolve (Greenwald, 1976; Keren, 1993). Crowdsourcing may be especially worthwhile when the experiment requires no special apparatus, as is often the case in usability research. Examples of research suited for crowdsourcing are perceptual tasks, cognitive tasks, and questionnaires (Buhrmester et al., 2011; De Winter et al., 2015; Rand, 2012).

3.2.6 Supplementary material

The CrowdFlower survey, raw data, and scripts can be found online: https://doi.org/10.1016/j.apergo.2017.05.001

3.2.7 References

Arik, S. Ö., Chrzanowski, M., Coates, A., Diamos, G., Gibiansky, A., Kang, Y., et al. (2017). Deep Voice: Real-time neural text-to-speech. Available at https://arxiv.org/pdf/1702.07825.pdf

Arrabito, G. R. (2009). Effects of talker sex and voice style of verbal cockpit warnings on performance. *Human Factors, 51*, 3–20.

Arrowsmith, J. (2011). Trial watch: phase II failures: 2008–2010. *Nature Reviews Drug Discovery, 10*, 328–329.

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., ... & Perugini, M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*, 108–119.

Baldwin, C. L. (2011). Verbal collision avoidance messages during simulated driving: perceived urgency, alerting effectiveness and annoyance. *Ergonomics, 54*, 328–337.

Bazilinskyy, P., & De Winter, J. (2015). Auditory interfaces in automated driving: an international survey. *PeerJ Computer Science*, *1*, e13. https://doi.org/10.7717/peerj-cs.13

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature, 483*, 531–533.

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43*, 800–813.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America, 114*(3), 1600–

1610.
Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5.
Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365–376.
Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*, 112–130.
Cooke, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauert, B., & Tang, Y. (2013). Evaluating the intelligibility benefit of speech modifications in known noise conditions. *Speech Communication, 55*, 572–585.
Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE, 8*, e57410.
Damböck, D., Weißgerber, T., Kienle, M., & Bengler, K. (2013). Requirements for cooperative vehicle guidance. Proceedings of *the 16th International IEEE Annual Conference on Intelligent Transportation Systems*, The Hague, the Netherlands, 1656–1661.
De Winter, J. C. F., & Dodou, D. (2010). Five-point Likert items: t test versus Mann-Whitney-Wilcoxon. *Practical Assessment, Research & Evaluation, 15*, 11.
De Winter, J. C. F., & Dodou, D. (2016). National correlates of self-reported traffic violations across 41 countries. *Personality and Individual Differences, 98*, 145–152.
De Winter, J. C. F., Kyriakidis, M., Dodou, D., & Happee, R. (2015). Using CrowdFlower to study the relationship between self-reported violations and traffic accidents. Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics (AHFE), Las Vegas, NV*, 2518–2525. https://doi.org/10.1016/j.promfg.2015.07.514
Edworthy, J., Hellier, E., & Rivers, J. (2003). The use of male or female voices in warnings systems: A question of acoustics. *Noise and Health, 6*, 39–50.
Edworthy, J., Hellier, E., Walters, K., Clift-Mathews, W., & Crowther, M. (2003). Acoustic, semantic and phonetic influences in spoken warning signal words. *Applied Cognitive Psychology, 17*, 915–933.
Eichelberger, A. H., & McCartt, A. T. (2014). Volvo drivers' experiences with advanced crash avoidance and related technologies. *Traffic Injury Prevention, 15*, 187–195.
Eriksson, A., & Stanton, N. A. (in press). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors*.
Fastl, H. (2006). Psychoacoustic basis of sound quality evaluation and sound engineering. Proceedings of *the Thirteenth International Congress on Sound and Vibration, Vienna, Austria*.
Freedman, L. P., Cockburn, I. M., & Simcoe, T. S. (2015). The economics of reproducibility in preclinical research. *PLOS Biology, 13*, e1002165.
Gadbury, G. L., & Allison, D. B. (2012). Inappropriate fiddling with statistical analyses to obtain a desirable p-value: tests to detect its presence in

published literature. *PLOS ONE, 7*, e46363.
Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). “Take over!” How long does it take to get the driver back into the loop? Proceedings of *the Human Factors and Ergonomics Society Annual Meeting, 57*, 1938–1942.
Gold, C., Berisha, I., & Bengler, K. (2015). Utilization of drivetime – Performing non-driving related tasks while driving highly automated. Proceedings of *the Human Factors and Ergonomics Society 59th Annual Meeting, 59*, 1666–1670.
Greenwald, A. G. (1976). Within-subject designs: To use or not to use? *Psychological Bulletin, 83*, 314–320.
Gururaj, G. (2008). Road traffic deaths, injuries and disabilities in India: current scenario. *National Medical Journal of India, 21*, 14–20.
Hellier, E., Edworthy, J., Weedon, B., Walters, K., & Adams, A. (2002). The perceived urgency of speech warnings: Semantics versus acoustics. *Human Factors*, *44*, 1–17.
Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–135.
Hergeth, S., Lorenz, L., Krems, J. F., & Toenert, L. (2015). Effects of take-over requests and cultural background on automation trust in highly automated driving. Proceedings of *the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Salt Lake City, UT*, 331–337.
Hollander, T. D., & Wogalter, M. S. (2000). Connoted hazard of voiced warning signal words: An examination of auditory components. Proceedings of *the Human Factors and Ergonomics Society Annual Meeting, 44*, 702–705.
Ioannidis, J. P. (2005). Why most published research findings are false. *PLOS Medicine, 2*, e124.
Ioannidis, J., & Doucouliagos, C. (2013). What’s to know about the credibility of empirical economics? *Journal of Economic Surveys, 27*, 997–1004.
Jang, P. S. (2007). Designing acoustic and non-acoustic parameters of synthesized speech warnings to control perceived urgency. *International Journal of Industrial Ergonomics, 37*, 213–223.
Keren, G. (1993). Between-or within-subjects design: A methodological dilemma. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behaviorial sciences: Methodological issues* (pp. 257–272). Hillsdale, NJ: Erlbaum.
Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., ... & Cemalcilar, Z. (2014). Investigating variation in replicability: A “many labs” replication project. *Social Psychology, 45*, 142–152.
Kuchar, J. K., & Yang, L. C. (2000). A review of conflict detection and resolution modeling methods. *IEEE Transactions on Intelligent Transportation Systems, 1*, 179–189.
Kyriakidis, M., Happee, R., & De Winter, J. C. F. (2015). Public opinion on automated driving: Results of an international questionnaire among 5,000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour, 32*, 127–140.
Langlois, S., Suied, C., Lageat, T., & Charbonneau, A. (2008). Cross cultural study of auditory warnings. Proceedings of *the 14th International Conference on Auditory Display (ICAD), Paris, France*.
Large, D. R., & Burnett, G. E. (2013). Drivers’ preferences and emotional

responses to satellite navigation voices. *International Journal of Vehicle Noise and Vibration*, *9*, 28–46.
Lerner, N., Singer, J., Kellman, D., & Traube, E. (2015). In-vehicle noise alters the perceived meaning of auditory signals. Proceedings of *the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Salt Lake City, UT, 401–407.
Ljungberg, J. K., Parmentier, F. B., Hughes, R. W., Macken, W. J., & Jones, D. M. (2012). Listen out! Behavioural and subjective responses to verbal warnings. *Applied Cognitive Psychology, 26*, 451–461.
Ljungberg, J. K., & Parmentier, F. (2012). The impact of intonation and valence on objective and subjective attention capture by auditory alarms. *Human Factors, 54*, 826–837.
Machado, S., Duarte, E., Teles, J., Reis, L., & Rebelo, F. (2012). Selection of a voice for a speech signal for personalized warnings: the effect of speaker's gender and voice pitch. *Work, 41*, 3592–3598.
Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist, 70*, 487–498.
Meng, F., & Spence, C. (2015). Tactile warning signals for in-vehicle systems. *Accident Analysis & Prevention, 75*, 333–346.
Merat, N., & Jamson, A. H. (2009). How do drivers behave in a highly automated car. Proceedings of *the 5th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Big Sky, MT, 514–521.
Mok, B., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., & Ju, W. (2015). Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles. Proceedings of *the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Canary Islands, Spain*, 2458–2464.
Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*, 73–97.
Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 28*, 111–131.
Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (Eds), *Phonology and Second Language Acquisition* (pp. 193–218). Philadelphia: John Benjamins.
Naujoks, F., Mai, C., & Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. In T. Ahram, W. Karwowski, & T. Marek (Eds.), Proceedings of *the 5th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Kraków, Poland.
Nilsson, J., Strand, N., Falcone, P., & Vinter, J. (2013). Driver performance in the presence of adaptive cruise control related failures: Implications for safety analysis and fault tolerance. Proceedings of *the 43rd Annual IEEE/IFIP Conference on Dependable Systems and Networks Workshop, Budapest, Hungary*.
Nixon, C. W., Morris, L. J., McCavitt, A. R., McKinley, R. L., Anderson, T. R., McDaniel, M. P., & Yeager, D. G. (1998). Female voice communications in high levels of aircraft cockpit noises–part I: spectra, levels, and

microphones. *Aviation, Space, and Environmental Medicine*, *69*, 675–683.
Noyes, J. M., Hellier, E., & Edworthy, J. (2006). Speech warnings: a review. *Theoretical Issues in Ergonomics Science*, *7*, 551–571.
Oedegaarde (2015). Tesla Autopilot test (EU version) - on country road, highway and in town. Available at https://www.youtube.com/watch?v=7YVNW7-IFx8
Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716.
Özkan, T., & Lajunen, T. (2007). The role of personality, culture, and economy in unintentional fatalities: An aggregated level analysis. *Personality and Individual Differences, 43*, 519–530.
Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230–253.
Park, K. S., & Jang, P. S. (1999). Effects of synthesized voice warning parameters on perceived urgency. *International Journal of Occupational Safety and Ergonomics, 5*, 73–95.
Petermeijer, S., De Winter, J. C. F., & Bengler, K. (2016). Vibrotactile displays: A survey with a view on highly automated driving. *IEEE Transactions on Intelligent Transportation Systems, 17*, 897–907.
Pfromm, M., Khan, R., Oppelt, S., Abendroth, B., & Brudera, R. (2015). Investigation of take-over performance of driving tasks by the driver due to system failure of semi-automated and assisted driving. Proceedings of *19th Triennial Congress of the IEA*, Melbourne, Australia.
Podlipský, V. J., Šimáčková, Š., & Petráž, D. (2016). Is there an interlanguage speech credibility benefit? *Topics in Linguistics, 17*(1), 30–44.
Politis, I., Brewster, S., & Pollick, F. (2015). Language-based multimodal displays for the handover of control in autonomous cars. Proceedings of *the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Application*, Nottingham, UK.
Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology, 299*, 172–179.
Smith, L. E., & Rafiqzad, K. (1979). English for cross-cultural communication: The question of intelligibility. *TESOL Quarterly, 13*, 371–380.
SAE International (2016). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles (Standard No. J3016).
Stewart, N., Ungemach, C., Harris, A. J., Bartels, D. M., Newell, B. R., Paolacci, G., & Chandler, J. (2015). The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. *Judgment and Decision Making*, *10*, 479–491.
Stroebe, W. (2016). Are most published social psychological findings false? *Journal of Experimental Social Psychology, 66*, 134–144.
Wagenmakers, E. J., Verhagen, J., Ly, A., Bakker, M., Lee, M. D., Matzke, D., ... & Morey, R. D. (2015). A power fallacy. *Behavior Research Methods, 47*, 913–917.
Wogalter, M. S., & Silver, N. C. (1995). Warning signal words: connoted strength and understandability by children, elders, and non-native English speakers. *Ergonomics, 38*, 2188–2206.
Wogalter, M. S., Conzola, V. C., Smith-Jackson, T. L. (2002). Research-based guidelines for warning design and evaluation. *Applied Ergonomics, 33*,

219–230.
World Health Organization (2015). *WHO global status report on road safety 2015.* Geneva, Switzerland: World Health Organization.
Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident Analysis & Prevention, 78*, 212–221.

3.3 Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop

Corresponding article: Petermeijer, S. M., Bazilinskyy, P.*, Bengler, K., & De Winter, J. C. F. (2017). Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. Applied Ergonomics, 62, 204–215.
* Joint first author.

3.3.1 Abstract

When a highly automated car reaches its operational limits, it needs to provide a take-over request (TOR) in order for the driver to resume control. The aim of this simulator-based study was to investigate the effects of TOR modality and left/right directionality on drivers' steering behaviour when facing a head-on collision without having received specific instructions regarding the directional nature of the TORs. Twenty-four participants drove three sessions in a highly automated car, each session with a different TOR modality (auditory, vibrotactile, and auditory-vibrotactile). Six TORs were provided per session, warning the participants about a stationary vehicle that had to be avoided by changing lane left or right. Two TORs were issued from the left, two from the right, and two from both the left and the right (i.e., nondirectional). The auditory stimuli were presented via speakers in the simulator (left, right, or both), and the vibrotactile stimuli via a tactile seat (with tactors activated at the left side, right side, or both). The results showed that the multimodal TORs yielded statistically significantly faster steer-touch times than the unimodal vibrotactile TOR, while no statistically significant differences were observed for brake times and lane change times. The unimodal auditory TOR yielded relatively low self-reported usefulness and satisfaction ratings. Almost all drivers overtook the stationary vehicle on the left regardless of the directionality of the TOR, and a post-experiment questionnaire revealed that most participants had not realized that some of the TORs were directional. We conclude that between the three TOR modalities tested, the multimodal approach is preferred. Moreover, our results show that directional auditory and vibrotactile stimuli do not evoke a directional response in uninstructed drivers. More salient and semantically congruent cues, as well as explicit instructions, may be needed to guide a driver into a specific direction during a take-over scenario.

3.3.2 Introduction

3.3.2.1 Highly automated driving and the importance of take-over requests

Research in automated driving is on the rise. Many car manufacturers, OEMs, universities, and federal research institutes are now developing and testing

automated driving systems. There appears to be a consensus that fully automated cars will be prevalent on public roads by the year 2030 (Kyriakidis, Happee, & De Winter, 2015; Underwood, 2014), yet some experts have argued that it will take many more decades before fully automated driving becomes ubiquitous (Shladover, 2015).

Before full automation ('level 5 automation'; SAE International, 2014) is technically feasible, conditional ('level 3') and high ('level 4') automation will probably be deployed. At levels 3 and 4, the automation is not perfectly capable and reliable, meaning that the driver will sometimes have to take back control. If the automation recognizes that it is unable to handle a traffic situation, it provides a warning, also called a take-over request (TOR).

The take-over process is an important topic in human factors research. A substantial number of researchers have studied how drivers behave after receiving a TOR (Clark & Feng, 2015; Gold, Damböck, Lorenz, & Bengler, 2013; Lorenz, Kerschbaum, & Schumann, 2014; Louw, Merat, & Jamson, 2015; Merat, Jamson, Lai, Daly, & Carsten, 2014; Mok et al., 2015; Petermann-Stock, Hackenberg, Muhr, & Mergl, 2013; Payre, Cestac, & Delhomme, 2016; Telpaz, Rhindress, Zelman, & Tsimhoni, 2015; Walch, Lange, Baumann, & Weber, 2015; Zeeb, Buchner, & Schrauf, 2015; for reviews see De Winter, Happee, Martens, & Stanton, 2014; Lu, Happee, Cabrall, Kyriakidis, & De Winter, 2016). The time buffer within which the driver has to perform a steering manoeuvre or a braking action can range from long (e.g., upcoming highway exit) to short (e.g., accident happening in front of the vehicle). In emergency scenarios, in which the time buffer is short, it is important that the TOR causes the driver to resume control as quickly and safely as possible. For example, the automation may provide a take-over request when it cannot handle an impending collision, such as when a stationary obstacle is present on the road. It is then up to the driver to take over control and execute a proper maneuver, such as to evade a stationary object on the left or right.

3.3.2.2 Auditory and vibrotactile TORs

In manual driving, information is typically presented to the driver via visual displays (e.g., low fuel indicator) and auditory displays (e.g., navigational instructions). Auditory and vibrotactile displays have the advantage over visual displays of being 'gaze-free', which means that the information can be detected by the driver irrespective of head or eye position (Meng & Spence, 2015; Stanton & Edworthy, 1999; Stokes, Wickens, & Kite, 1990). During highly automated driving, the driver is likely to be occupied with non-driving activities. Therefore, auditory and vibrotactile stimuli are promising as TORs (Bazilinskyy & De Winter, 2015).

Vibrations are a relatively underused modality in the automotive industry (Meng & Spence, 2015) but are gaining interest (for a review, see Petermeijer, Abbink, Mulder, & De Winter, 2015). For example, BMW (2015) and Mercedes-Benz (2015) have introduced a vibrating steering wheel for lane departure warnings, whereas Citroën (2007) and Chevrolet (General Motors, 2014) have introduced a lane departure warning system that provides vibrations in the driver seat. Compared to vibrations on the steering wheel, seat vibrations are a promising means of conveying TORs to the driver, because the driver of an

automated car will usually be in contact with the seat but not with the steering wheel (Petermeijer, De Winter, & Bengler, 2016).

3.3.2.3 Multimodal feedback

Psychophysics research has shown that multimodal warnings (i.e., combinations of visual, auditory, and vibrotactile stimuli) are perceived as more urgent than their unimodal constituents (e.g., Van Erp, Toet, & Janssen, 2015). In a self-report questionnaire among 1,692 respondents investigating the public opinion on visual, auditory, and vibrotactile displays during highly automated driving, it was found that people are more likely to prefer a multimodal TOR when the urgency of the takeover is higher (Bazilinskyy, Petermeijer, Petrovych, Dodou, & De Winter, 2018). Consistent with these findings, driving simulator research has shown that a combination of visual, auditory, and vibrotactile TORs led to higher perceived urgency and perceived alerting effectiveness than the corresponding unimodal warnings (Politis, Brewster, & Pollick, 2014).

Multimodal warnings not only enhance subjective urgency, but also elicit faster reaction times than unimodal warnings. Burke et al. (2006) found in a meta-analysis of 43 studies on various types of human-machine interaction that visual-auditory and visual-tactile feedback yield faster reaction times than visual feedback alone. An experimental study by Diederich and Colonius (2004) found that trimodal stimuli (vibration, light, & tone) consistently evoked faster reaction times than bimodal stimuli, which in turn were faster than unimodal ones. Additionally, a review by Spence and Santangelo (2009) concluded that multimodal stimuli are more effective in capturing a person's attention than unimodal ones, especially when the person is engaged in a concurrent attention-demanding task.

Although the benefits of multimodal feedback are well established, such benefits are not necessarily obtained in manual driving with a driver assistance system. A driving simulator study by Tijerina et al. (1996) on a lane departure warning system concluded that a bimodal auditory-vibrotactile display "may be a source of overload to a driver", whereas a study in a driving simulator by Lees, Cosman, Lee, Vecera, Dawson, and Rizzo (2012) found that bimodal auditory-visual cues yielded higher reaction times than auditory-only cues. In another simulator study investigating warnings when Adaptive Cruise Control (ACC) exceeded its functional limits, Lee, McGehee, Brown and Marshall (2006) observed brake reaction times that were 400 ms slower for a combination of a visual warning, auditory warning, vibratory seat, and brake pulse feedback compared to a visual-auditory warning. A detrimental effect of a multimodal warnings may occur when the cues from the different sources are semantically, temporally, and/or spatially incongruent, as a result of which they are perceived as a series of cues rather than a single cue (Diaconescu, Alain, & McIntosh, 2011; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010).

3.3.2.4 Directional warnings

In most of the available research in automated driving, the TORs are provided in a nondirectional manner, meaning that the warning is used for alerting the driver without conveying any extra information. In particular, many studies on the take-over process have used a nondirectional auditory TOR (e.g., a double beep) often in combination with a nondirectional visual notification (e.g., Damböck,

2013; Gold et al., 2013; Melcher, Rauh, Diederichs, Widlroither, & Bauer, 2015; Naujoks, Mai, & Neukum, 2014; Naujoks, Purucker, Neukum, Wolter, & Steiger, 2015; Radlmayr, Gold, Lorenz, Farid, & Bengler, 2014).

Several researchers have demonstrated the potential of directional warnings in manual car driving, whereby the location of the warning signal indicates a location or direction to which the driver needs to focus his/her attention (Weller, Heyne, Feige, Bretschneider, Oeser, & Schlag, 2013; Zarife, 2014; Zhang, Yan, & Yang, 2015). For example, Gray, Ho, and Spence (2014) tested a forward collision warning system that used vibrotactile stimili that were linked to the closing velocity and which travelled upward or downward on the human body using three tactors. Nukarinen, Rantala, Farooq, and Raisamo (2015) tested left/right directional visual cues versus directional vibrotactile cues provided via eyeglasses and the driver seat in a simulated lane change test. Their results showed that the vibrotactile cues yielded faster response times than the visual cues. Schwalk, Kalogerakis, and Maier (2015) provided dynamic directional vibrotactile cues via the driver seat, which participants rated as appropriate for TORs, whereas Telpaz, Rhindress, Zelman, and Tsimhoni (2015) provided seat vibrations that made drivers aware of surrounding traffic during automated driving. More generally, research in a variety of applications areas has demonstrated the effectiveness of visual, auditory, and vibrotactile directional cues regarding reaction times and situation awareness (Houtenbos et al., 2017; Naujoks & Neukum, 2014; Prewett, Elliott, Walvoord, & Coovert, 2012).

One issue in the design of left/right directional warnings is that a distinction can be made between an ipsilateral mapping, requiring the driver to steer in the direction of the stimulus (i.e., steer towards the right when the stimulus comes from the right), and a contralateral mapping, requiring the driver to steer away from the direction of the stimulus. Early studies investigating directional cueing in abstract laboratory environments found that ipsilateral mapping yields faster reaction times, a phenomenon also known as the spatial stimulus-response compatibility effect (e.g., Simon, Hinrichs, & Craft, 1970; Umiltá & Nicoletti, 1990). However, in realistic driving scenarios, in which there is a dangerous situation and the driver is able to visually assess the driving scene before responding, a contralateral mapping has been found to yield faster reaction times (Beruscha, Wang, Augsburg, & Wandke, 2010; Wang, Pick, Proctor, & Ye, 2007; Müsseler, Aschersleben, Arning, & Proctor, 2009; Straughn, Gray, & Tan, 2009). This is also the approach used in the majority of lane departure warning systems (Meng & Spence, 2015).

Manual and automated driving are different with respect to the role of the driver, and to our knowledge, the effects of directional auditory or vibrotactile warnings have not been investigated in a highly automated driving context. One particular difference between warnings in manual driving (e.g., lane departure warnings, forward collision warnings) and warnings in automated driving (TORs) is that the latter warnings may occur when the driver is not engaged in the driving task at all. In automated driving, the driver may be performing a distracting non-driving task, and should be able to effectively reclaim control and intuitively interpret the directional feedback within a matter of seconds. At present, it is unknown whether directional auditory or directional vibrotactile cues have the potential to guide a driver towards a left or right direction in a take-over scenario.

3.3.2.5 Aim of the study

The aim of this study was to evaluate drivers' reaction times and self-reported usefulness and satisfaction of unimodal (i.e., auditory or vibrotactile) versus bimodal (i.e., auditory-vibrotactile) TORs. Furthermore, this study aimed to investigate whether the directionality of the TOR (left, right, or nondirectional) evokes a spontaneous ipsi- or contralateral response. That is, we investigated whether uninstructed drivers execute an ipsilateral or a contralateral response in situations where both responses are valid for avoiding a collision. In our experiment, the drivers of the highly automated car were biomechanically, visually, and cognitively engaged with a secondary task (Surrogate Reference Task [SuRT]; ISO/DTS 14198, 2012) prior to receiving the TOR. The consistency of the participants' steering reaction to the directional TOR after this period of distraction is informative about whether directional feedback is effective in guiding action.

3.3.3 Method

3.3.3.1 Participants

Twenty-four participants (16 male; 8 female) holding a driver's license participated in the experiment. The participants were students and employees of the Technical University of Munich, and were between 24 and 35 years old (M = 27.9 years; SD = 3.0 years). Their mean driving experience was 10.1 years (SD = 3.1). Four of the participants reported a mileage of 1–1000 km, 9 participants reported a mileage of 1,001–5,000 km, and 11 participants reported a mileage of 5,001–25,000 km in the past 12 months (Table S5). Twelve participants had participated in more than five previous driving simulator experiments. Eight participants reported wearing glasses or contact lenses while driving a car, and none of the participants reported colour blindness. One participant was left-handed.

3.3.3.2 Simulator

The study was conducted in a fixed-base driving simulator consisting of a complete BMW 6 Series (Fig. 1). The front (ca. 180 deg) and rear views (perceivable via the rear mirrors) of the environment were presented using six LCD projectors (Technical University of Munich, 2015). Mounted LCD screens represented the dashboard and on-board computer. Road noise and engine noise were played back, and low frequency vibrations were provided in the driver seat via a bass speaker. The participant could override the automation by braking or turning the steering wheel. The automation could be engaged and disengaged by pressing a diamond-shaped ACC button on the steering wheel. Pushing the brake pedal with more than 25% depression or steering so that the deviation from lane centre was greater than approximately 0.5 m would also disengage the automation. No visual indication of the automation status was provided.

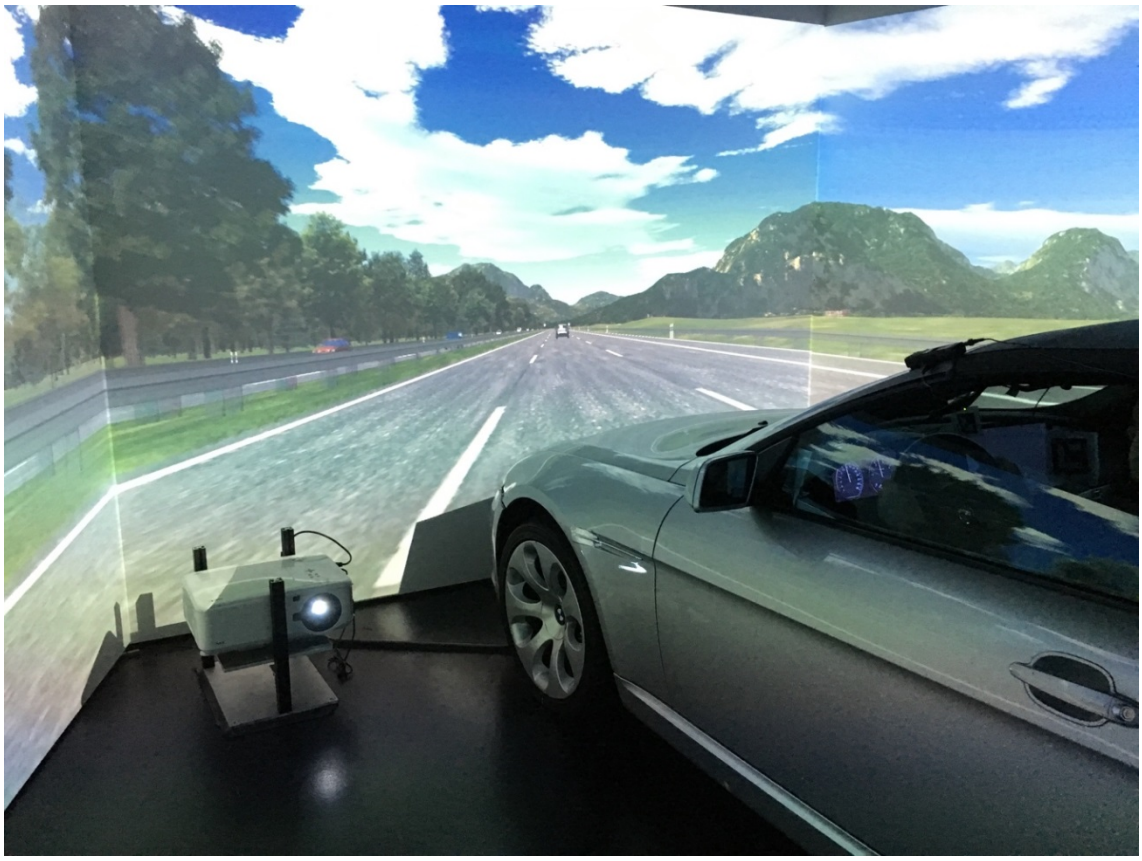


Figure 1. Driving simulator used in the experiment.

3.3.3.3 Independent variables

The independent variable was the type of TOR. The TORs were auditory beeps (A), vibrations in the driver seat (V), or their combination (AV). Both directional and nondirectional TORs were provided. In the directional AV TORs, the beeps and the vibrations were provided from the same side (left or right). The TORs did not include a visual notification, because our aim was to study the effectiveness of auditory and vibrotactile feedback while drivers were visually distracted (see Section 2.5 for a description of the visually demanding secondary task).

An auditory TOR (A) was a single pair of 240 ms beeps of 2,700 Hz, with a 100 ms interval between the two beeps. Directional auditory TORs produced the sound from the in-vehicle sound system. Specifically, the sounds were produced from left or right speakers located in the front and rear doors and a subwoofer located in the upper part of the driver seat. Nondirectional TORs were generated from both speakers simultaneously. The loudness of the auditory TOR was 105 dB (measured with Decibel Ultra iOS application).

Figure 2. Vibration mat on the driver seat of the simulator. Red circles indicate the approximate locations of the vibration motors.

A vibrotactile TOR (V) was a single pair of vibrotactile pulses having a frequency of approximately 60 Hz, using the same temporal pattern as the auditory TOR (pulse duration = 240 ms, interval between pulses = 100 ms). Twelve vibration motors (Pico Vibe 9 mm, model number: 307-103) were configured in two 3 x 2 matrices on the driver seat back and bottom (Fig. 2). Directional TORs were provided by vibrating the left or right column of motors in the seat back and bottom simultaneously, whereas nondirectional TORs were provided by vibrating all 12 motors. The sound produced by the vibration motors was negligible compared to the engine noise and road rumble of the simulation.

3.3.3.4 Driving scenario in experimental sessions

The sessions involved driving on a highway with three 4 m wide lanes. At the beginning of each of the three sessions, the participant's car was positioned in the middle of the three lanes. The participants were asked via the intercom to accelerate to 100 km/h and engage the automation. The automated system controlled both lateral and longitudinal motion at a constant speed of 120 km/h.

In each session, a distance of approximately 21.9 km was driven in about 11.5 min. A total of six stationary cars were positioned in the middle lane, between 3,000 m and 4,000 m apart. Accordingly, the time interval between the TORs was between 1.5 and 2 min. When the participant's car was 223 m in front of the stationary car, a TOR was provided. At a speed of 120 km/h, this implies a lead time of about 7 s (see also Gold et al., 2013). All TORs were provided on straight road segments. A video illustration of a TOR is provided as supplementary material.

During the entire experiment, there was no other traffic in the participants' direction of travel, so that the participant could avoid the stationary vehicle by changing to either one of the adjacent lanes. Consequently, either an ipsilateral or contralateral response to a directional stimulus were valid. No collisions with objects in the simulated environment (e.g., guardrails or other vehicles) were possible, which means that the participants could drive through these objects.

3.3.3.5 Secondary task

The participants were asked to perform the SuRT shown on a 14-inch tablet on the central console, at the position of the car radio. The SuRT is a self-paced task that requires visual search and manual input. The task of the participant was to identify a circle that is larger (target) among other smaller circles (distractors). The participant used a keypad, located next to the handbrake, to select the column of circles that contained the target circle. The target size was 14 mm, the distractor size 12 mm, and the number of columns was 6.

3.3.3.6 Procedures and instructions to participants

All participants were given an instructions form describing that the purpose of the experiment was to investigate driving behaviour, subjective experience, and workload for three types of TOR in a highly automated vehicle. Moreover, the form explained that the participants would drive three 12-min sessions, each session with a different TOR (sounds, vibrations, or sound and vibrations combined). The form also introduced the SuRT.

The participants completed an intake questionnaire, after which they proceeded to the simulator, where they adjusted the seat, steering wheel, and mirrors according to their liking. The participants were verbally told how to (dis)engage the automation and how to perform the SuRT. Furthermore, they were asked to focus on the secondary task during automated mode. Participants were instructed that when a TOR was provided, they had to take the steering wheel with both hands and avoid the obstacle. Additionally, they were asked to return to the centre lane and re-engage the automation after having taken over the stationary car. The experimenter also asked the participants to behave as if driving on a real highway in a real car.

After receiving these instructions, participants drove a 3-min familiarisation session in which they practiced how to control the car, (dis)engage the automation, and perform the SuRT. In this session, one nondirectional AV TOR was provided. The participants were not informed that they would be exposed to directional TORs in the forthcoming sessions.

The participants drove three experimental sessions, with one TOR modality (A, V, or AV) per session. Each session featured six TORs: four directional (two from the left, two from the right) and two nondirectional ones. The three conditions (A, V, AV) as well as the directionality of the TORs within a session were randomized between the participants.

Each session was followed by a break of up to 5 min outside the simulator. During this break, the participants filled out a questionnaire on usefulness and satisfaction (Van der Laan, Heino, & De Waard, 1997) and a NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). The participants were asked to consider only the TORs (which were presented in either the A, V, or AV modality) when answering the usefulness/satisfaction questionnaire, and

all activities during the session when answering the NASA-TLX. After the third experimental session, a post-experiment questionnaire was filled out about participants' preference, perceived urgency, and perceived directionality of the TORs they received during the experiment. The experiment took approximately 1 hour per participant to complete.

3.3.3.7 Dependent variables

3.3.3.7.1 Objective measures

The direction (left or right) on which participants overtook the stationary vehicle was used to assess whether participants chose that direction more likely based on the directionality of the TOR. Furthermore, reaction times were calculated to assess how quickly participants provided steering or braking input after receiving a TOR. The following reaction times were calculated:

(1) Steer touch: absolute steering wheel velocity greater than 1 deg/s. During automated driving, the steering wheel hardly moves. An absolute steering velocity of 1 deg/s was the minimum value which could be reliably attributed to human input. Accordingly, the steer-touch reaction time was regarded as a measure of how quickly participants touched the steering wheel after receiving the TOR.
(2) Steer initiate: absolute steering wheel angle greater than 0.25 deg. This 0.25 deg threshold represents the minimum that could be reliably detected by the steering sensor as being different from the steering angles that were measured during automated driving. This measure may represent the initiation of a steering action or stabilization movement. Out of the 376 registered lane changes, 287 (i.e., 74%) lane changes were made in the same direction as the steering wheel angle when it first exceeded the 0.25 deg threshold.
(3) Steer turn: absolute steering wheel angle greater than 2 deg. The 2 deg threshold was used to represent the initiation of a 'conscious' steering action (Gold et al., 2013). Gold et al. reported that steering angles under this threshold are used to stabilize the vehicle and do not generate notable acceleration forces. The fact that steering actions greater than 2 deg correspond to conscious steering actions was confirmed by the results of the present experiment. Out of 376 registered lane changes, 371 were made in the same direction as the direction of the steering wheel when it first exceeded a 2 deg threshold after a take-over request. In other words, participants made a steering correction in the opposite direction in only 1% of the cases (5/376).
(4) Car avoid: absolute deviation from the lane centre greater than 1.00 m.
(5) Lane change: absolute deviation from the lane centre greater than 2.00 m.
(6) Brake: pedal depression greater than 0%. Similar to the steer-touch reaction time, the brake reaction time represents the initial movement of the brake pedal.
(7) Steer touch or brake: the minimum of the brake time and steer-touch time.

3.3.3.7.2 Self-report measures

Table S1 shows the intake questionnaire and the coding of the responses. The first part of the questionnaire (Q1–Q13) contained general questions about gender, age, driving experience, accident history, vision quality, and handedness. The second part (Q14–Q20) measured the participant's driving style using the violations scale of the Driver Behaviour Questionnaire (DBQ) as in De Winter (2013). Q21 asked about past experience with driving simulator experiments, and the remaining questions (Q22–Q25) polled the participants' preference and perceived urgency of auditory TORs, vibrotactile TORs, visual TORs, and combinations thereof. The intake questionnaire was offered in digital form.

A questionnaire on usefulness and satisfaction (Van der Laan et al., 1997) was offered in paper format. The mean usefulness score was determined across the following five items: 1. useful–useless; 3. bad–good; 5. effective–superfluous; 7. assisting–worthless; and 9. raising alertness–sleep-inducing. The mean satisfaction score was determined from the following four items: 2. pleasant–unpleasant; 4. nice–annoying; 6. irritating–likeable; and 8. undesirable–desirable. All items were on a five-point Likert scale. Sign reversals were conducted for items 1, 2, 4, 5, 7, and 9, so that a higher score indicates higher usefulness/satisfaction.

The NASA-TLX questionnaire included the following six aspects of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. All aspects were marked on a 21-tick horizontal bar with anchors on the left (0% = very low) and right (100% = very high) sides. For the performance item, the anchors (0% = perfect) and (100% = failure) were used. The questionnaire was offered in an online software application provided by Sharek (2011).

Table S2 shows the questions of the post-experiment questionnaire and the corresponding coding of the responses. Q2–Q5 were identical to Q22–Q25 of the intake questionnaire, in order to detect eventual changes in the participants' perception on the different TOR modalities after these were experienced during the experiment. The post-experiment questionnaire was offered in digital form.

3.3.3.8 Statistical analyses

Comparisons of the independent variables between the conditions were conducted using a repeated measures analysis of variance. In addition, paired comparisons between the three conditions (A vs. V, A vs. AV, & V vs. AV) were conducted using paired t-tests, with a significance level of 0.01. A low significance level was used to minimize the probability of false positives.

3.3.4 Results

3.3.4.1 Missing data

The driving simulator did not store data for the first two participants, and during the A and V conditions for the third participant. For one participant, the AV condition was excluded because the automation drove at a speed that was different from the target speed of 120 km/h. Therefore, the effective sample size for all conditions was 21. Additionally, for one participant in the AV condition, the

third take-over manoeuvre was excluded because the automation did not drive in the middle lane. Moreover, for one participant in the AV condition, the first take-over manoeuvre was excluded because the participant was already pressing the brake when the TOR was provided. The total number of take-overs included in the analysis was therefore 126, 126, and 124 for the A, V, and AV conditions, respectively.

3.3.4.2 Take-over direction

Table 1 shows that almost all participants overtook the car on the left side, regardless of the direction of the TOR and regardless of TOR modality (A, V, or AV). Figure 3 shows the mean deviation from the lane centre for all take-overs, separated into overtaking manoeuvres on the left lane versus the right lane. The black vertical line represents the location of the stationary vehicle. It can be seen that in the AV condition the drivers increased their lateral position slightly earlier than in the two unimodal conditions.

Table 1. Number of take-overs as a function of take-over request (TOR) modality, TOR direction, and participant's lane change direction

TOR direction	Participant's lane change	A	V	AV
Left	Right	3	2	5
None	Right	4	3	6
Right	Right	4	2	7
Left	Left	39	40	37
None	Left	38	39	34
Right	Left	38	40	35
	Total	126	126	124

Note. A = Auditory, V = Vibrotactile, AV = Auditory & vibrotactile

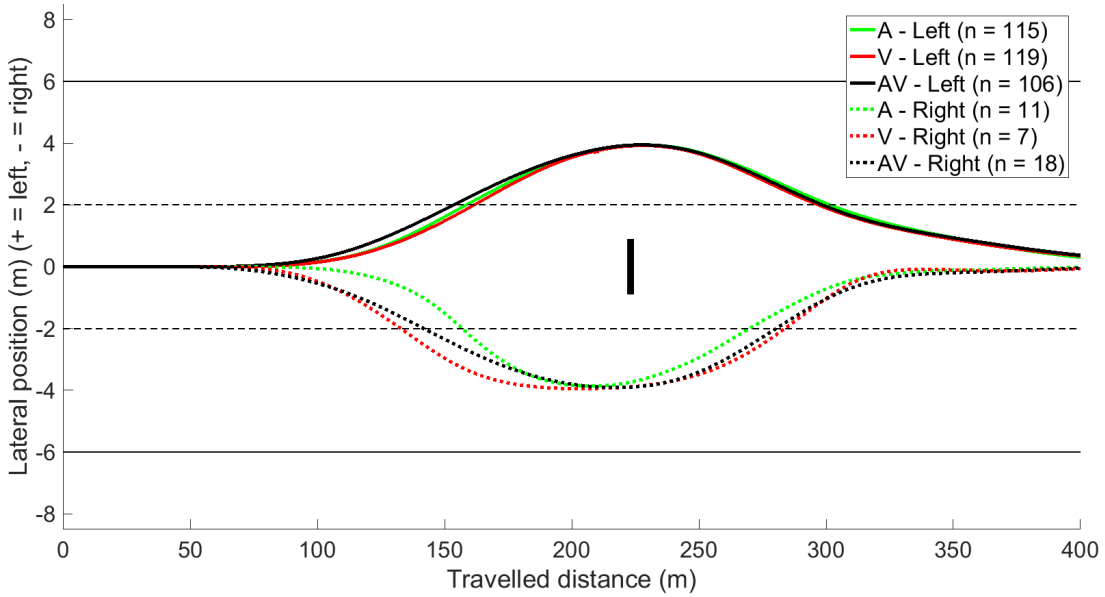


Figure 3. Mean deviation from the lane centre across all take-overs, for left and right lane changes. A = Auditory, V = Vibrotactile, AV = Auditory-vibrotactile.

3.3.4.3 Reaction times

The participants successfully evaded the stationary obstacle in all cases. The steer-touch and steer-initiate reactions were faster for the AV condition than for the V condition (Table 2). Figure 4 shows the mean steering wheel angle across all take-overs, distinguishing between TORs in which the participant braked (bottom figure) or not (top figure). It can be seen that the steering reaction in the AV condition starts earlier compared to the unimodal ones, which is in line with the results in Table 2 and Figure 3. There were no statistically significant differences between the three conditions regarding the reaction times for steer-turn, avoiding the car, changing lanes, and braking.

Table 2. Mean reaction times with standard deviations in parentheses (N = 21), results of the repeated measures ANOVA, and results of paired comparisons using t tests.

	A ***M* (*SD*)**	**V** ***M* (*SD*)**	**AV** ***M* (*SD*)**	**Repeated measures ANOVA**	**A vs. V p**	**A vs. AV p**	**V vs. AV p**
Steer touch or brake (s)	1.69 (0.39)	1.80 (0.38)	1.57 (0.38)	$F(2,38) = 5.53$, $p = 0.008$	0.013	0.221	0.005
Steer touch (s)	1.83 (0.43)	1.92 (0.46)	1.67 (0.48)	$F(2,38) = 5.08$, $p = 0.011$	0.151	0.117	0.002
Steer initiate (s)	2.00 (0.47)	2.03 (0.43)	1.80 (0.49)	$F(2,38) = 4.22$, $p = 0.022$	0.596	0.073	0.009
Steer turn (s)	2.91 (0.86)	2.90 (0.89)	2.67 (0.95)	$F(2,38) = 2.73$, $p = 0.078$	0.876	0.064	0.047
Brake (s)	1.91 (0.45)	1.88 (0.25)	1.87 (0.46)	$F(2,14) = 1.16$, $p = 0.342$	0.191	0.939	0.276
Car avoid (s)	4.57 (0.94)	4.56 (1.08)	4.27 (1.13)	$F(2,38) = 1.93$, $p = 0.159$	0.984	0.098	0.124
Lane change (s)	5.22 (0.94)	5.20 (1.08)	4.96 (1.19)	$F(2,38) = 1.22$, $p = 0.307$	0.882	0.175	0.230

Note. A = Auditory, V = Vibrotactile, AV = Auditory & vibrotactile. $N = 10, 8, 10$ for the brake reaction time for the A, V, and AV conditions, respectively. The repeated measures ANOVA was performed for the 20 participants (8 participants for the brake reaction time) for whom data were available for each of the three conditions. For the paired comparisons, p-values smaller than 0.01 are listed in boldface.

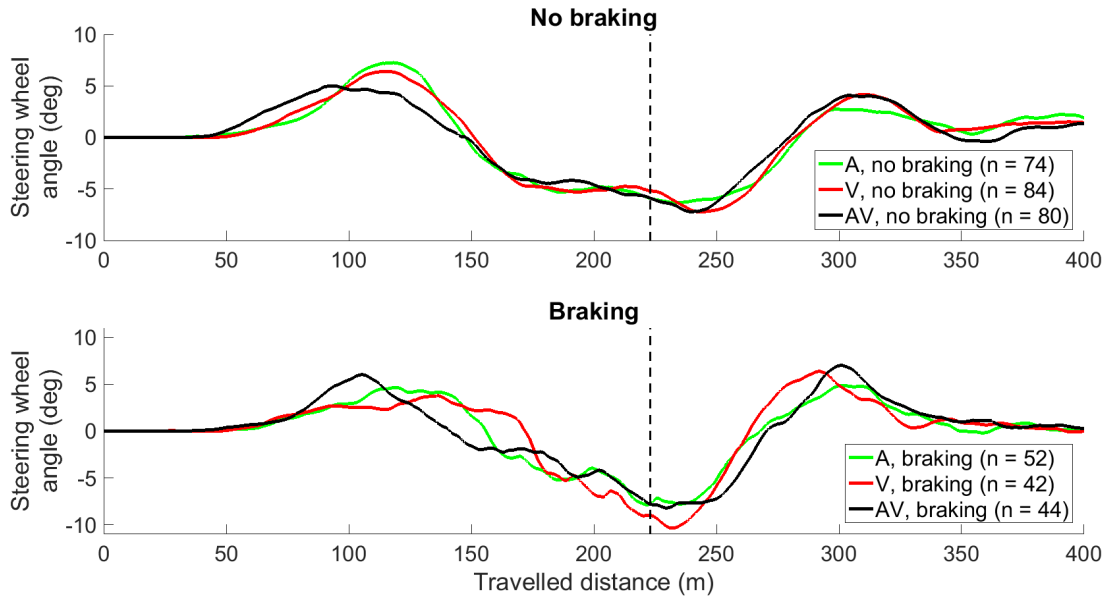


Figure 4. Mean steering wheel angle across all take-overs for take-overs in which the driver braked (top) or did not brake (bottom). If the participant overtook the car on the right, the sign of the steering wheel angle was reversed. The dotted black line represents the location of the stationary vehicle. A = Auditory, V = Vibrotactile, AV = Auditory-vibrotactile.

There were no statistically significant differences between the steer-touch reaction time for 'congruent' responses (i.e., TOR from the left & overtake on the left) and 'incongruent' responses (i.e., TOR from the right & overtake on the left). Specifically, the mean steer-touch reaction times were 1.77 s (SD = 0.50), 1.92 s (SD = 0.53), and 1.70 s (SD = 0.43) for congruent responses, and 1.91 s (SD = 0.50), 1.89 s (SD = 0.53), and 1.67 s (SD = 0.43) for incongruent responses, for the A, V, and AV conditions, respectively (p = 0.298, 0.818, 0.768, for congruent vs. incongruent responses per condition, respectively).

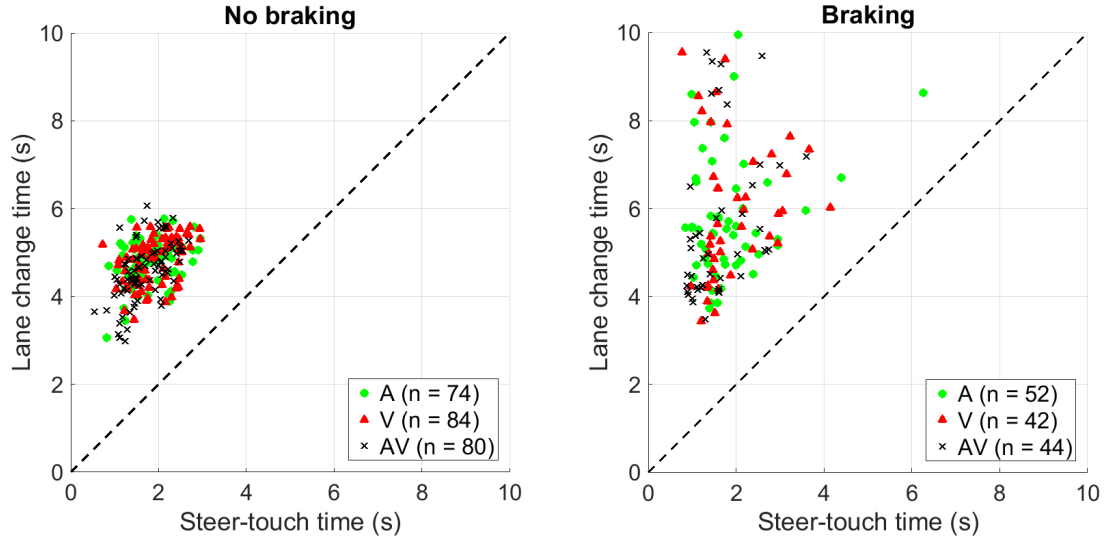


Figure 5. Scatter plot of lane change time versus steer-touch time, distinguishing between trials in which the participant did not brake (left panel) and trials in which the participants braked (right panel). A = Auditory, V = Vibrotactile, AV = Auditory-vibrotactile.

As a supplementary analysis, Figure 5 shows a scatter plot of the lane change versus steer-touch reaction times for take-overs that involved braking versus no braking. It can be seen that the reaction times in which no braking was involved had considerably smaller standard deviations than the reaction times with braking (SD steer-touch time = 0.49 vs. 0.81 s; SD lane change time = 0.58 vs. 1.57 s). The results in Figure 5 illustrate that by braking, participants 'buy time' in order to resolve the conflict. Moreover, a learning effect was observed for the percentage of TORs that involved braking and for the mean maximum brake position across TORs, as well as for the reaction time measures (Fig. 6).

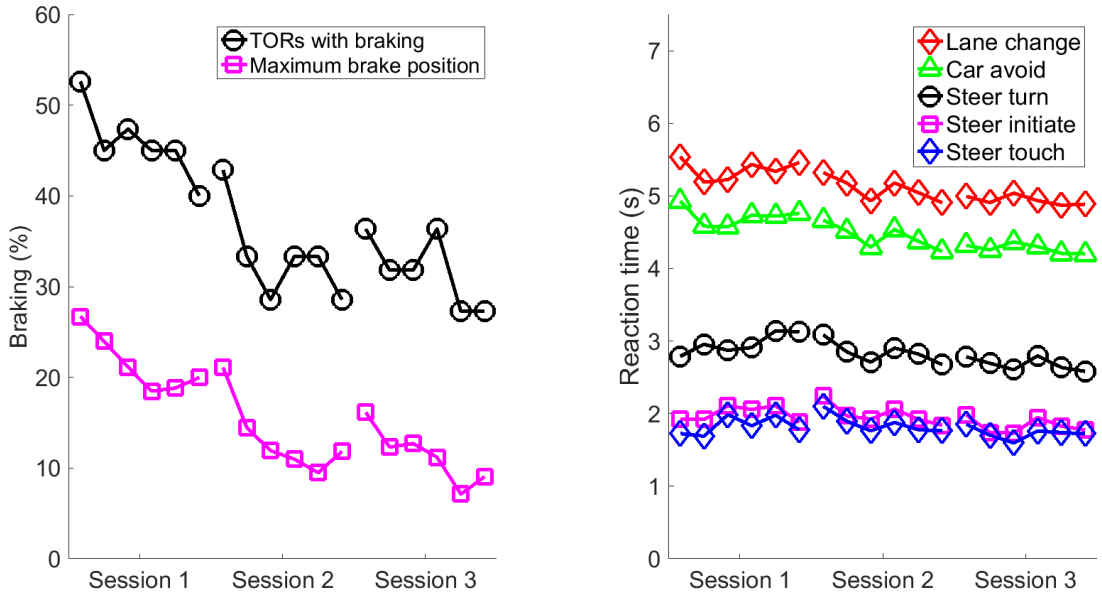


Figure 6. Left: Percentage of take-over requests (TORs) that involved braking and mean maximum brake position across TORs as a function of the TOR number (3 sessions * 6 TORs per session equals 18 TORs per participant). Right: Mean lane change times, mean car avoid times, mean steer turn times, mean steer initiate times, and mean steer touch times as a function of the TOR number. Markers of the six TORs within the same session are connected by a line.

3.3.4.4 Self-report measures

Table 3. Mean NASA Task Load Index (TLX) scores with standard deviations in parentheses (N = 18), results of the repeated measures ANOVA, and results of paired comparisons using t tests.

	A ***M* (*SD*)**	**V** ***M* (*SD*)**	**AV** ***M* (*SD*)**	
Mental demand (%)	38 (26)	37 (29)	34 (24)	
Physical demand (%)	17 (16)	18 (20)	20 (18)	
Temporal demand (%)	35 (25)	31 (26)	32 (23)	
Performance (%)	26 (18)	27 (18)	23 (15)	
Effort (%)	36 (25)	35 (26)	29 (21)	
Frustration (%)	19 (17)	20 (22)	19 (15)	
				Repeated measures ANOVA
Average (%)	28 (17)	28 (17)	26 (15)	$F(2,34) = 1.38$, $p = 0.265$

Note. A = Auditory, V = Vibrotactile, AV = Auditory & vibrotactile.

Table 3 shows the mean NASA-TLX scores with standard deviations in parentheses. No statistically significant differences between the three modalities of TORs were observed. Table 4 presents the mean scores of the usefulness/satisfaction questionnaire with standard deviations in parentheses. The bimodal TORs (AV) were more highly rated in terms of usefulness than the unimodal TORs (A & V). Furthermore, A was considered more annoying and unpleasant than V.

Table 4. Mean usefulness and satisfaction scores with standard deviations in parentheses, results of the repeated measures ANOVA, and results of paired comparisons using t tests.

Negative (-2)	**Positive (+2)**	**A** ***M* (*SD*)**	**V** ***M* (*SD*)**	**AV** ***M* (*SD*)**
Useless	Useful	1.17 (0.65)	1.04 (0.81)	1.63 (0.49)
Unpleasant	Pleasant	–0.09 (0.90)	1.00 (0.88)	0.50 (0.72)
Bad	Good	0.65 (0.78)	0.88 (0.85)	1.00 (0.78)
Annoying	Nice	–0.17 (0.98)	0.83 (1.09)	0.25 (0.85)
Superfluous	Effective	1.26 (0.81)	0.92 (0.88)	1.50 (0.51)
Irritating	Likeable	0.22 (0.80)	0.58 (0.88)	0.29 (0.91)
Worthless	Assisting	1.09 (0.79)	1.29 (0.62)	1.38 (0.58)
Undesirable	Desirable	0.22 (0.60)	0.71 (0.86)	0.63 (0.77)
Sleep-inducing	Raising Alertness	1.04 (0.93)	0.63 (0.82)	1.38 (0.88)

				Repeated-measures ANOVA	A vs. V p	A vs. AV p	V vs. AV p
Overall usefulness score		1.04 (0.51)	0.95 (0.60)	1.38 (0.43)	$F(2,44) = 4.71$, $p = 0.014$	0.573	< 0.001	0.013
Overall satisfaction score		0.04 (0.64)	0.78 (0.72)	0.42 (0.63)	$F(2,44) = 7.47$, $p = 0.002$	0.001	0.043	0.058

Note. A = Auditory, V = Vibrotactile, AV = Auditory & vibrotactile. N = 23, 24, and 24, respectively. The repeated-measures ANOVA was performed for the 23 participants for whom data were available for each of the three conditions. For the paired comparisons, p-values smaller than 0.01 are listed in boldface.

Figure 7 shows the participants' opinion on the use of A, V, and AV TORs for the intake- and the post-experiment questionnaires. A large shift in opinion was observed for the option "Sound message and vibrations message (in any order)": 5 participants chose this TOR type before the experiment, whereas 12 selected this option after the experiment. None of the participants selected "Visual message", either before or after the experiment. The figure also shows respondents' opinion from a previous online questionnaire study (Bazilinskyy, Kyriakidis, & De Winter, 2015; N = 1,692), where similar questions were asked for a scenario with a high level of urgency: a traffic accident happening in front of a highly automated vehicle. It can be seen that the responses from that

questionnaire were comparable to the results from the intake questionnaire in the current study (Fig. 7).

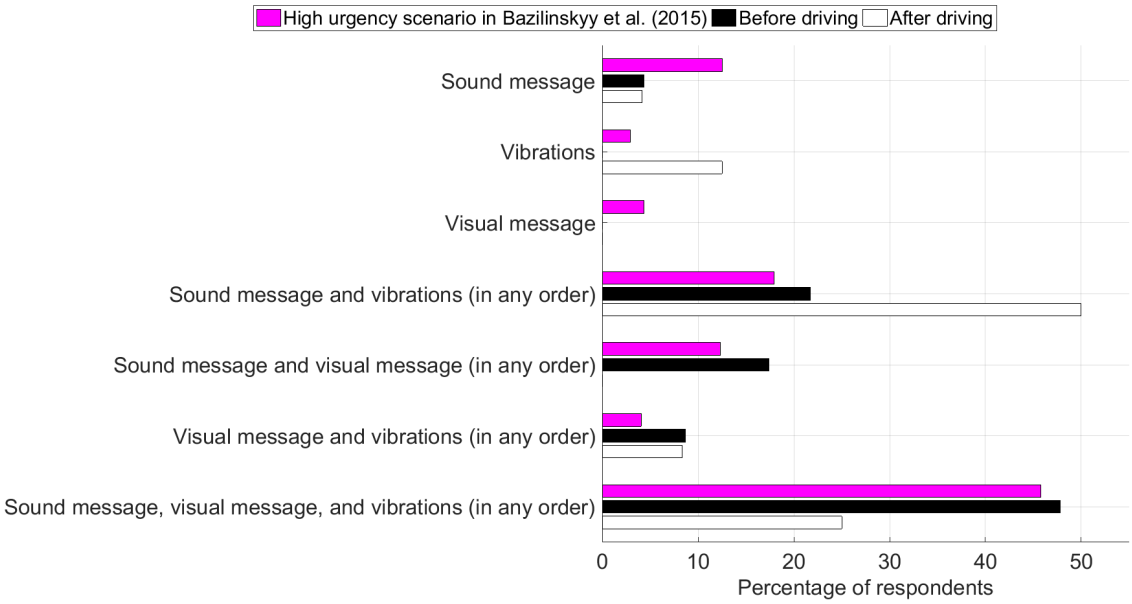


Figure 7. Participants' opinion on various types of take-over requests (TORs) before (*N* = 23; one missing value) and after (*N* = 24) the experiment. Magenta bars represent the respondents' opinion from a previous online questionnaire (Bazilinskyy et al., 2015) for the question "*What take-over request would you like to receive in this scenario?*", where the scenario was described as "You are driving on the highway in the automated mode and you see a traffic accident happening in front of you. You have little time to take over control".

The participants were also asked both before and after the experiment whether they considered A, V, and AV TORs as urgent (1 = Disagree strongly to 5 = Agree strongly). No large differences were found between the pre- and post-experiment responses (Tables S3 and S4). The majority of participants had not perceived that the TORs sometimes came from the left or right: MA = 1.88 and MV = 2.00, on a scale from 1 = Disagree strongly to 5 = Agree strongly.

3.3.5 Discussion

The aim of this study was to investigate the steer reaction time, steering behaviour, and self-reported usefulness and satisfaction scores of auditory, vibrotactile, and bimodal (i.e., auditory-vibrotactile) TORs. We also investigated whether a directional cue (left, right, or nondirectional) of the TORs evoke a consistent contra- or ipsilateral response if drivers are not informed about the presence and meaning of this directional feedback.

The results showed that all types of TOR were effective in ensuring that participants did not collide with the stationary car. Moreover, the results showed that the reaction times for steer, brake, and lane change were not significantly different between the auditory and vibratory tactile modalities. The similarity of the reaction times for auditory and vibrotactile TORs is consistent with basic psychophysical research on simple reaction times (Woodworth & Schlosberg, 1954). Similarly, previous research in manual driving has found that auditory and vibrotactile warnings yielded equivalent reaction times (e.g., Scott & Gray, 2008); however, there have also been cases where auditory warnings (e.g., Ho, Tan, &

Spence, 2006) or vibrotactile warnings (e.g., Mohebbi, Gray, & Tan, 2009; Cao, Van der Sluis, Theune, op den Akker, & Nijholt, 2010) yielded the faster reaction times. It seems that the specific task conditions (e.g., auditory demands, such as talking on the phone) and the physical intensity of the stimulus (i.e., vibration amplitude, sound pressure) can explain these differential findings (see Dodonova, & Dodonov, 2013 and Jensen, 2006 for overviews of task-related factors that influence reaction times). In summary, the results of the current study show HMI designers that vibrotactile and auditory feedback are both effective in alerting the driver of a take-over request.

Bimodal TORs yielded mean steering reaction times that were about 0.2 s faster than unimodal TORs (Table 2). It should be noted that the differences were statistically significant only for the steer-touch and steer-initiate times, and not for the steer-turn time, car avoid time, lane change time, or brake time. The lack of significant effects may be explained because participants were inclined to grab the steering wheel as quickly as possible, while effects diluted afterwards (i.e., there may be no need to change lanes as quickly as possible, given the fairly large time budget of 7 s). This dilutive effect can be seen when comparing the magnitude of the standard deviations between the steer-touch and steer-initiate time versus the other measures in Table 2. Although the difference in steer-touch times was only 0.2 s, it could have large safety consequences if this effect transpires to evasive manoeuvring or braking in a truly urgent condition. For example, if decelerating with 8 m/s2, a 0.2 s reaction time advantage implies a speed reduction of 6 km/h, which in turn has strong effects on the probability of surviving a collision (Joksch, 1993). More research is needed to investigate whether multimodal feedback offers safety benefits on roads. Specifically, driving safety is not only determined by take-over time, but also by take-over quality (cf. Radlmayr et al., 2014).

One of the goals of this study was to determine whether directional feedback without any prior instructions about the directionality caused drivers to follow the direction of the feedback. The results showed that almost all participants overtook the stationary car on the left, regardless of the directionality of the TOR. This result is consistent with German traffic rules stating that overtaking on the right is prohibited on highways, and suggests that rules and habits are dominant performance-shaping factors regarding whether a driver steers to the left or right. The post-experiment questionnaire showed that only a few participants reported to have noticed that feedback was directional (Table S6). These findings may be attributable to a lack of saliency of the directionality of the stimuli, or by the cognitively engaging secondary task (SuRT). Future studies should investigate whether more salient cues, instructions, or a higher level of semantics might allow the driver to perceive directional cues in a TOR. In more recent work (Petermeijer, Cieler, & De Winter, 2017) we found correct left/right response rates in the order of 80% to 90% after participants had been trained and instructed about the meaning of directional vibrotactile stimuli. In order to improve drivers' responsiveness to directional feedback, future research could investigate the effectiveness of verbal warnings (e.g., "left!", "right!", see Gold, Berisha, & Bengler, 2015). Such directional voice cueing is also used in traffic alert and collision avoidance systems (TCAS), a technology that is mandatory in most airplanes. Another promising solution for object avoidance is to use continuous force feedback on the steering wheel, an approach that may work both when the driver touches and when the driver

releases the steering wheel. This approach, also known as haptic shared control or haptic steering guidance, has been previously shown to support effective left/right steering decisions in a head-on collision scenario (Della Penna, Van Paassen, Abbink, Mulder, & Mulder, 2010). Yet another strategy is to apply small oscillatory movements on the steering wheel to prime the driver to steer in a particular direction (Navarro, Mars, and Hoc, 2007; Navarro et al., 2010).

Our results further suggest that the initial steering reaction represents only a portion of the behaviours that occur in a conflict resolution scenario. For example, we showed that drivers can increase their own temporal demands by braking (Fig. 5), and that they become more efficient at resolving the conflicts with increasing experience (Fig. 6, see also Young, 2000). This latter finding is in line with research showing that practice and mental model forming are crucial determinants of the use of automated driving systems (Beggiato, Pereira, Petzoldt, & Krems, 2015).

Several limitations have to be considered when interpreting the results of this experiment. First, participants experienced a high number of TORs per time unit (i.e., every 1.5 to 2 min), and each take-over scenario was identical (i.e., no other road users, time budget of 7 s, and a stationary vehicle in the middle lane). The high consistency and low level of ambiguity may explain why none of the participants missed a TOR or crashed into the stationary car. Future studies on driving behaviour in a take-over scenario should consider that unsuspecting drivers are unlikely to react quickly and consistently. For example, an on-road study into drivers' reactions to truly unexpected events found that drivers reacted in 2.5 s on average (Summala, 1981), which is considerably higher than professional drivers' average reaction time of 0.84 s to system failures of automated vehicles on the roads (Dixit, Chand, & Nair, 2016). Our approach has advantages from a statistical viewpoint because we obtained as much as 125 reaction times for each of the three TOR conditions. According to our literature survey, over 70 studies have been published on take-over performance in highly automated driving, yet almost all of them included only one TOR per experimental condition. An exception is Young (2000), who found a learning effect in a critical event scenario that required a braking intervention, with 16 of 44 participants applying the brakes in trial 1, and 36 of 44 participants in trial 2. Similarly, Hergeth, Lorenz, and Krems (in press) found that the take-over time reduced from the first to the second TOR. In our study, we established learning curves across as much as 18 TORs.

Second, although simulators are useful tools because they offer safety and a high degree of controllability of the environment, by definition simulators have limited fidelity (Boer, Della Penna, Utz, Pedersen, & Sierhuis, 2015; De Winter, Van Leeuwen, & Happee, 2012). Our simulator had a realistic visual projection with a large field of view, but did not provide vestibular motion feedback. It is known that participants brake harder and more abruptly in simulators than in real cars, especially when the simulator has no motion platform (Boer, Yamamura, Kuge, & Girshick, 2000; De Groot, De Winter, Mulder & Wieringa, 2011; Siegler, Reymond, Kemeny & Berthoz, 2001). Klüver, Herrigel, Heinrich, Schöner, and Hecht (2016) found that drivers showed higher standard deviation of lateral position (SDLP) in fixed base simulators than in moving simulators and in a real car (i.e., a violation of absolute fidelity because of the discrepancy in SDLPs). However, these authors showed that the fixed base simulators were still useful for assessing the distractive effect of secondary tasks

(i.e., a confirmation of relative fidelity). Another factor is that our experiment occurred in the summer period with high temperatures in the lab. Because the air conditioning in the car was not functional and the car windows were closed, it is possible that the observed reaction times may have been slower than reaction times in real cars that are equipped with air conditioning (but see Teichner, 1954 claiming that ambient temperature has little to no effect on reaction times). On the other hand, it is also possible that the haptic seat or auditory warnings are actually more difficult to detect in real cars, due to environmental noise and vibrations that may be more intense on the road than in the simulator.

Vibrotactile TORs were rated as more satisfactory than auditory TORs, which is consistent with results of Stanley (2006) and Calhoun et al. (2005) on haptic and auditory warnings. Moreover, multimodal TORs were rated as more useful than auditory ones. The questionnaire data showed that participants became more appreciative towards TORs they were exposed to (Fig. 7). These results per se do not imply that multimodal TORs are preferred over visual TORs; participants may have rated the vibrotactile and multimodal TORs highly for the reason that they had experienced them in the experiment.

Finally, the participants were mostly researchers and students from the Technical University of Munich, many of whom had previously participated in driving simulator studies and were familiar with the principles of highly automated driving. Further research could investigate the effects of TORs in different samples of the driving population. Note that it is likely that drivers of future highly automated cars will also be familiar with the technology in their cars, and so testing naïve participants may not be a recommended approach either.

In summary, our results showed that multimodal TORs yielded a faster steer-touch times and higher self-reported usefulness than unimodal TORs, and the directional cue evoked no spontaneous contra- or ipsilateral response of the drivers. Our results complement the literature on multimodal warnings in general (Bazilinsksy et al., 2015; Burke et al., 2006; Diederich & Colonius, 2004; Oviatt, 1999; Petermeijer et al., 2015; Van Erp, Toet, & Janssen, 2015), and suggest that in a take-over scenario, a TOR should be multimodal rather than unimodal.

3.3.6 Supplementary material

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.apergo.2017.02.023

3.3.7 References

Bazilinskyy, P., & De Winter, J. (2015). Auditory interfaces in automated driving: an international survey. *PeerJ Computer Science*, *1*, e13. https://doi.org/10.7717/peerj-cs.13

Bazilinskyy, P., Kyriakidis, M., & De Winter, J. C. F. (2015). An international crowdsourcing study into people's statements on fully automated driving. Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Las Vegas, NV, 2534–2542.

Bazilinskyy, P., Petermeijer, S. M., Petrovych, V., Dodou, D., & De Winter, J. C. F. (2018). Take-over requests in highly automated driving: A crowdsourcing

survey on auditory, vibrotactile, and visual displays. *Transportation Research Part F: Traffic Psychology and Behaviour*, *56*, 82–98.

Beggiato, M., Pereira, M., Petzoldt, T., & Krems, J. (2015). Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *35*, 75–84.

Beruscha, F., Wang, L., Augsburg, K., & Wandke, H. (2010). Do drivers steer toward or away from lateral directional vibrations at the steering wheel. Proceedings of *European Conference on Human Centred Design for Intelligent Transport Systems, 2*, 227–236.

BMW. (2015). The BMW 5 series Gran Turismo. Owner's manual. Available at https://carmanuals2.com/get/bmw-5-series-gran-turismo-2015-owner-s-manual-63422

Boer, E. R., Della Penna, M., Utz, H., Pedersen, L., Sierhuis, M. (2015). The role of driving simulators in developing and evaluating autonomous vehicles. Proceedings of *the Driving Simulation Conference Europe* (pp. 3–10). Paris, France.

Boer, E. R., Yamamura, T., Kuge, N., & Girshick, A. (2000). Experiencing the same road twice: A driver centered comparison between simulation and reality. Proceedings of *Driving Simulation Conference* (pp. 33–55). Paris, France.

Burke, J. L., Prewett, M. S., Gray, A. A., Yang, L., Stilson, F. R. B., Coovert, M. D., ... & Redden, E. (2006). Comparing the effects of visual-auditory and visual-tactile feedback on user performance: A meta-analysis. Proceedings of *the 8th International Conference on Multimodal Interfaces*, 108–117, New York, NY: ACM.

Calhoun, G. L., Draper, M. H., Guilfoos, B. J., & Ruff, H. A. (2005). Tactile and aural alerts in high auditory load UAV control environments. Proceedings of *the Human Factors and Ergonomics Society Annual Meeting, 49*, 145–149.

Cao, Y., Van der Sluis, F., Theune, M., Op den Akker, R., & Nijholt, A. (2010). Evaluating informative auditory and tactile cues for in-vehicle information systems. Proceedings of *the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 102–109, ACM.

Citroën. (2007). Citroën C5. Owner's manual. Available at https://carmanuals2.com/get/citroen-c5-2007.5-owner-s-manual-63737

Clark, H., & Feng, J. (2015). Semi-autonomous vehicles: Examining driver performance during the take-over. Proceedings of *the Human Factors and Ergonomics Society 59th Annual Meeting, 59*, 781–785.

Damböck, D. (2013). *Automationseffekte im Fahrzeug–von der Reaktion zur Übernahme* (Doctoral dissertation). Technical University of Munich.

De Groot, S., De Winter, J. C. F., Mulder, M., & Wieringa, P. A. (2011). Nonvestibular motion cueing in a fixed-base driving simulator: Effects on driver braking and cornering performance. *Presence: Teleoperators and Virtual Environments, 20*, 117–142.

De Winter, J. C. F. (2013). Predicting self-reported violations among novice license drivers using pre-license simulator measures. *Accident Analysis and Prevention*, *52*, 71–79.

De Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation*

Research Part F: Traffic Psychology and Behaviour, *27*, 196–217.
De Winter, J. C. F., Van Leeuwen, P. M., & Happee, R. (2012). Advantages and disadvantages of driving simulators: A discussion. Proceedings of *the Measuring Behavior Conference*, Utrecht, the Netherlands, 47–50.
Della Penna, M., Van Paassen, M. M., Abbink, D. A., Mulder, M., & Mulder, M. (2010). Reducing steering wheel stiffness is beneficial in supporting evasive maneuvers. *IEEE International Conference on Systems Man and Cybernetics*, 1628–1635.
Diaconescu, A. O., Alain, C., & McIntosh, A. R. (2011). The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *Journal of Neurophysiology*, *106*, 2896–2909.
Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*, 1388–1404.
Dixit, V. V., Chand, S., & Nair, D. J. (2016). Autonomous vehicles: disengagements, accidents and reaction times. *PLOS ONE*, *11*, e0168054.
Dodonova, Y. A., & Dodonov, Y. S. (2013). Is there any evidence of historical slowing of reaction time? No, unless we compare apples and oranges. *Intelligence*, *41*, 674–687.
General Motors. (2014). Guide du propriétaire Chevrolet Tahoe/Suburban 2015 [Owner's manual Chevrolet Tahoe/Suburban 2015]. Available at https://carmanuals2.com/get/chevrolet-tahoe-2015-manuel-du-proprietaire-56514
Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? Proceedings of *the Human Factors and Ergonomics Society Annual Meeting, 57*, 1938–1942.
Gold, C., Berisha, I., & Bengler, K. (2015). Utilization of drivetime – Performing non-driving related tasks while driving highly automated. Proceedings of *the Human Factors and Ergonomics Society 59th Annual Meeting, 59*, 1666–1670.
Gray, R., Ho, C., & Spence, C. (2014). A comparison of different informative vibrotactile forward collision warnings: does the warning need to be linked to the collision event. *PLOS ONE*, *9*, e87070.
Hart, S. G., & Staveland. L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (139–183). Amsterdam, the Netherlands: North Holland Press.
Hergeth, S., Lorenz, L., & Krems, J. F. (in press). Prior familiarization with takeover requests affects drivers' takeover performance and automation trust. *Human Factors*.
Ho, C., Tan, H. Z., & Spence, C. (2006). The differential effect of vibrotactile and auditory cues on visual spatial attention. *Ergonomics*, *49*, 724–738.
Houtenbos, M., De Winter, J. C. F., Hale, A. R., Wieringa, P. A., & Hagenzieker, M. P. (2017). Concurrent audio-visual feedback for supporting drivers at intersections: A study using two linked driving simulators. *Applied Ergonomics, 60*, 30–42.
ISO/DTS 14198. (2012). Road vehicles - Ergonomic aspects of transport information and control systems - Calibration tasks for methods which assess driver demand due to the use of in vehicle systems. Available at http://www.din.de/en/getting-involved/standards-

committees/naautomobil/din-spec/wdc-beuth:din21:169391345/toc-1933931/download

Jensen, A. R. (2006). Clocking the mind: Mental chronometry and individual differences. Amsterdam: Elsevier.

Joksch, H. C. (1993). Velocity change and fatality risk in a crash—a rule of thumb. *Accident Analysis and Prevention*, *25*, 103–104.

Klüver, M., Herrigel, C., Heinrich, C., Schöner, H. P., & Hecht, H. (2016). The behavioral validity of dual-task driving performance in fixed and moving base driving simulators. *Transportation Research Part F: Traffic Psychology and Behaviour*, *37*, 78–96.

Kyriakidis, M., Happee, R., & De Winter, J. C. F. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour, 32*, 127–140.

Lee, J., McGehee, D., Brown, T., & Marshall, D. (2006). Effects of adaptive cruise control and alert modality on driver performance. *Transportation Research Record: Journal of the Transportation Research Board*, *1980*, 49–56.

Lees, M. N., Cosman, J., Lee, J. D., Vecera, S. P., Dawson, J. D., & Rizzo, M. (2012). Cross-modal warnings for orienting attention in older drivers with and without attention impairments. *Applied Ergonomics, 43*, 768–776.

Lorenz, L., Kerschbaum, P., & Schumann, J. (2014). Designing take over scenarios for automated driving How does augmented reality support the driver to get back into the loop? Proceedings of *the Human Factors and Ergonomics Society Annual Meeting, 58*, 1681–1685.

Louw, T. L., Merat, N., & Jamson, A. H. (2015). Engaging with highly automated driving: To be or not to be in the loop? *Proceedings of the 8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 190–196). Salt Lake City, UT.

Lu, Z., Happee, R., Cabrall, C. D. D., Kyriakidis, M., & De Winter, J. C. F. (2016). Human factors of transitions in automated driving: A general framework and literature survey. *Transportation Research Part F: Traffic Psychology and Behaviour, 43*, 183–198.

Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., & Bauer, W. (2015). Take-over requests for automated driving. Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Las Vegas, NV, 2867–2873.

Meng, F., & Spence, C. (2015). Tactile warning signals for in-vehicle systems. *Accident Analysis & Prevention, 75*, 333–346.

Merat, N., Jamson, A. H., Lai, F. C., Daly, M., & Carsten, O. M. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 274–282.

Mercedes-Benz (2015). E-Class. Sedan. Operator's manual. Available at https://carmanuals2.com/get/bmw-5-series-gran-turismo-2015-owner-s-manual-63422

Mohebbi, R., Gray, R., & Tan, H. Z. (2009). Driver reaction time to tactile and auditory rear-end collision warnings while talking on a cell phone. *Human Factors*, *51*, 102–110.

Mok, B., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., & Ju, W. (2015).

Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles. Proceedings of *the 2015 IEEE 18th International Conference on Intelligent Transportation Systems,* Canary Islands, Spain, 2458–2464.

Müsseler, J., Aschersleben, G., Arning, K., & Proctor, R. W. (2009). Reversed effects of spatial compatibility in natural scenes. *American Journal of Psychology, 122*, 325–336.

Naujoks, F., Mai, C., & Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. In T. Ahram, W. Karwowski, & T. Marek (Eds.), Proceedings of *the 5th International Conference on Applied Human Factors and Ergonomics (AHFE)*, *Kraków, Poland*.

Naujoks, F., & Neukum, A. (2014). Specificity and timing of advisory warnings based on cooperative perception. In A. Butz, M. Koch, & J. Schlichter (Eds.), Mensch & Computer 2014 – Workshopband (pp. 229–238). Berlin: De Gruyter Oldenbourg.

Naujoks, F., Purucker, C., Neukum, A., Wolter, S., & Steiger, R. (2015). Controllability of partially automated driving functions – Does it matter whether drivers are allowed to take their hands off the steering wheel? *Transportation Research Part F: Traffic Psychology and Behaviour*, *35*, 185–198.

Navarro, J., Mars, F., Forzy, J. F., El-Jaafari, M., & Hoc, J. M. (2010). Objective and subjective evaluation of motor priming and warning systems applied to lateral control assistance. *Accident Analysis & Prevention, 42*, 904–912.

Navarro, J., Mars, F., Hoc, J.-M., (2007). Lateral control support for car drivers: a human-machine cooperation approach. Proceedings of *the 14th European Conference on Cognitive Ergonomics: Invent! Explore!*, 249–252, New York, NY: ACM.

Nukarinen, T., Rantala, J., Farooq, A., & Raisamo, R. (2015). Delivering directional haptic cues through eyeglasses and a seat. *IEEE World Haptics Conference* (pp. 345–350). Evanston, IL.

Oviatt, S. (1999). Ten myths of multimodal interaction. *Communications of the ACM*, *42*, 74–81.

Payre, W., Cestac, J., & Delhomme, P. (2016). Fully automated driving impact of trust and practice on manual control recovery. *Human Factors*, *58*, 229–241.

Petermann-Stock, I., Hackenberg, L., Muhr, T., & Mergl, C. (2013). Wie lange braucht der Fahrer? Eine Analyse zu Übernahmezeiten aus verschiedenen Nebentätigkeiten während einer hochautomatisierten Staufahrt. *Tagung Fahrerassistenzsysteme. Der Weg zum automatischen Fahren*. TÜV SÜD Akademie GmbH.

Petermeijer, S. M., Abbink, D. A., Mulder, M., & De Winter, J. C. F. (2015). The effect of haptic support systems on driver performance: A literature survey. *IEEE Transactions on Haptics*, *PP*.

Petermeijer, S. M., Cieler, S., & De Winter, J. C. F. (2017). Comparing spatially static and dynamic vibrotactile take-over requests in the driver seat. *Accident Analysis and Prevention, 99*, 218–227.

Petermeijer, S. M., De Winter, J. C. F., & Bengler, K. (2016). Vibrotactile displays: A survey with a view on highly automated driving. *IEEE Transactions on Intelligent Transportation Systems, 17(4),* 897–907.

Politis, I., Brewster, S. A., & Pollick, F. (2014). Evaluating multimodal driver displays under varying situational urgency. Proceedings of *the SIGCHI Conference on Human Factors in Computing Systems*, 4067–4076, New York, NY: ACM.

Prewett, M. S., Elliott, L. R., Walvoord, A. G., & Coovert, M. D. (2012). A meta-analysis of vibrotactile and visual information displays for improving task performance. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *42*, 123–132.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. Proceedings of *the Human Factors and Ergonomics Society Annual Meeting, 58*, 2063–2067.

Scott, J. J., & Gray, R. (2008). A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. *Human Factors*, *50*, 264–275.

SAE International (2014). Taxonomy and definitions for terms related to onroad motor vehicle automated driving systems (Standard No. J3016). SAE International. Available at http://standards.sae.org/j3016_201401.

Schwalk, M., Kalogerakis, N., & Maier, T. (2015). Driver support by a vibrotactile seat matrix – Recognition, adequacy and workload of tactile patterns in take-over scenarios during automated driving. Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Las Vegas, NV, 1427–1434.

Sharek, D. (2011). A useable, online NASA-TLX tool. Proceedings of *the Human Factors and Ergonomics Society Annual Meeting*, 55, 1375–1379.

Shladover, S. E. (2015, November). Road vehicle automation history, opportunities and challenges. *Mini-seminar 'developments selfdriving vehicles in USA'.* Delft, the Netherlands.

Siegler, I., Reymond, G., Kemeny, A., & Berthoz, A. (2001). Sensorimotor integration in a driving simulator: Contributions of motion cueing in elementary driving tasks. Proceedings of *the Driving Simulation Conference*, Sophia-Antipolis, France, 21–32.

Simon, J. R., Hinrichs, J. V., & Craft, J. L. (1970). Auditory SR compatibility: reaction time as a function of ear-hand correspondence and ear-response-location correspondence. *Journal of Experimental Psychology, 86*, 97–102.

Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. *Hearing Research, 258*, 134–142.

Stanley, L. M. (2006). Haptic and auditory cues for lane departure warnings. Proceedings of *the Human Factors and Ergonomics Society Annual Meeting*, *50*, 2405–2408.

Stanton, N. A., & Edworthy, J. (1999). Auditory warnings and displays: An overview. In N. A. Stanton & J. Edworthy (Eds.), *Human factors in auditory warnings* (pp. 3–30). Aldershot: Ashgate Publishing Limited.

Stokes, A., Wickens, C., & Kite, K. (1990). *Display technology-human factors concepts* (Technical Report No. SAE R-102). Warrendale, PA: Society of Automotive Engineers, Inc.

Straughn, S. M., Gray, R., & Tan, H. Z. (2009). To go or not to go: Stimulus-response compatibility for tactile and auditory pedestrian collision warnings. *IEEE Transactions on Haptics*, *2*, 111–117.

Summala, H. (1981). Drivers' steering reaction to a light stimulus on a dark

road. *Ergonomics, 24*, 125–131.
Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*, 400–410.
Technical University of Munich (2015). Static driving simulator. Available at http://www.lfe.mw.tum.de/en/research/methods-and-lab-equipment/static-driving-simulator
Teichner, W. H. (1954). Recent studies of simple reaction time. *Psychological Bulletin*, *51*, 128–149.
Telpaz, A., Rhindress, B., Zelman, I., & Tsimhoni, O. (2015). Haptic seat for automated driving: preparing the driver to take control effectively. Proceedings of *the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 23–30. New York, NY: ACM.
Tijerina, L, Jackson, J. L., Pomerleau, D. A., Romano, R. A., & Petersen, A. D. (1996). Driving simulator tests of lane departure collision avoidance systems. *Intelligent Transportation: Realizing the Benefits*. Proceedings of *the 1996 Annual Meeting of ITS America,* 636–648.
Umiltá, C., & Nicoletti, R. (1990). Spatial stimulus-response compatibility. *Advances in Psychology*, *65*, 89–116.
Underwood, S. E. (2014). Automated vehicles forecast vehicle symposium opinion survey. Presentation at *the 2014 Automated Vehicles Symposium*, San Francisco, CA.
Van der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*, 1–10.
Van Erp, J. B. F., Toet, A., & Janssen, J. B. (2015). Uni-, bi-and tri-modal warning signals: Effects of temporal parameters and sensory modality on perceived urgency. *Safety Science*, *72*, 1–8.
Walch, M., Lange, K., Baumann, M., & Weber, M. (2015). Autonomous driving: investigating the feasibility of car-driver handover assistance. Proceedings of *the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 11–18. New York, NY: ACM.
Wang, D. Y. D., Pick, D. F., Proctor, R. W., & Ye, Y. (2007). Effect of a side collision-avoidance signal on simulated driving with a navigation system. Proceedings of *the 4th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 206–211, Stevenson, WA.
Weller, G., Heyne, F., Feige, T., Bretschneider, H., Oeser, H., & Schlag, B. (2013). Die Wirkung gerichteter Warnungen von Fahrerassis-tenzsystemen auf die Blickzuwendungs- und Reakti-onszeiten von Autofahrern. In E. Brandenburg, L. Doria, A. Gross, T. Günzler & H. Smieszek (Eds.), Grundlagen und Anwendungen der Mensch-Maschine-Interaktion 10. Berliner Werkstatt Mensch-Maschine-Systeme (pp. 376–382). Berlin: Universitätsverlag der TU Berlin.
Woodworth, R. S., & Schlosberg, H. (1954). *Experimental Psychology* (Revised edition). New York, NY: Henr y Holt & Co.
Young, M. S. (2000). Attention, automaticity, and automation: New perspectives on mental underload and performance (Doctoral dissertation). University of Southampton, UK.

Zarife, R. (2014). *Integrative warning concept for multiple driver assistance systems* (Doctoral dissertation). University of Würzburg.
Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident Analysis and Prevention*, *78*, 212–221.
Zhang, Y., Yan, X., & Yang, Z. (2015). Discrimination of effects between directional and nondirectional information of auditory warning on driving behavior. *Discrete Dynamics in Nature and Society*.

3.4 Usefulness and satisfaction of take-over requests for highly automated driving

Corresponding article: Bazilinskyy, P., Eriksson, A., Petermeijer, S. M., & De Winter, J. C. F. (2017). Usefulness and satisfaction of take-over requests for highly automated driving. Proceedings of Road Safety and Simulation conference (RSS) 2017, The Hague, the Netherlands.

The present section contains text from a review article. The author was involved in the mentioned below studies: Bazilinskyy & De Winter, 2017; Petermeijer et al., 2017a (Experiment 1); Bazilinskyy et al., 2018. Other studies mentioned in the section were conducted within the Work Package 2 of the HFAuto project.

3.4.1 Abstract

This paper summarizes our results from survey research and driving simulator experiments on auditory, vibrotactile, and visual take-over requests in highly automated driving. Results showed that vibrotactile take-over requests in the driver's seat yielded relatively high ratings of self-reported usefulness and satisfaction. Auditory take-over requests in the form of beeps were regarded as useful but not satisfactory, and it was found that an increase of beep rate yields an increase of self-reported urgency. Visual-only feedback in the form of LEDs was regarded by participants as neither useful nor satisfactory. Augmented visual feedback was found to support effective steering and braking actions, and may be a useful complement to vibrotactile take-over requests. The present findings may be useful in the design of take-over requests.

3.4.2 Introduction

Automated driving is being pursued at a large scale by various vehicle manufacturers. However, fully automated driving, in which the driver never has to intervene, does not exist yet on public roads. Between September 2014 to December 2016, 2,616 disengagements of control were recorded in on-road automated test vehicles. These were due to a human factor (e.g., driver discomfort with the vehicle's behaviour) in 30% of the cases, 52% were due to system failures (e.g., software discrepancy), and 11% were due to external conditions (e.g., poorly marked lanes) (Favarò et al., 2018).

When a highly automated car reaches its operational limits, the driver has a certain time buffer to reclaim control. This time buffer (also called lead time, transition time, or time budget) may range from long (for non-urgent situations) to short (for high-urgent situations, such as an impending collision) (Bazilinskyy

et al., 2018). Prior research has measured drivers' behaviour in take-over situations involving time buffers ranging between 2 and 10 seconds (e.g. Gold et al., 2013, for a review see Eriksson & Stanton, 2017).

The design of take-over requests (TORs) is a crucial factor in the safety of automated driving systems, because a late or wrong response may lead to incidents and accidents. If the time buffer is short, the driver could benefit from receiving a take-over request (a warning) that conveys a high level of urgency. On the other hand, if the automated vehicle can anticipate when a transition to manual control will be needed, a take-over request can be issued well in advance in a more discretionary manner.

New types of in-vehicle feedback, such as take-over requests, can be rated along two dimensions: (1) usefulness (quality) and (2) satisfaction (pleasure) (Van der Laan et al., 2016). Both dimensions are regarded as important. That is, the feedback should be useful in that it supports drivers in making a safe and timely response, and it should be satisfactory: if it is not, then drivers may become annoyed and disable the feedback system altogether (Parasuraman & Riley, 1997).

Within the project HFauto (Human Factors of Automated Driving), we have investigated how drivers perceive and respond to different auditory, vibrotactile, and visual take-over requests for highly automated driving. The aim of the present paper is to summarize our contributions regarding the effects of take-over request modality on drivers' self-reported usefulness and satisfaction, as well as their response.

3.4.3 Auditory take-over requests

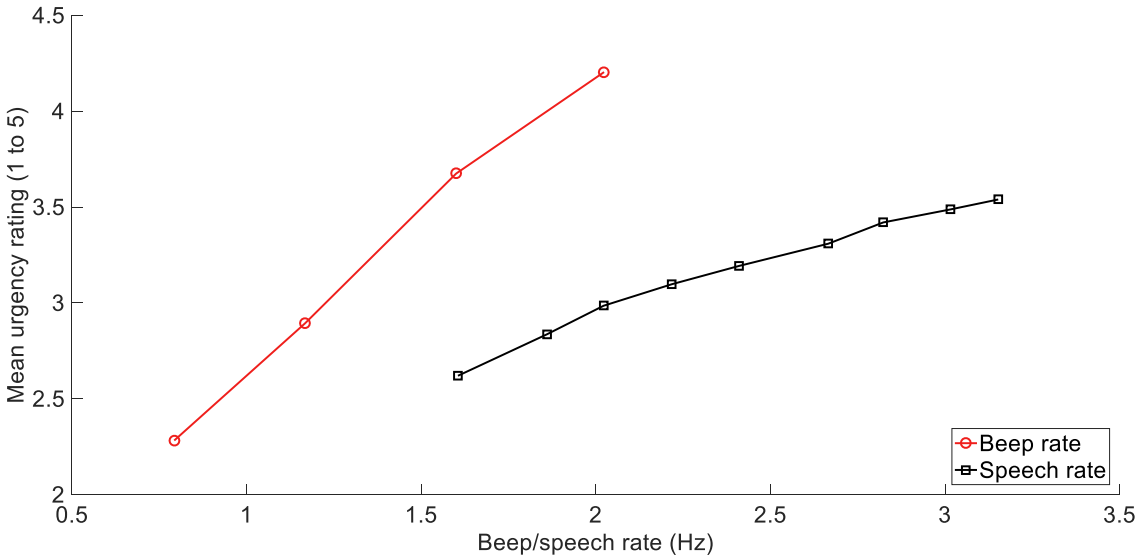


Figure 1. Relationship between self-reported urgency and beep rate / speech rate (Bazilinskyy et al., 2018; Bazilinskyy & De Winter, 2017). The beep rate is the number of beeps per second (beeps were presented in pairs, with a 0.11 s pause in between); speech rate is expressed in syllabi per second for the uttered phrase "*take over please*" (i.e., 4 syllabi). Participants responded to the statement 'I consider such a sound as urgent' or 'The message is urgent' on a scale from 1 = Disagree strongly, 5 = Agree strongly.

Beeps are a commonly used type of auditory feedback in automated driving (Cabrall et al., 2017). In a crowdsourcing study with 1,692 participants, we

replicated previous experimental results by showing that there exists a clear monotonic relationship between self-reported urgency and inter-beep duration (Bazilinskyy et al., 2018), see Figure 1.

Auditory take-over requests can also be provided in the form of speech (Gold et al., 2015; Politis et al., 2015). In the same large-scale international online survey, we found that the female and male voice ("Please take over!") were rated as more preferred than beeps (Bazilinskyy et al., 2018). In another large-sample crowdsourcing study (Bazilinskyy & De Winter, 2017), 2,669 participants from 95 countries listened to a random 10 out of 140 take-over requests, and rated each take-over request on urgency, commandingness, pleasantness, and ease of understanding. We found that differences in speech intelligibility and self-reported urgency between take-over requests in male versus female voice are generally small. Additionally, in agreement with earlier findings by Hellier et al., 2002, we found that the spoken phrase (e.g., "Note, take over" versus "Take over now") affects self-reported urgency. Furthermore, it was shown that an increase of speech rate yielded increased self-reported urgency (Fig. 1).

Figure 2. A take-over scenario. Participants could avoid the stationary car by steering left or right (Petermeijer et al., 2017a; Experiment 1).

In an experiment in a driving simulator (Experiment 1), we found that drivers responded effectively (i.e., average steer initiation times of about 2.0 s, which was well within the 7 s time buffer) to an auditory take-over request (double beep) (Petermeijer et al., 2017a). However, *directional* auditory feedback provided via the car's speakers on either the left or the right was not noticed by drivers. Drivers who received directional feedback almost always steered to the left in a scenario where a stationary car blocked the middle lane (Fig. 2), just as

did drivers who received non-directional feedback (i.e., TOR provided via the left and right speakers simultaneously).

3.4.4 Vibrotactile take-over requests

Vibrotactile messages can be perceived even in the presence of auditory distractions such as a phone call or radio music (Petermeijer et al., 2016). In a driving simulator experiment involving take-over situations with 7 s time buffer, we found that vibrotactile-only warnings in the driver seat are effective for ensuring that drivers reclaim control of the steering wheel in time (Experiment 1) (Petermeijer et al., 2017a). However, *directional* vibrotactile cues embedded in the take-over request did not elicit a directional response in uninstructed drivers (Petermeijer et al., 2017a). In a follow-up study (Experiment 2) (Petermeijer et al., 2017b), it was evaluated how well drivers recognized directional cues presented via the vibrotactile seat, when they were explicitly instructed about the meaning of the directional cues. Here, the participants received a static (i.e., left or right) or dynamic (i.e., moving left or right) take-over request and were asked to change lane according to the directional cue. Results showed that participants did not accurately detect the directional vibrotactile cues (correct response rates of about 80%). Furthermore, it was found that static take-over requests yielded faster reaction times than dynamic ones. In summary, vibrotactile take-over requests are useful for alerting a driver, but the amount of information that can be communicated via a vibrotactile seat may be limited (Petermeijer et al., 2016).

3.4.5 Visual augmented feedback and visual take-over requests

In a driving simulator study (Experiment 3, Eriksson et al., 2017) we assessed the effectiveness of visual augmented feedback for supporting vibrotactile take-over requests. Four the types of visual feedback were evaluated during lane change and braking take-over scenarios: (1) a baseline condition without visual support, (2) a sphere highlighting a slow-moving vehicle ahead (Fig. 3, top middle), (3) a green carpet in the left lane for the lane change scenario (Fig. 3, bottom left) and a red barrier covering the lane markings for the braking scenario, and (4) a green arrow pointing left for the lane change scenario (Fig. 3, bottom right) and red arrow pointing backwards, for the braking scenario. We found that the carpet feedback and arrow feedback facilitated accurate braking and lane changing behaviour compared to the baseline condition, whereas the sphere feedback appeared to cause confusion in that drivers showed unnecessary braking in a scenario in which they only had to change lanes.

In another driving simulator study (Experiment 4, Petermeijer et al., 2017c), we measured driver response times to take-over requests provided via (1) a vibrotactile seat, (2) auditory beeps, and (3) visual LEDs surrounding the secondary task display and above the steering wheel, while drivers were performing different types of secondary tasks (watching a video on the display, reading on the display, or performing a simulated hands-free phone task). The results of this study showed that the initial steering response times were about 0.6 s slower for the visual take-over requests than for the vibrotactile and auditory take-over requests. It was concluded that visual warnings convey a low sense of urgency or may go unnoticed even when in the driver's visual field of view. In summary, visual messages are prone to be overlooked, especially during

highly automated driving in which drivers will be allowed to take their eyes off the road and engage in non-driving tasks.

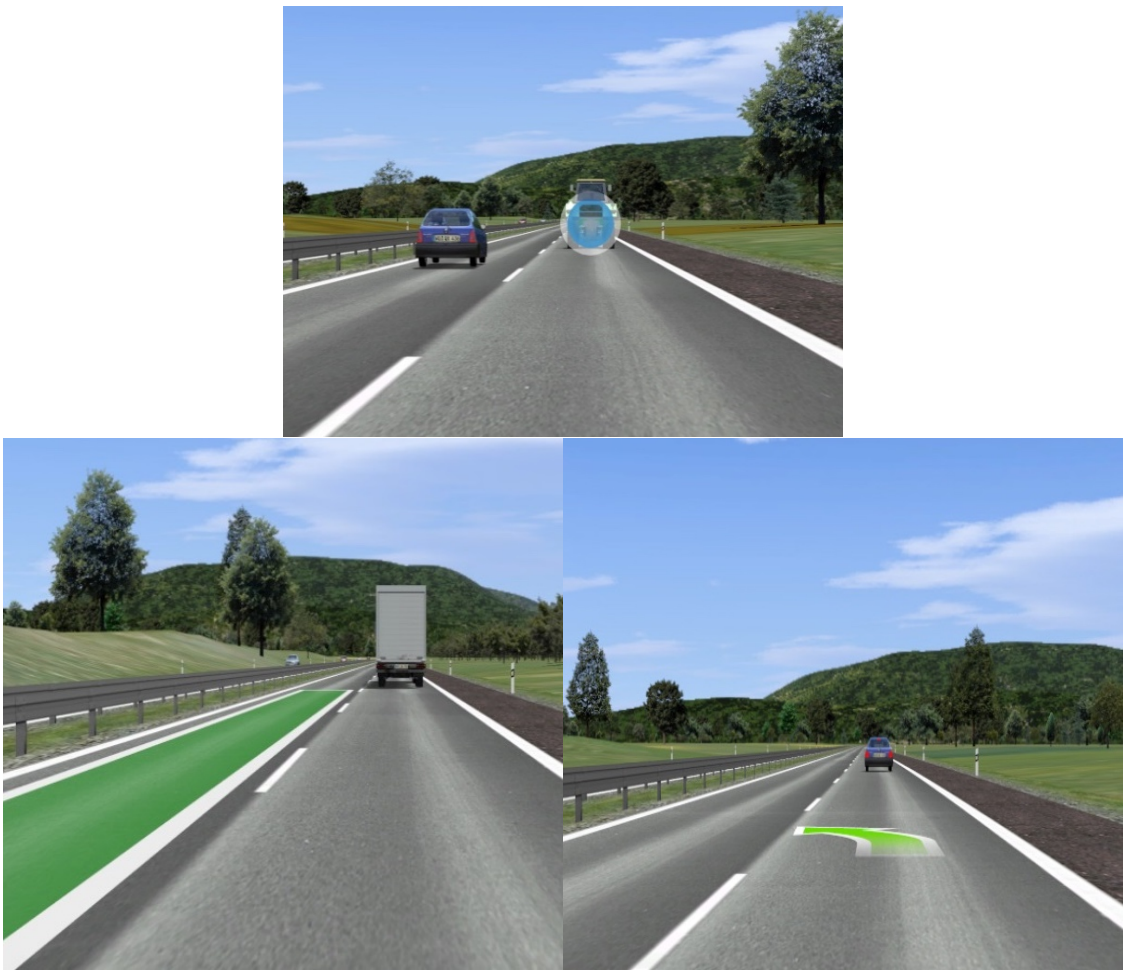


Figure 3. Examples of the visual interfaces used in Eriksson et al. (2017; Experiment 3). Top middle (sphere): a sphere highlighting the slow-moving vehicle ahead in both scenarios. Bottom left (carpet): a green carpet in the left lane for the lane change scenario. Bottom right (arrow): a green arrow pointing left for the lane change scenario.

3.4.6 Comparing auditory, visual, and vibrotactile take-over requests

Figure 4 summarizes the results of the same usefulness and satisfaction questionnaire (Van der Laan et al., 1997), for all four driving simulator experiments reviewed above. All experiments were performed with the same driving simulator software (SILAB) and with equivalent simulator hardware (i.e., full passenger vehicle with surround projection).

Several findings stand out: visual-only take-over requests (i.e., the LEDs) was not regarded as useful by participants (Experiment 4, Petermeijer et al., 2017c). These subjective findings mirror the objective take-over response times for visual-only take-over requests, which were found to be slower than the response times to vibrations-only and auditory-only take-over requests (Petermeijer et al., 2017c). Furthermore, auditory-only feedback (Experiments 1 & 4; Petermeijer et al., 2017a and Petermeijer et al., 2017c) was useful, but not satisfactory. In our experiments, the auditory take-over request consisted of loud beeps. Vibrations were overall regarded as both useful and satisfactory (Experiments 1–4). However, vibrations combined with ambiguous visual

information (Sphere) reduced overall usefulness and satisfaction as compared to vibrations-only take-over requests (Experiment 3, Eriksson et al., 2017).

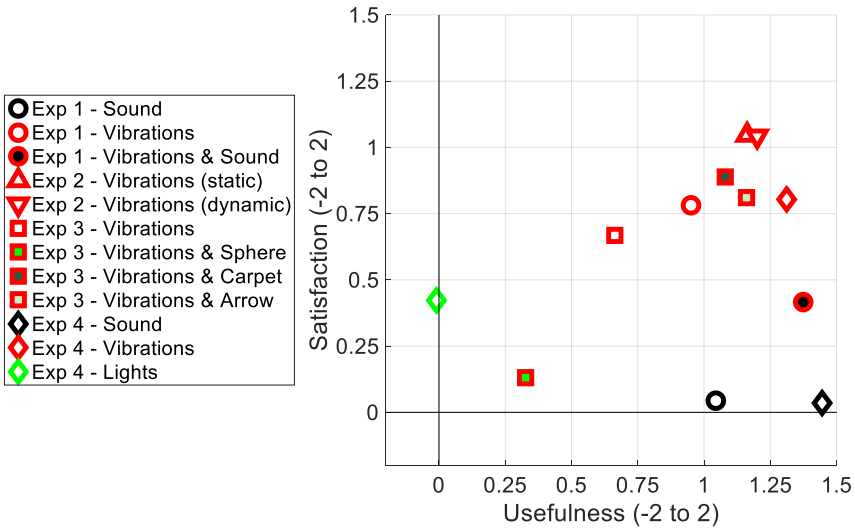


Figure 4. Self-reported satisfaction and usefulness for take-over requests tested in four driving simulator experiments.

3.4.7 Discussion

We designed and evaluated various auditory, visual, and vibrotactile take-over requests. Results showed that visual-only take-over requests in the form of LEDs yielded low ratings of usefulness and high steer-touch response times compared to sound-only and vibration-only take-over requests. Augmented visual feedback (Carpet, Arrow) has the potential to enhance decision making (e.g., whether the driver appropriately implements a steering or braking action), but should be implemented with care. Visual feedback tends to be dominant over feedback in other modalities, and if augmented visual feedback does not provide semantically meaningful information (cf. Sphere), then the driver may respond inappropriately and self-report ratings of usefulness may be impaired. It should be noted that the present results do not necessarily generalize to all types of visual feedback. For example, recently the use of ambient light was found to be promising as take-over request (Borojeni et al., 2016).

Vibrotactile take-over requests were found to be useful for getting the driver back into the loop, even when presented in isolation. However, the effectiveness of directional (left vs. right) vibrations in the driver seat may be limited as compared to non-directional vibrations. Another limitation of vibrotactile feedback is that the driver and the source of vibrations have to be in physical contact with each other (Petermeijer et al., 2016).

The beeps yielded low satisfaction ratings, and were rated as less satisfactory than speech-based take-over requests. However, beeps may be a useful channel for conveying a sense of urgency, and the inter-beep interval is a useful moderator variable in this regard (see Fig. 1, cf. parking sensors). Our results may be specific to the type of beeps used in the experiments. It is possible that other types of beeps, earcons, or speech-based take-over requests would yield high satisfaction ratings (for more research into speech-based take-over requests, see Bazilinskyy & De Winter, 2017; Gold et al., 2015; Politis et al., 2015; Walch et al., 2016; Mok et al., 2015).

In order to counteract the limitations of unimodal take-over requests, the use of multimodal take-over requests may be promising. Multimodal feedback increases the redundancy of the warning and consequently reduces the probability of misses, as compared to unimodal feedback. By means of a crowdsourced online questionnaire, we asked the opinion of 1,692 people on auditory, visual, and vibrotactile take-over requests in highly automated driving (Bazilinskyy et al., 2018). The survey included recordings of auditory messages and illustrations of visual and vibrational messages. The results of the survey showed, consistent with the literature, that multimodal take-over requests were the most preferred option in high urgency scenarios. Furthermore, in a driving simulator experiment (Petermeijer et al., 2017a), we found that drivers showed a faster initial response to multimodal (i.e., auditory and vibrotactile) than vibrotactile-only take-over requests.

Future research should seek ways to maximize the usefulness and satisfaction of take-over requests by finding the right combination of auditory, vibrotactile, and visual feedback. Here attention should be paid to temporal and semantic congruence.

3.4.8 Supplementary material

Supplementary material for this paper is accessible online: https://www.dropbox.com/sh/sb2180f8t27hw3x/AAAMIuifV7NlVv6T3xqpXW8ja?dl=0

3.4.9 References

Bazilinskyy, P., & De Winter, J. C. F. (2017). Analyzing crowdsourced ratings of speech-based take-over requests for automated driving. *Applied Ergonomics, 64*, 56–64.

Bazilinskyy, P., Petermeijer, S. M., Petrovych, V., Dodou, D., & De Winter, J. C. F. (2018). Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays. *Transportation Research Part F: Traffic Psychology and Behaviour*, *56*, 82–98.

Borojeni, S. S., Chuang, L., Heuten, W., & Boll, S. (2016). Assisting drivers with ambient take-over requests in highly automated driving. Proceedings of *the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 237–244.

Cabrall, C., De Winter, J. C. F., Manca, L., Petermeijer, S., Eriksson, A., Bazilinskyy, P., Petrovych, V., & Happee, R. (2017). *A survey of human machine interfaces in automated driving*. Manuscript in preparation.

Gold, C., Berisha, I., & Bengler, K. (2015). Utilization of drivetime: Performing non-driving related tasks while driving highly automated. Proceedings of *the Human Factors and Ergonomics Society 59th Annual Meeting, 59*, 1666–1670.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). Take over!' How long does it take to get the driver back into the loop? Proceedings of *the Human Factors and Ergonomics Society 57th Annual Meeting*, 1938–1942.

Eriksson, A., Petermeijer, S. M., Zimmerman, M., De Winter, J. C. F., Bengler, K. J., & Stanton, N. (2017). *Rolling out the red (and green) carpet: supporting driver decision making in automation to manual transitions*.

Manuscript submitted for publication.
Eriksson, A., & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors, 59*, 689–705.
Favarò, F., Eurich, S., & Nader, N. (2018). Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations. *Accident Analysis & Prevention, 110*, 136–148.
Hellier, E., Edworthy, J., Weedon, Walters, K., & Adams, A. (2002). The perceived urgency of speech warnings: Semantics versus acoustics. *Human Factors, 44*, 1–17.
Mok, B., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., & Ju, W. (2015). Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles. Proceedings of *the 2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2458–2464.
Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230–253.
Petermeijer, S., De Winter, J. C. F., & Bengler, K. (2016). Vibrotactile displays: A survey with a view on highly automated driving. *IEEE Transactions on Intelligent Transportation Systems, 17*, 897–907.
Petermeijer, S. M., Bazilinskyy, P., Bengler, K. J., & De Winter, J. C. F. (2017a). Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics, 62*, 204–215.
Petermeijer, S. M., Cieler, S., & De Winter, J. C. F. (2017b). Comparing spatially static and dynamic vibrotactile take-over requests in the driver seat. Accident Analysis and Prevention, 99, 218–227.
Petermeijer, S. M., Doubek, F., & De Winter, J. C. F. (2017c). Driver response times to auditory, visual and tactile take-over requests: A simulator study with 101 participants. *IEEE International Conference on Systems, Man, and Cybernetics*, Banff, Canada, 1505–1510.
Politis, I., Brewster, S., & Pollick, F. (2015). Language-based multimodal displays for the handover of control in autonomous cars. Proceedings of *the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 3–10.
Van der Laan, J. D., Heino, A. & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research Part C: Emerging Technologies, 5, 1–10.
Walch, M., Lange, K., Baumann, M., & Weber, M. (2016). Autonomous driving: investigating the feasibility of car-driver handover assistance. Proceedings of *the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 11–18.

4 AUDITORY FEEDBACK FOR SITUATION AWARENESS DURING HIGHLY AUTOMATED DRIVING

4.1 Sonifying the location of an object: A comparison of three methods

Corresponding article: Bazilinskyy, P., Van Haarlem, W., Quraishi, H., Berssenbrugge, C., Binda, J., & De Winter, J. C. F. (2016). Sonifying the location of an object: A comparison of three methods. Proceedings of 13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kyoto, Japan.

4.1.1 Abstract

Auditory displays are promising for informing operators about hazards or objects in the environment. However, it remains to be investigated how to map distance information to a sound dimension. In this research, three sonification approaches were tested: Beep Repetition Rate (BRR) in which beep time and inter-beep time were a linear function of distance, Sound Intensity (SI) in which the digital sound volume was a linear function of distance, and Sound Fundamental Frequency (SFF) in which the sound frequency was a linear function of distance. Participants (N = 29) were presented with a sound by means of headphones and subsequently clicked on the screen to estimate the distance to the object with respect to the bottom of the screen (Experiment 1), or the distance and azimuth angle to the object (Experiment 2). The azimuth angle in Experiment 2 was sonified by the volume difference between the left and right ears. In an additional Experiment 3, reaction times to directional audio-visual feedback were compared with directional visual feedback. Participants performed three sessions (BRR, SI, SFF) in Experiments 1 and 2 and two sessions (visual, audio-visual) in Experiment 3, 10 trials per session. After each trial, participants received knowledge-of-results feedback. The results showed that the three proposed methods yielded an overall similar mean absolute distance error, but in Experiment 2 the error for BRR was significantly smaller than for SI. The mean absolute distance errors were significantly greater in Experiment 2 than in Experiment 1. In Experiment 3, there was no statistically significant difference in reaction time between the visual and audio-visual conditions. The results are interpreted in light of the Weber-Fechner law, and suggest that humans have the ability to accurately interpret artificial sounds on an artificial distance scale.

4.1.2 Introduction

Auditory displays can be of value in a broad spectrum of applications, especially in situations where visual feedback is restricted, when the visual system is overburdened, or when the message is short and calls for immediate action (Stanton & Edworthy, 1999). Adding auditory feedback to a human-machine interface may shorten visual search times and reduce the workload compared to using vision only (Perrot et al., 1990; Wickens, 1984).

Usually, auditory feedback takes the form of short warning signals (Patterson, 1982; Stanton & Edworthy, 1999). For example, auditory warnings are used in blind spot monitoring and forward collision warning systems in modern cars (Bazilinskyy et al., 2015; Jamson et al., 2008).

Auditory feedback can also be used to perceptualize objects or activity in the environment, a method which is called sonification (Hermann et al., 2011). One of the earliest known applications of sonification is an optophone. The device, used by the blind, was developed in 1913; it scans text and generates time-varying chords of tones to identify letters (Capp & Picton, 2000). One of the most successful examples of sonification is the Geiger counter, in which auditory clicks are produced to represent ionization events. The Geiger counter was developed in the early 1900s, and is still used today to measure the level of radiation in the environment (Knoll, 2010). An auditory pulse-oximeter, a device similar to the Geiger-counter, was used in hospitals in the United States in 1980's. It generated a tone that varied in pitch based on the level of oxygen in patient's blood (Kramer et al., 1999). Spain et al. (2007) investigated the implications of the use of sonified feedback during a patient monitoring task. They found that a short inter-pulse time contributes to a higher level of perceived urgency.

Sonification is also useful in the field of data analysis, in which case it is sometimes called audification or auditory graphing (Flowers, 2005). During the Voyager 2 space mission, the control encountered a problem when the spacecraft was going through the rings of Saturn. The unexpected behaviour could not be explained by means of a visual analysis of the data. When the data was played through a music synthesizer, a 'machine gun' sound was heard, leading to the conclusion that the problem was caused by collisions with electromagnetically charged micrometeoroids (Barrass & Kramer, 1999; Kramer et al., 1999).

Sensory substitution of visual information may be of value in supporting persons in locomotion tasks (e.g., Hussain et al., 2014). As early as 1936, De Florez suggested that pilots of aircrafts can benefit from the support of sonified instruments in so-called "blind flying" (De Florez, 1936). Parseihian et al. (2012) studied the mapping of the sonified distance to the actual object's location, and developed a sonified device for visually impaired persons. In the automotive industry, the parking sensor of a modern car is another example of the use of sonification, where an increasingly frequent beep is emitted to indicate that the car approaches an object. Although a parking sensor is a successful demonstration of sonification, it remains to be investigated which sonification method is the most effective for conveying information about distance or the degree of hazard.

Haas & Edworthy (1996) showed that sounds producing the highest level of perceived urgency are sounds of a high beep rate, a high intensity, and high

frequency. This suggests that each of these three dimensions may be intuitive for sonification purposes. A review article of 179 publications related to sonification of physical quantities concurs that pitch (frequency), loudness (e.g., volume, intensity), and duration (e.g., beep time, inter-beep time) are the most often used auditory dimensions for sonification (Dubus & Bresin, 2013). Sanders & McCormick (1987; as cited in Stanton and Edworthy, 1999) on the other hand suggested that the auditory discrimination power of humans is rather limited, and contended that humans can identify only 2 to 3 levels of sound duration, 4 to 5 levels of sound intensity (at a given frequency), and 4 to 7 levels of sound frequency. Zahorik (2002) and Loomis et al. (1998) found that participants consistently underestimated the distance in auditory distance perception tasks. Thus, more fundamental research into the topic of mapping of given auditory cues to the distance needs to be conducted.

As mentioned above, beep time, intensity, and frequency are primary sonification dimensions. The aim of this study was to investigate which of these three sonification dimensions allows a person to most accurately indicate the location of an object. Participants completed two experiments; the first experiment involved one-dimensional distance estimation, whereas the second experiment involved the localization of an object in a two-dimensional plane. The participants were presented with sounds without visual feedback, and subsequently had to click on the screen to locate the object. In an additional Experiment 3 we sought to determine whether directional auditory feedback improves reaction times compared to visual-only feedback.

4.1.3 Method

4.1.3.1 Apparatus

The research was conducted using a computer program created with the Unity game engine (version 4.6.1f1). Razer Electra headphones were used.

4.1.3.2 Auditory feedback

Three types of auditory feedback were tested. The first type was Beep Repetition Rate (BRR), in which the beep time was linearly related to distance with respect to the bottom of the screen. For the closest distance (bottom of the screen), the beep time and inter-beep time were 0.05 s (i.e., 10 beeps per second). For the farthest distance (top of the screen), the beep time and inter-beep time were 0.55 s (i.e., 0.91 beeps per second). BRR resembled the feedback in a parking sensor, in that it ‘beeps’ faster as you are closer to an object. In the BRR condition, the sound volume was 100%, and the frequency of the beeps was 460 Hz. The volume of the laptop computer was set so that 100% sound volume generated by the software was regarded as loud but not uncomfortable.

Second, we tested Sound Intensity (SI), where the volume intensity was linearly related to the distance to the object. The volume was 0% at the top of the screen and 100% at the bottom of the screen. The frequency of the sound was 460 Hz.

Third, we tested the Sound Fundamental Frequency (SFF), where the frequency of the sound was linearly related to the distance. The frequency was 1,076 Hz at the bottom of the screen and 184 Hz at the top of the screen. The volume of the sound was 100%.

4.1.3.3 Participants

Twenty-nine persons (8 females) participated in the experiment. Most participants were students and employees of Delft University of Technology, and were on average 29.6 years old (SD = 15.7 years). None of the participants had a hearing disorder or used hearing aids.

4.1.3.4 Procedure

The participants conducted three experiments in the following order: Experiment 1: Distance estimation, Experiment 2: Distance and angle estimation, and Experiment 3: Reaction time. In Experiment 1 and Experiment 2, the participants completed three sessions, each session with a different sound condition (BRR, SI, SFF). To neutralize the effects of a learning curve, we randomized the order of the three sound conditions. We did, however, have the same order for the sessions in Experiments 1 and 2, to prevent participants from experiencing the same sound method right after each other. In Experiment 3, the participants completed two sessions: No Sound and Sound, in randomized order. Each session consisted of 10 trials. Accordingly, each of the participants completed 80 trials in total (30 in Experiment 1, 30 in Experiment 2, and 20 in Experiment 3). The three experiments are explained below.

4.1.3.5 Experiment 1: Distance estimation

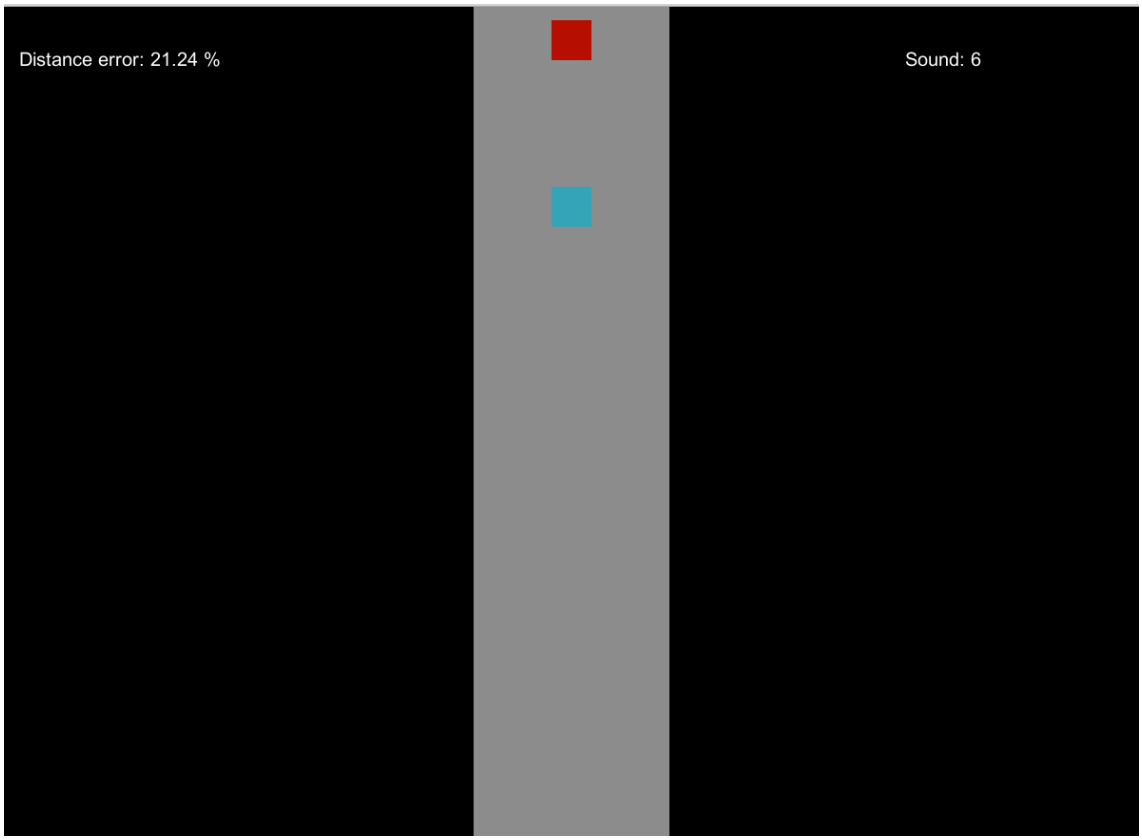


Figure 1. Interface used in Experiment 1.

In the first experiment the participant heard a sound, equally loud in both ears of the headphones. The duration of the sound was 1.0 s for SI and SFF, and 3 beeps for BRR. The participant had to locate the object as accurately as possible by clicking on the screen. Immediately afterwards, the participants were shown

the chosen location (cyan square) and the actual location of the object (red square), as well as an absolute distance error score expressed as a percentage shown in the left top of the screen (Fig. 1). The experiment was preceded by a short automated demonstration in which the participants were presented with 11 sounds from low to high intensity (0%, 10%, …, 100%), together with a corresponding red square on the screen from top to bottom.

4.1.3.6 Experiment 2: Distance and angle estimation

The second experiment was the same as the first, but this time the participant had to locate the object in a two-dimensional plane (Fig. 2). Not only the distance but also the azimuth angle had to be estimated. To represent the angle, we used the volume per ear linearly mapped from the azimuth angle. If the sound volume was 100% in both ears, the object was in front of the participant. If there was only sound in the right ear (right volume = 100%, left volume = 0%), the object was on the right. Sound only in the left ear (left volume = 100%, right volume = 0%) meant that the object was located on the left. Experiment 2 was also preceded by a short demonstration of different distances as in Experiment 1, followed by a presentation of 10 different angles from right to left in 20 deg increments.

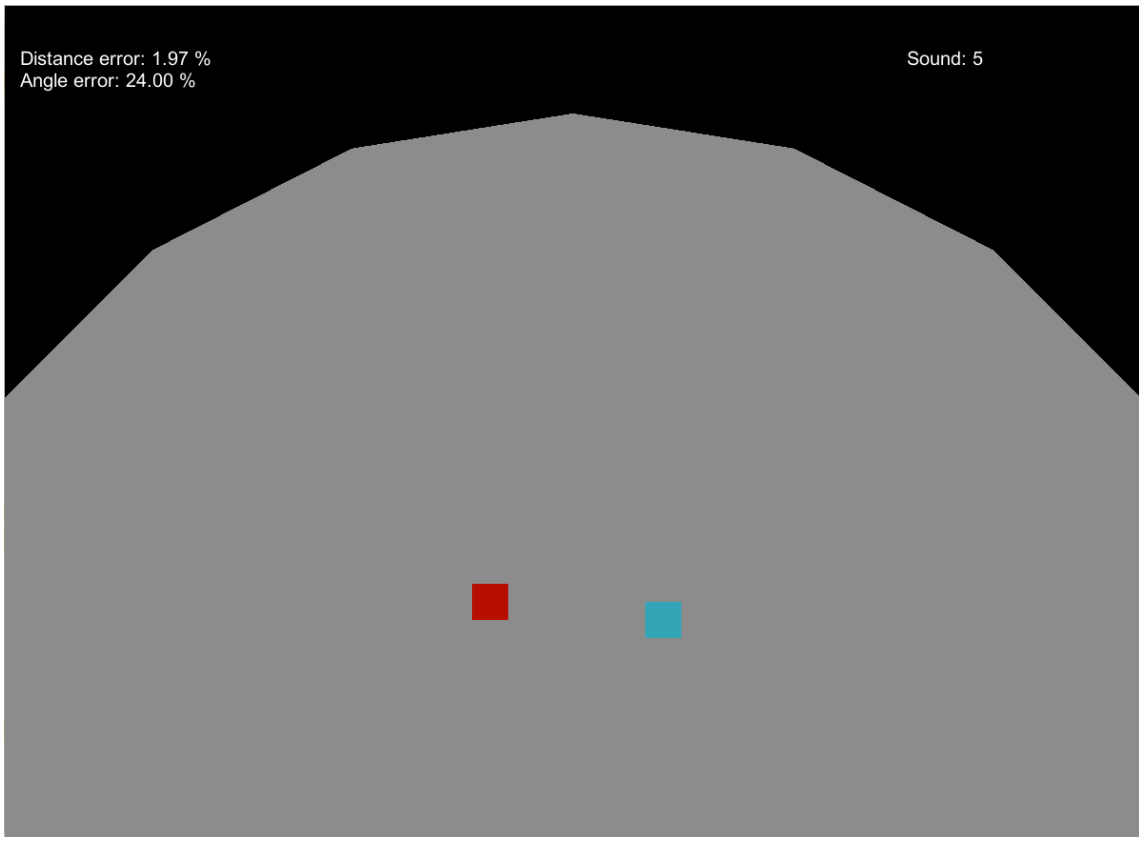


Figure 2. Interface used in Experiment 2.

4.1.3.7 Experiment 3: Reaction time

The third experiment was divided into two sessions. In one session, the participant was presented with a block on the screen (Fig. 3). It could appear on the left, on the right, or in front of the participant. The participant had to press the left, right, or up arrow key as fast as possible. During the second session the

participant both heard a sound corresponding to the location of the block and was presented with a visual representation of the block. After each trial, the reaction time was shown (Fig. 3).

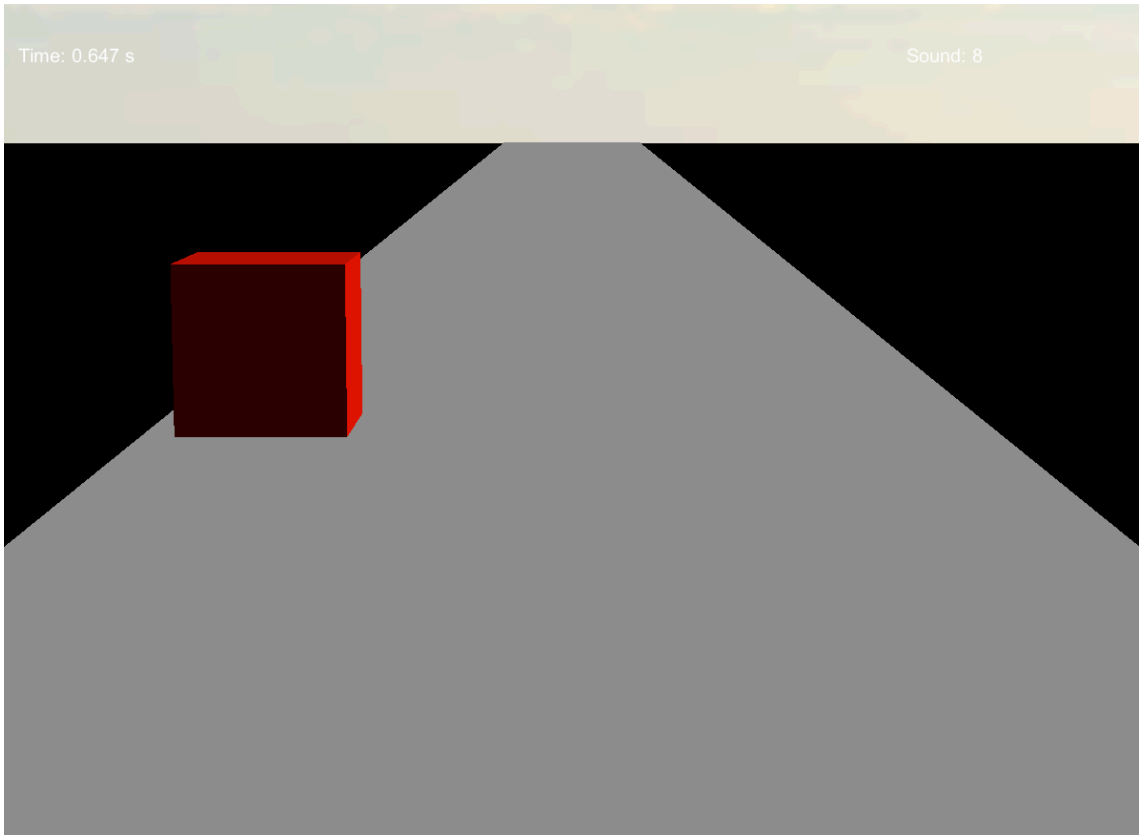


Figure 3. Interface used in Experiment 3.

4.1.3.8 Self-report questionnaires

At the end of the study, participants filled out the NASA TLX for measuring workload (Hart, 2006), complemented with two extra questions “To what extent did you feel motivated while testing?” and “I experienced discomfort (eyestrain, difficulty focussing, pain in ears and/or headache)”. The items consisted of 21-tick scales running from *very low* to *very high*.

4.1.3.9 Procedure and instructions

Prior to the experiment, the participants received a leaflet explaining the three experiments. They were informed that the goal was to locate an object as accurately as possible based on the sound they heard by clicking on the screen, in a one-dimensional space (Experiment 1) or a two-dimensional space (Experiment 2). Regarding Experiment 3, the form stated: “The aim is for you to press on the left, right or up arrow as fast as possible (without making an error).”

4.1.3.10 Statistical analyses

The mean absolute distance error (Experiments 1 & 2), mean absolute angular error (Experiment 2), and mean reaction time (Experiment 3), averaged across trials of a session, were compared between sound conditions with paired *t* tests ($df = N - 1 = 28$).

4.1.4 Results

4.1.4.1 Experiment 1: Distance estimation

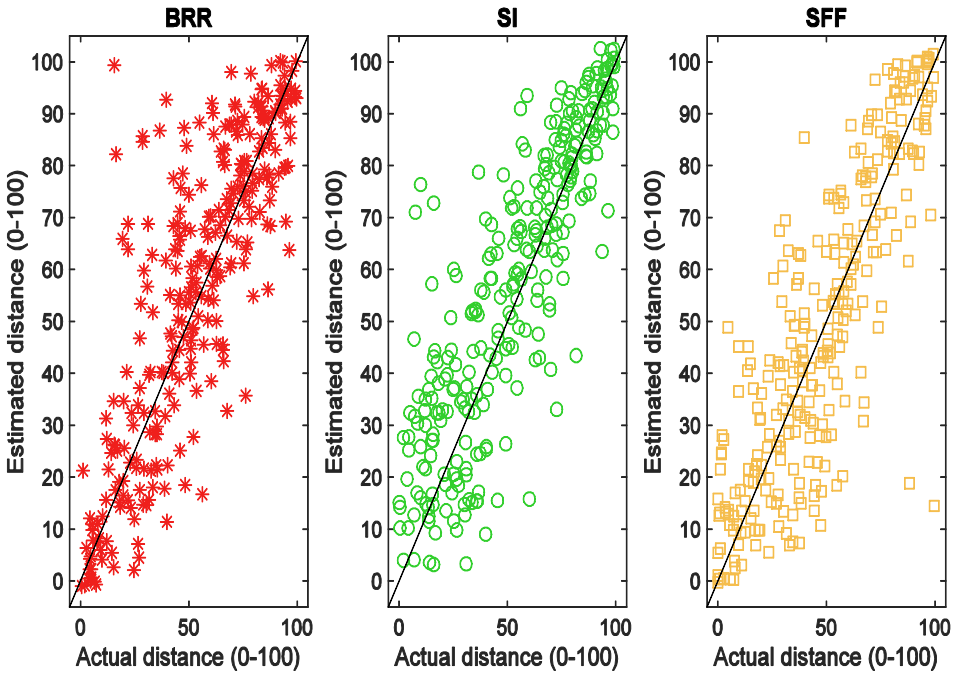


Figure 4. Estimated distance versus actual distance for the three conditions in Experiment 1. 290 values (29 participants * 10 trials) are shown per plot.

The mean absolute distance error on a scale from 0 to 100 was 11.88 (*SD* = 5.26), 11.76 (*SD* = 3.87), and 12.03 (*SD* = 4.10), for BRR, SI, and SFF, respectively. Paired *t* tests revealed no significant differences between BRR and SI (p = .926), BRR and SFF (p = .906), and SI and SFF (p = .791). Figure 4 shows there were no structural under- or overestimations of the error, nor floor or ceiling effects. Figure 5 shows that BRR yielded particularly low errors when the target was near, whereas SI yielded relatively large errors in that case. Figure 6 shows the learning curves. A performance improvement was observed between trials 1–5 versus trials 6–10 (p = .001 for BRR, p = .026 for SI, p = .543 for SFF).

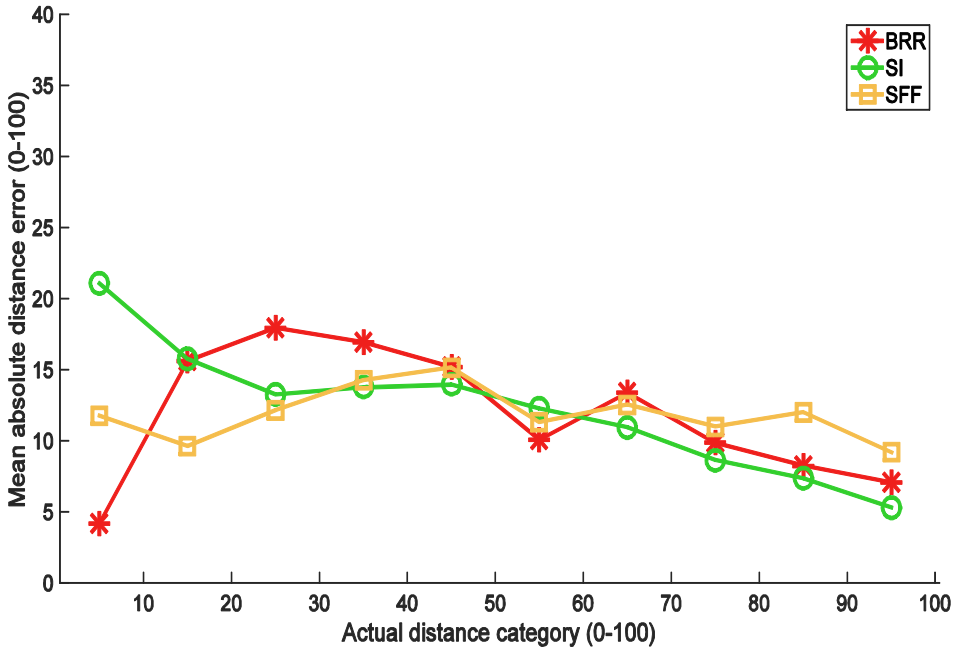


Figure 5. Mean absolute distance error as a function of actual distance in Experiment 1. Ten categories of actual distances were created (0–10%, 10–20%, …, 90–100%).

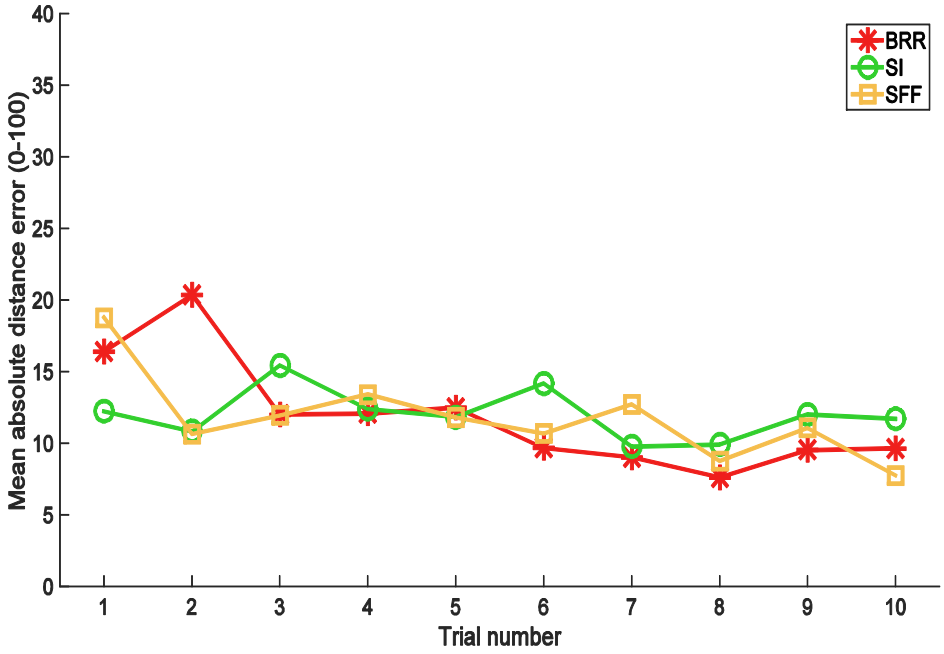


Figure 6. Mean absolute distance error versus trial number in Experiment 1.

4.1.4.2 Experiment 2: Distance and angle estimation

The mean absolute distance error on the scale from 0 to 100 was 16.04 (*SD* = 6.35) for BRR, 20.20 (*SD* = 7.60) for SI, and 18.54 (*SD* = 7.52) for SFF. Paired *t* tests revealed a significant difference between BRR and SI (p = .016), and no significant difference between BRR and SFF (p = .174) nor between SI and SFF (p = .430). Figures 7 and 8 are consistent with the results of Experiment 1 (Figs. 4 and 5), in the sense that BRR performed particularly well when the distance was small, whereas SI performed relatively poor in that case. Figure 9 shows the

learning curve of the distance estimation. A significant improvement was observed for SFF between trials 1–5 versus trials 6–10 (p = .733 for BRR, p = .137 for SI, p = .001 for SFF).

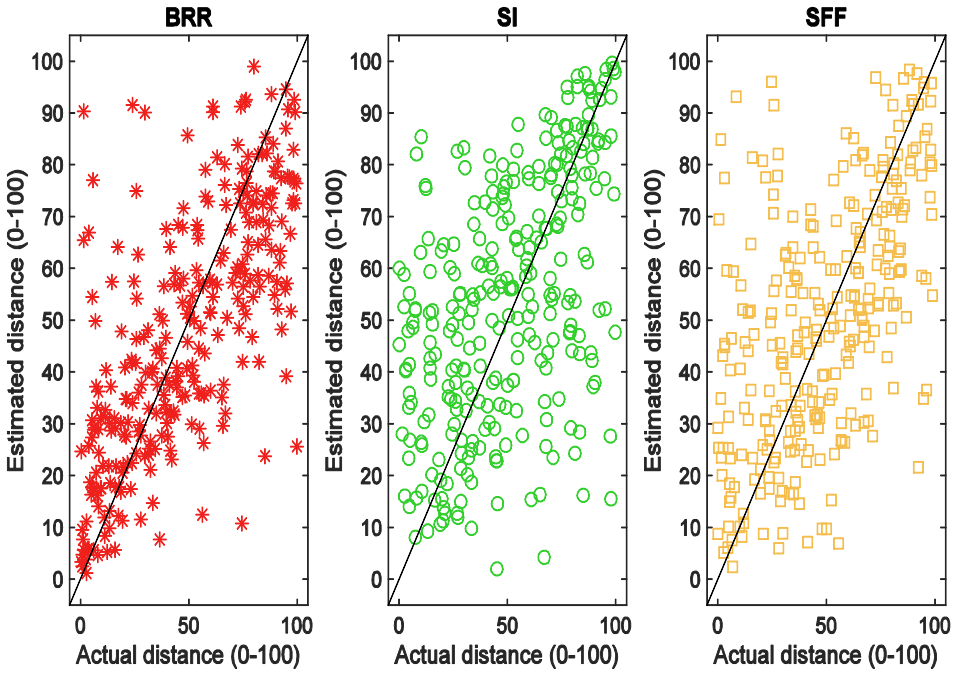


Figure 7. Estimated distance versus actual distance for the three conditions in Experiment 2. 290 values (29 participants * 10 trials) are shown per plot.

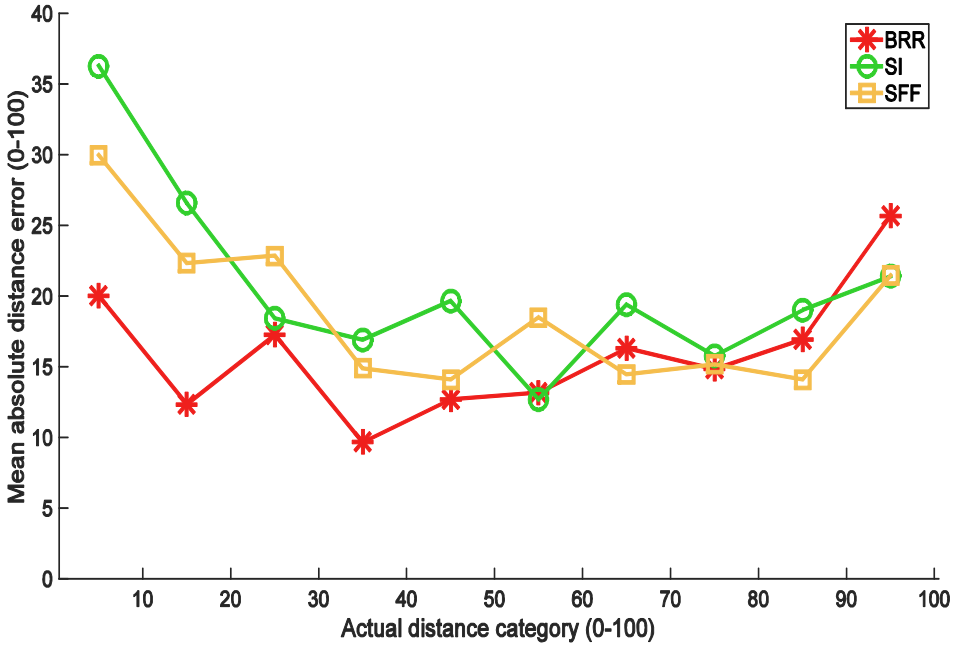


Figure 8. Mean absolute distance error as a function of actual distance in Experiment 2. Ten categories of actual distances were created (0–10%, 10–20%, …, 90–100%).

A comparison of the mean absolute distance errors revealed significant differences between Experiments 1 and 2 (p < .001 for BRR, SI, & SFF), see also Figures 7 and 8 versus Figures 4 and 5.

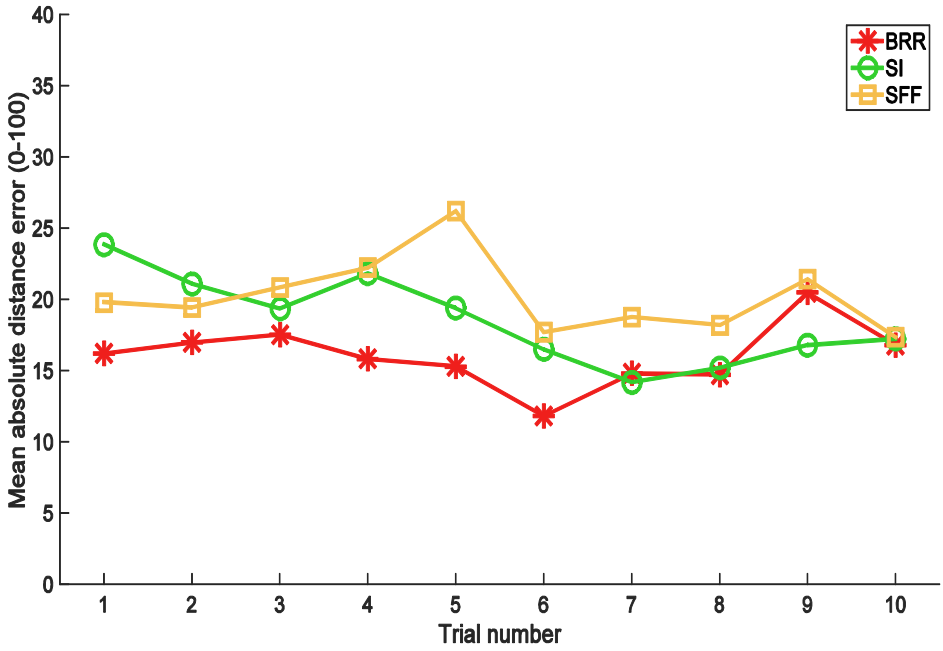


Figure 9. Mean absolute distance error versus trial number in Experiment 2.

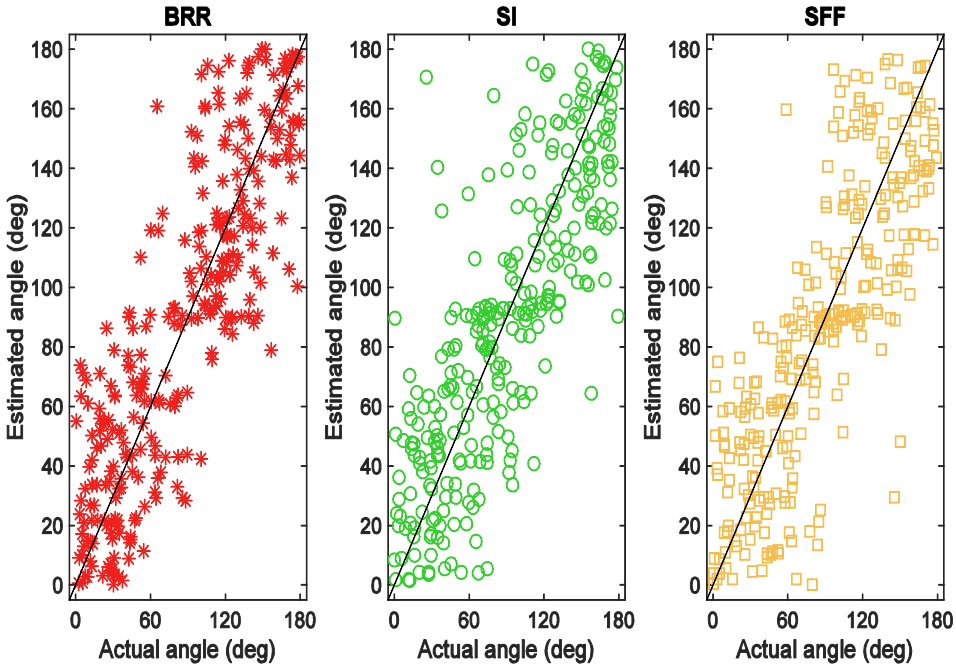


Figure 10. Estimated angle versus actual angle for the three conditions in Experiment 2. 290 values (29 participants * 10 trials) are shown per plot. (0 deg = far right; 180 deg = far left).

The mean absolute angle error was 22.74 deg (*SD* = 7.47) for BRR, 24.53 deg (*SD* = 8.90) for SI, and 24.26 deg (*SD* = 6.08) for SFF. There were no significant differences between BRR and SI (p = .431), between BRR and SFF (p = .325), and between SI and SFF (p = .899). The angular errors are illustrated in Figure 10, and the experience effects are illustrated in Figure 11. There was no significant performance improvement between the first five trials and the second five trials (p = .454, .893, & .564 for BRR, SI, & SFF, respectively).

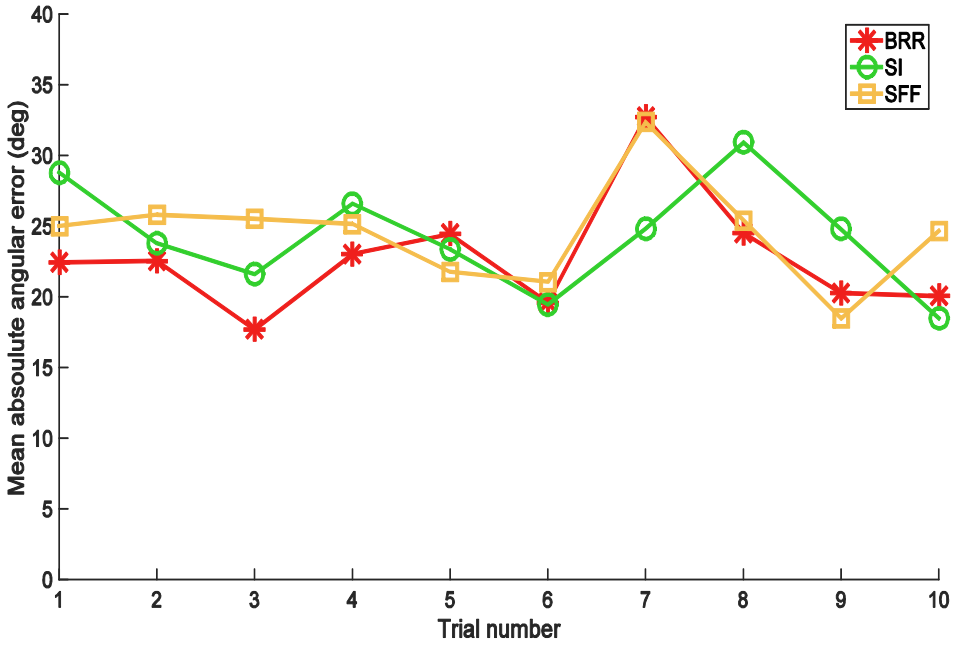


Figure 11. Mean absolute angular error versus trial number in Experiment 2.

4.1.4.3 Experiment 3: Reaction time

The mean reaction times were 0.619 s with sound and 0.666 s without sound. The respective standard deviations among the 29 participants were 0.233 s and 0.267 s. A paired *t* test showed no significant difference in reaction time between the tests with and without sound (p = .366). The error rates were 3.8% (SD = 6.2%) with sound and 4.1% (SD = 6.3) without sound. Figure 12 illustrates the experience effect. A substantial performance improvement can be observed (a comparison of trials 1–5 with trials 6–10 yielded p = .011 for No Sound, and p = .004 for Sound).

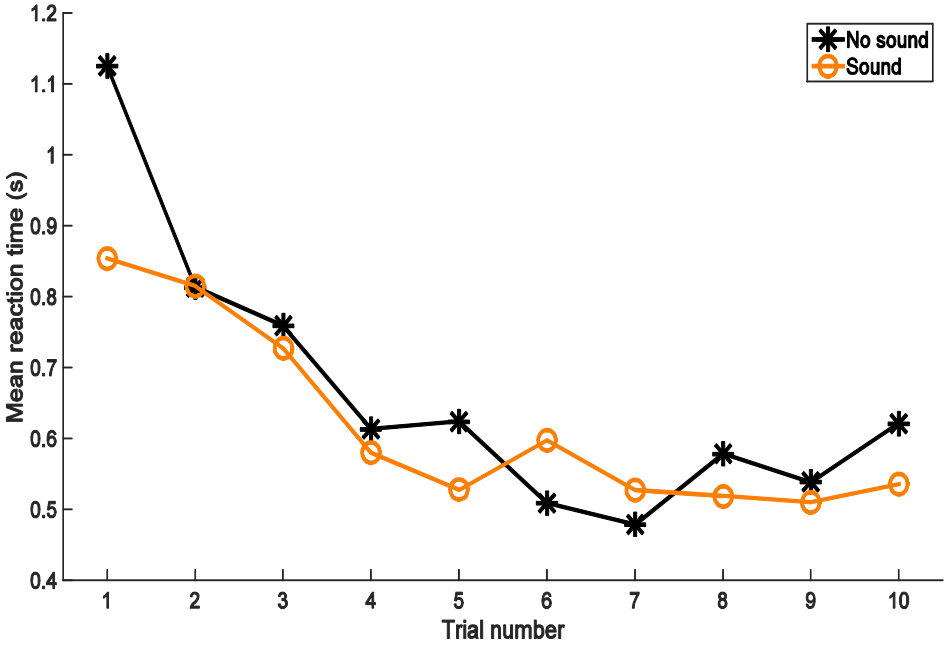


Figure 12. Mean reaction time versus trial number in Experiment 3.

4.1.4.4 Self-report questionnaires

Figure 13 provides a boxplot of the eight questionnaire items. It can be seen that the task was regarded as somewhat mentally demanding, and that people were overall motivated.

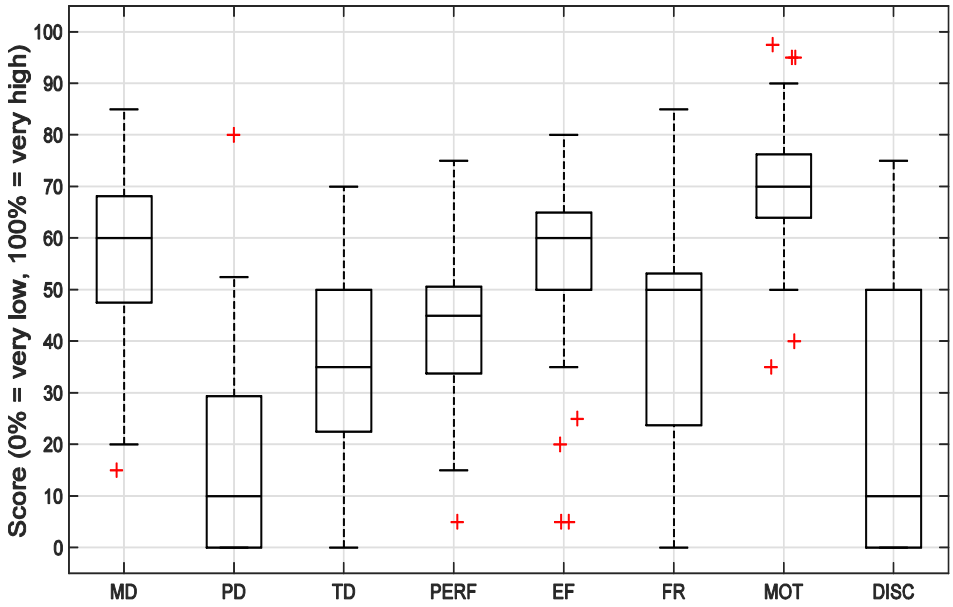


Figure 13. Boxplots for the NASA-TLX and two additional questions. MD = Mental demand, PD = Physical demand, TD = Temporal demand, PERF = Performance, EF = Effort, FR = Frustration, MOT = Motivation, DISC = Discomfort.

4.1.5 Discussion

The aim of this research was to determine which auditory method yields the smallest error between the object's actual and estimated location. For this purpose, in Experiments 1 and 2, we sonified the distance to the object along three primary dimensions: Beep Repetition Rate (BRR), Sound Intensity (SI), and Sound Fundamental Frequency (SFF). Additionally, in Experiment 2 the azimuth angle was sonified to the volume difference between the two ears.

The results revealed no clear-cut differences between the three sound conditions. The three proposed methods (BRR, SI, and SFF) yielded close to equal performance in the distance estimation task in the first experiment. However, in Experiment 2 (the distance and angle estimation task), BRR resulted in a significantly smaller percentage error than SI. BRR performed particularly well compared to SI when the actual distance was small (Figs. 5 and 8). This finding can be explained by the Weber-Fechner law, which states that the just noticeable difference between two stimuli increases linearly with stimulus intensity (e.g., Dehaene, 2003). For example, a difference between 10% and 20% volume is easier to distinguish than a difference between 90% and 100% volume. It is also possible that the sound level was saturated, and that ceiling effects may be the cause of the relatively poor performance in the SI condition when the distance was small. Moreover, according to the sone scale of loudness, how loud a sound is subjectively perceived is nonlinearly related to the physical sound intensity as well as sound frequency (Stevens, 1936). A related practical issue is that absolute sound level is difficult to control and

reproduce on different desktop computers, which each have their idiosyncratic hardware and software configurations. For BRR it was beep time (rather than its reciprocal beep rate) that was linearly related to distance. At small distances, a small increase in beep time represents a large increase in beep rate (e.g., at 0% distance the beep rate was 10 Hz, and at 10% distance, the beep rate was 5 Hz). Thus the sonification of distance to beep rate may allow for sensitive discrimination at small distances. Moreover, BRR is an easy to reproduce and standardize means of sonification.

The participants became better at the tasks with increasing trial number. Experiments 1 and 2 used the same distance sonification. If the learning effect were the only factor affecting the error, one would expect the mean percentage error in the second experiment to be lower than in the first experiment. The results, however, showed the contrary. The mean distance error was statistically significantly higher in the second experiment than in the first experiment. This can be explained by the requirement to multitask in the second experiment: the participants had to divide their attention to determine both distance and angle. The difference in sound intensity in the ears could make it harder for the participants to estimate the distance of the object.

In Experiment 3, we observed no statistically significant difference in the reaction times between visual directional feedback and audio-visual feedback. This may be due to a lack of statistical power, or due to the visual dominance in these types of tasks (Posner et al., 1976).

In conclusion, with appropriate instructions and knowledge-of-results feedback, humans have a discriminating power of beep rate, sound volume, and sound frequency that allows them to map these sound dimensions to a virtual distance.

4.1.6 References

Barrass, S., & Kramer, G. (1999). Using sonification. *Multimedia Systems, 7*(1), 23–31.

Bazilinskyy, P., Petermeijer, S. M., Petrovych, V., Dodou, D., & De Winter, J. C. F. (2015). Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays. *Transportation Research Part F: Traffic Psychology and Behaviour, 56*, 82–98.

Capp, M., & Picton, P. (2000). The optophone: an electronic blind aid. *Engineering Science & Education Journal, 9*(3), 137–143.

De Florez, L. (1936). True blind flight. *Journal of Aeronautical Sciences, 3*(5), 168–170.

Dehaene, S. (2003). The neural basis of the Weber–Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences, 7*(4), 145–147.

Dubus, G., & Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLOS ONE, 8*(12), e82491.

Flowers, J. H. (2005). Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions. Proceedings of *the International Conference on Auditory Display (ICAD)*, Limerick, Ireland, 406–409.

Haas, E. C., & Edworthy, J. (1996). Measuring perceived urgency to create safe auditory warnings. Proceedings of *the Human Factors and Ergonomics Society Annual Meeting*, *40*(16), 845–849.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. Proceedings o*f the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904–908.

Hellier, E., & Edworthy, J. (1999). On using psychophysical techniques to achieve urgency mapping in auditory warnings. *Applied Ergonomics, 30*(2), 167–171.

Hermann, T., Hunt, A., & Neuhoff, J. (2011). The sonification handbook. Logos Publishing House, Berlin.

Hussain, I., Chen, L., Mirza, H. T., Xing, K., & Chen, G. (2014). A comparative study of sonification methods to represent distance and forward-direction in pedestrian navigation. *International Journal of Human-Computer Interaction*, 30(9), 740–751.

Jamson, A. H., Lai, F. C. H., & Carsten, O. M. J. (2008). Potential benefits of an adaptive forward collision warning system. *Transportation Research Part C: Emerging Technologies, 16*(4), 471–484.

Knoll, G. F. (2010). Radiation detection and measurement. John Wiley & Sons, Hoboken, NJ.

Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J., Miner, N., ... & Evreinov, G. (1999). The sonification report: Status of the field and research agenda. report prepared for the national science foundation by members of the international community for auditory display. *International Community for Auditory Display (ICAD)*, Santa Fe, NM.

Loomis, J.M., Klatzky, R.L., Philbeck, J.W., & Golledge, R.G. (1998). Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics, 60*(6), 966–80.

Parseihian, G., Katz, B. F. G., & Conan, S. (2012). Sound effect metaphors for near field distance sonification. Proceedings of *International Conference on Auditory Display (ICAD)*, Atlanta, GA, 18–21 July, 6–13.

Patterson, R. D. (1982). Guidelines for auditory warning systems on civil aircraft (CAA paper 82017). Civil Aviation Authority, London.

Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search performance. *Perception & Psychophysics, 48*(3), 214–226.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review, 83*(2), 157–171.

Sanders, M. S. and McCormick, E. J. (1987). Human Factors in Engineering and Design. McGraw-Hill, New York, NY.

Spain, R. D., Bliss, J. P., & Newlin, E. T. (2007). The effect of sonification pulse rate on perceived urgency and response behaviors. Proceedings of *the Human Factors and Ergonomics Society*, 61(19), 1345–1348.

Stanton, N. A., & Edworthy, J. (1999). Auditory warnings and displays: An overview. In N. A. Stanton & J. Edworthy (Eds.), Human factors in auditory warnings (pp. 3–30). Ashgate Publishing Limited, Aldershot.

Stevens, S. S. (1936). A scale for the measurement of a psychological magnitude: Loudness. *Psychological Review, 43*(5), 405–416.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman and D.R. Davies (Eds.), Varieties of attention (pp. 63–102). Academic Press, New York.

Zahorik, P. (2002). Assessing auditory distance perception using virtual

acoustics. *The Journal of the Acoustical Society of America*, 111(4), 1832–1846.

4.2 Blind driving by means of auditory feedback

Corresponding article: Bazilinskyy, P., Geest, L. Van Der, Van Leeuwen, S., Numan, B., Pijnacker, J., & De Winter, J. C. F. (2016). Blind driving by means of auditory feedback. Proceedings of 13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kyoto, Japan.

4.2.1 Abstract

Driving is a safety-critical task that predominantly relies on vision. However, visual information from the environment is sometimes degraded or absent. In other cases, visual information is available, but the driver fails to use it due to distraction or impairment. Providing drivers with real-time auditory feedback about the state of the vehicle in relation to the environment may be an appropriate means of support when visual information is compromised. In this study, we explored whether driving can be performed solely by means of artificial auditory feedback. We focused on lane keeping, a task that is vital for safe driving. Three auditory parameter sets were tested: (1) predictor time, where the volume of a continuous tone was a linear function of the predicted lateral error from the lane centre 0 s, 1 s, 2 s, or 3 s into the future; (2) feedback mode (volume feedback vs. beep-frequency feedback) and mapping (linear vs. exponential relationship between predicted error and volume/beep frequency); and (3) corner support, in which in addition to volume feedback, a beep was offered upon entering/leaving a corner, or alternatively when crossing the lane centre while driving in a corner. A dead-zone was used, whereby the volume/beep-frequency feedback was provided only when the vehicle deviated more than 0.5 m from the centre of the lane. An experiment was conducted in which participants (N = 2) steered along a track with sharp 90-degree corners in a simulator with the visual projection shut down. Results showed that without predictor feedback (i.e., 0 s prediction), participants were more likely to depart the road compared to with predictor feedback. Moreover, volume feedback resulted in fewer road departures than beep-frequency feedback. The results of this study may be used in the design of in-vehicle auditory displays. Specifically, we recommend that feedback be based on anticipated error rather than current error.

4.2.2 Introduction

Worldwide, billions of people engage in driving at some stage in their lives. Driving is crucial for economic success, but the corresponding cost is substantial. Over 1 million people die in road traffic crashes each year, and millions more become injured (World Health Organisation, 2015).

Driving is primarily a visual task (Groeger, 2000; Sivak, 1996). To be able to drive safely, drivers need to have a valid estimate of their position in relation to other road users and the road boundaries (Groeger, 2000; Macadam, 2003). However, sometimes, such as in case of fog, rain, or darkness, the visual information from the environment is degraded or absent (e.g., Edwards, 1999; Smith, 1982). Relevant visual information may also be unavailable because of

occlusion by other road users or buildings, or when an object is in the blind spot (North, 1985; Staubach, 2009).

Even when visual information is available, the driver may fail to use it. In a naturalistic driving study, it was found that 78% of crashes involved a driver looking away from the forward road just prior to the crash (Klauer et al., 2006). This finding is consistent with a literature review of 50 years of driving safety research, which concluded that most crashes occur because "drivers fail to look at the right thing at the right time" (Lee, 2008, p. 525). Moreover, people tend to underestimate distance (Baumberger et al., 2005; Teghtsoonian and Teghtsoonian, 1970) and speed (Recarte and Nunes, 1996). In addition, there are large individual differences in visual ability. Contrast sensitivity, perceptual speed, and useful field of view decline substantially with age (Janke, 1994; Kline and Fuchs, 1993; Salthouse, 2009; Sekuler et al., 2000). Thus, there appears to be a need for assistive technology that supports the driver when visual information from the environment is degraded, or when the driver fails to process the available visual information.

The auditory modality is promising for warning or supporting human operators, because humans can receive auditory information from any direction, irrespective of the orientation of their head and eyes (Sanders and McCormick, 1987; Stanton and Edworthy, 1999). Furthermore, the ears can receive information at any moment, and humans have the ability to focus selectively on one sound in situations where multiple auditory signals are present (Hermann et al., 2011). Not surprisingly, various types of auditory displays (in the form of forward collision warning systems, parking assistance systems, and blind spot monitoring systems) are available on the market and have been found to improve road safety (e.g., Piccinini et al., 2012). Moreover, auditory feedback systems have been designed that support drivers in case visual information is unavailable (Colby, 2012; Hong et al., 2008; Verbist et al., 2009). As part of the Blind Driver Challenge, Hong et al. (2008) developed an auditory and vibrotactile feedback system that relays information to the driver about the car speed and movement direction. Verbist et al. (2009) proposed two continuous auditory displays based either on brown noise or a melody for supporting the lane-keeping task in the absence of visual information; both displays proved to be capable of supporting such a task.

Outside the domain of driving, the potential of auditory feedback has been studied as well. For example, auditory feedback was found to be effective for supporting blindfolded participants in steering a powered wheelchair (Vinod et al., 2010). In Simpson et al. (2008) the vision of pilots in actual flight was occluded by goggles, and an auditory artificial horizon was used for attitude identification and for recovering from displaced aircraft attitudes. The results showed that the pilots were able to manoeuver the aircraft within its flight envelope by means of auditory feedback only (and see De Florez, 1936, for a classic study on 'blind flight'; also Wickens, 1992, pp. 480–481). Vinje and Pitkin (1972) showed that participants performed a tracking task equally well when the tracking error information was provided via an auditory or a visual display.

Can driving be performed without any visual feedback? Without alternative feedback, this is impossible because drivers need to visually sample the road about every 4 s to keep the car on the road (Godthelp, Milgram, and Blaauw, 1984). Google put Steve Mahan, who lost 95% of his vision, behind the steering wheel of one of their prototypes of fully automated cars (Prince, 2012).

Mahan was able to get to a restaurant and pick up his dry cleaning. However, substantial technological advances are required before self-driving cars can be put on the road (Shladover, 2015). Unless the driving task is wholly automated, humans have a crucial role in the driving task, and could benefit from real-time feedback

This study explored whether driving can be performed as an auditory task without any visual feedback. Specifically, we looked at lane keeping, a task that has to be conducted permanently and is crucial for safe driving (Brookhuis and De Waard, 1993). By means of this research, we aimed to generate knowledge that may be of value in the design of in-vehicle auditory displays. One example of such an application may be a situation where a driver falls asleep behind the wheel or is visually distracted, in which case appropriate (directional) auditory feedback could warn and support him/her in regaining control.

In the design of driver support systems and in the modelling of driver behaviour, a predictor time is often used (e.g., Donges, 1978; Hellström et al., 2009; Hingwe and Tomizuka, 1998; Petermeijer et al., 2015). This means that the driver responds to a predicted error rather than to the current error. It has also been advised to use graded (i.e., increasing with deviation from a target) instead of binary feedback (e.g., Lee et al., 2004; Wolf and Nees, 2015). Therefore, we tested the effectiveness of graded predictor feedback in our research.

4.2.3 Method

4.2.3.1 Apparatus

Figure 1. The driving simulator used in this research. In all trials, the visual projection was shut down.

For this research, we used a fixed-base driving simulator (Fig. 1; Green Dino, the Netherlands). An interface was programmed in MATLAB/Simulink r2015a to retrieve data from the simulator and to generate audio output via Creative Sound Blaster Tactic 3D Alpha headphones. The participants were able to hear engine and tire sounds via loudspeakers mounted in the simulator. During the experiment, the LCD projectors of the simulator were turned off. The width of the car was 1.76 m and its length was 4.22 m.

4.2.3.2 Track

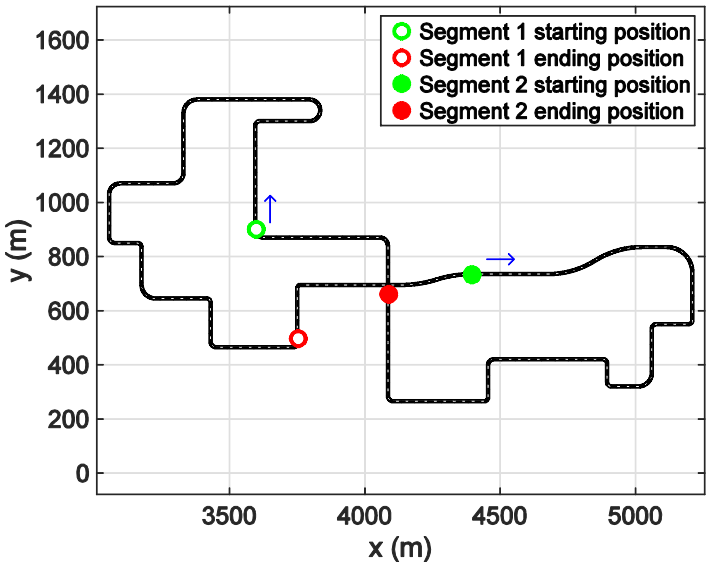


Figure 2. Top view of Segment 1 and Segment 2 of the test track. x and y are Cartesian coordinates in meters.

The track was a two-lane 7.5-km road without intersections and without other road users. It contained straight segments and sharp 90-degree corners, most of which had a radius of about 20 m (for research using the same track, see De Groot et al., 2012; Van Leeuwen et al., 2014, 2015). The lane width was 5 m. There were two starting points, yielding two different segments (Fig. 2). In each trial, the participant drove 3 km which took on average 4.80 min (*SD* = 0.72 min, *N* = 44).

4.2.3.3 Participants

The participants were two experienced drivers (two of the authors) with good knowledge of the auditory feedback concepts and the track.

4.2.3.4 Speed and gearbox settings

An automatic gearbox was used. The speed of the car was predetermined so that the participants did not use the pedals. Fig. 3 (top) illustrates the speed of the car in two left corners followed by a right corner.

4.2.3.5 Parameter sets

Three parameter sets were tested in the following order per participant: 1) predictor time (consisting of 4 conditions), 2) feedback mode and mapping (consisting of 4 conditions), and 3) corner support (consisting of 3 conditions).

Each participant tested each condition once on Segment 1 and once on Segment 2 (Fig. 2). The conditions and segments were randomized within each parameter set.

In all three parameter sets, a dead zone was used based on De Groot et al. (2011), see also Horiguchi et al. (2013) for the advantages of a dead zone in sonification for a manual control task. Thus, the volume and beep-frequency feedback were provided only when the predicted position of the car deviated from the centre of the right lane by more than 0.5 m.

Parameter set 1: Predictor time

The predicted lateral error was calculated by extrapolating the current position of the centre of the car (x, y), by t_{pred} seconds using the velocities in world coordinates (v_x, v_y). Volume feedback was used in this design. Specifically, the volume (on a scale from 0 to 1) of a 464 Hz tone became linearly louder with increasing predicted error with respect to the centre of the right lane (e) as follows: $0.1*|e-0.5|$ (Fig. 4, left). The participants had to steer away from the sound. That is, sound on the left was produced when the predicted lateral error was left of the centre of the right lane, whereas sound on the right was produced when the error was right of the lane centre.

Fig. 3 (middle & bottom) illustrates the working mechanism of the predictor. It can be seen that the larger the t_{pred}, the larger the difference between predicted and current error. For example, for t_{pred} = 3 s, at a travelled distance of 1700 m, the participant was left of the lane centre, while the predictor indicated that the car ends up to the right in 3 s time.

Parameter set 2: Feedback mode and mapping

We evaluated linearly graded volume feedback (VL), exponentially graded volume feedback (VE), linearly graded beep-frequency feedback (FL), and exponentially graded beep-frequency feedback (FE).

The linear volume was the same as in Parameter set 1, whereas the exponential volume was defined as $0.02*e^{|e-0.5|}$. In the beep feedback, the inter-beep time (IBT) was varied as a function of the predicted lateral error. In the FL condition, the reciprocal of the IBT was linearly related to the predicted error, whereas in the FE condition this was an exponential relationship. For both the FL and FE conditions, the beep duration was 0.14 s. Fig. 4 illustrates how the auditory feedback became louder, and the inter-beep interval became shorter, with predicted error.

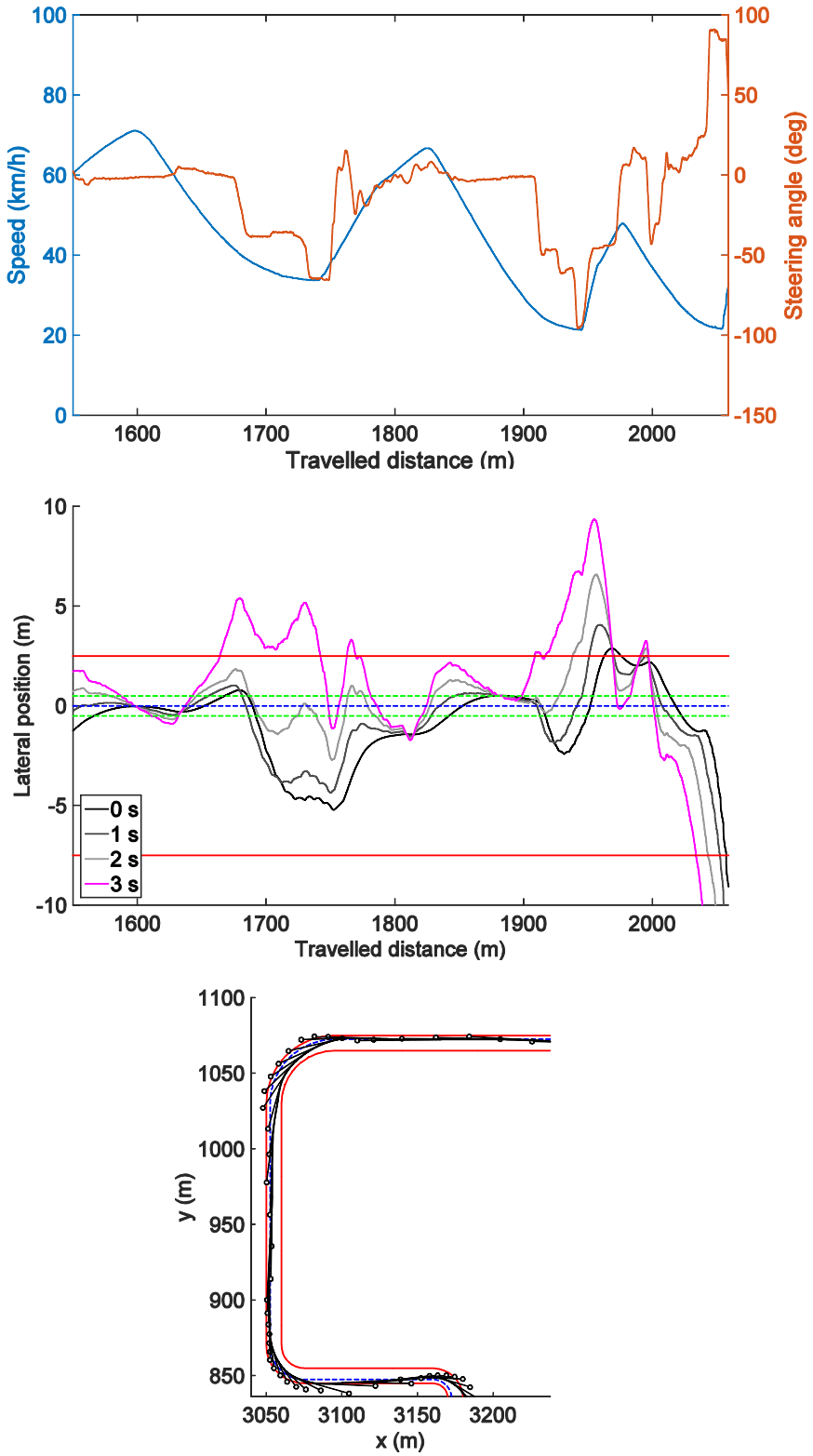


Figure 3. Working mechanism of the predictor feedback. The three figures correspond to the same selected part of the route (participant 1, Segment 1, used predictor time = 3 s) consisting of two left corners followed by a right corner (radii of the centre of the right lane = 42.5 m, 22.5 m, & 12.5 m, respectively). The participant departed the road in the right corner. *Top* = Speed and steering angle versus travelled distance. *Middle* = Lateral error with respect to the centre of the right lane versus travelled distance, for t_{pred} = 0 s (i.e., the actual lateral error), 1 s, 2 s, and 3 s. The green lines at -0.5 m and 0.5 m represent the bandwidth of the feedback. *Bottom* = The path driven by the participant. The circular markers represent the predicted position of the car with t_{pred} =

3 s, calculated every 1 s. For the middle and bottom figures, the blue dashed line represents the right lane centre, and red lines represent the lane boundaries. For these figures, a low-pass filter was used, as signals were somewhat noisy.

As in parameter set 1, the sounds were directional: sound on the right was produced when deviating to the right, and vice versa. In all cases, a predictor time of 2 s was used, and the sound was a 464 Hz tone as in Parameter set 1.

Parameter set 3: Corner support

In this design, t_{pred} was 2 s, and linear graded volume feedback (VL) was used. An additional corner support was implemented, allowing drivers to infer when to make small corrections on straights and when to make large required steering angles in corners. Three concepts were tested. The first concept did not involve corner support. In the second concept, a beep was produced when the car entered and when the car left a corner. When entering a left corner, a beep on the right was produced, while when entering a right corner, a beep on the left was produced. When leaving a corner, a beep was produced both on the left and on the right. The third concept provided a beep on the left and right when crossing the lane centre in corners (i.e., when the sign of the predicted error changed).

In Parameter set 3, the linear graded volume feedback with t_{pred} = 2 s was used. A 565 ms long beep of 2165 Hz was used as corner support. The volume of this beep was constant.

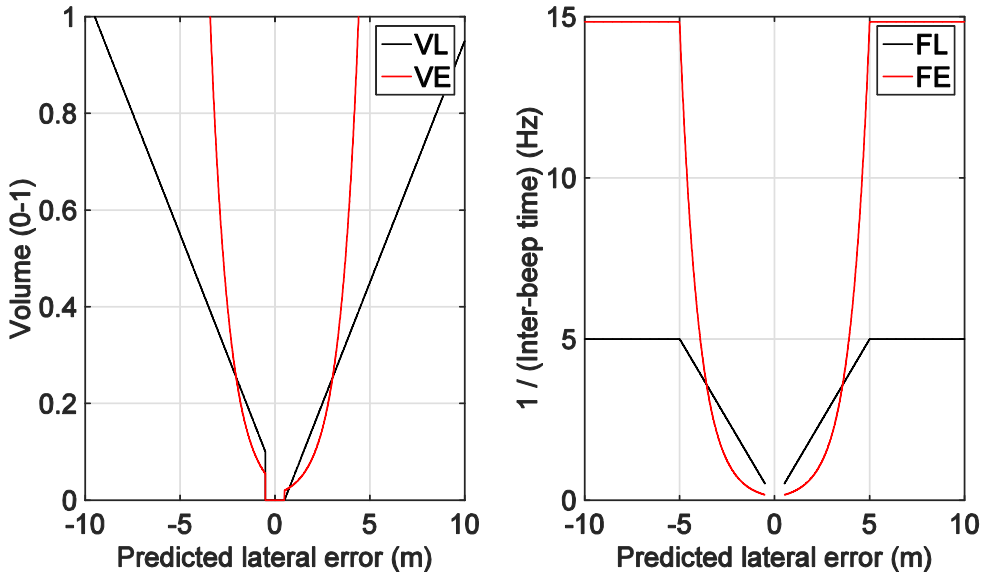


Figure 4. Relationship between predicted lateral error and volume (left) / beep frequency (right)

4.2.3.6 Dependent measures

The participants' driving performance was evaluated using: 1) the on-target percentage (OTP), being the percentage of time the centre of the car was within 0.5 m of the centre of the right lane, and (2) the number of resets. A reset (i.e., road departure) occurred when all four edges of the car were outside the road boundaries. After each reset, the car was automatically placed back in the centre of the right lane with zero speed. For calculating OTP, data between 3 s prior to

10 s after each reset were removed to prevent a causal influence from resets on OTP. The predetermined speed was reached about 5 s after the reset.

4.2.4 Results

Fig. 5 shows the effects of driving with different prediction times. It can be seen that the number of resets was highest when t_{pred} = 0 s. Specifically, there were about 30 resets per drive without prediction, and no more than 12 with prediction. There were no clear effects of predictor time on the OTP, with both participants driving within the 1-m wide dead-zone about 25 to 40 % of the time for all predictor times. Fig. 6 shows that the volume feedback was more effective than the beeping feedback in terms of the number of resets. The differences between the linear and exponentially graded feedback were small. Fig. 7 shows that the corner support had no consistent effect on the number of resets and the OTP.

To elucidate why the lack of prediction yielded a high number of resets, we inspected the driven paths. Fig. 8 shows the paths of the two participants for the same road segment as depicted in Fig. 3. It can be seen that with t_{pred} = 0 s the participants often left the road, even on straight road segments. The participants veered off the road on the outside of the corner, indicating that they were too late with providing a steering input.

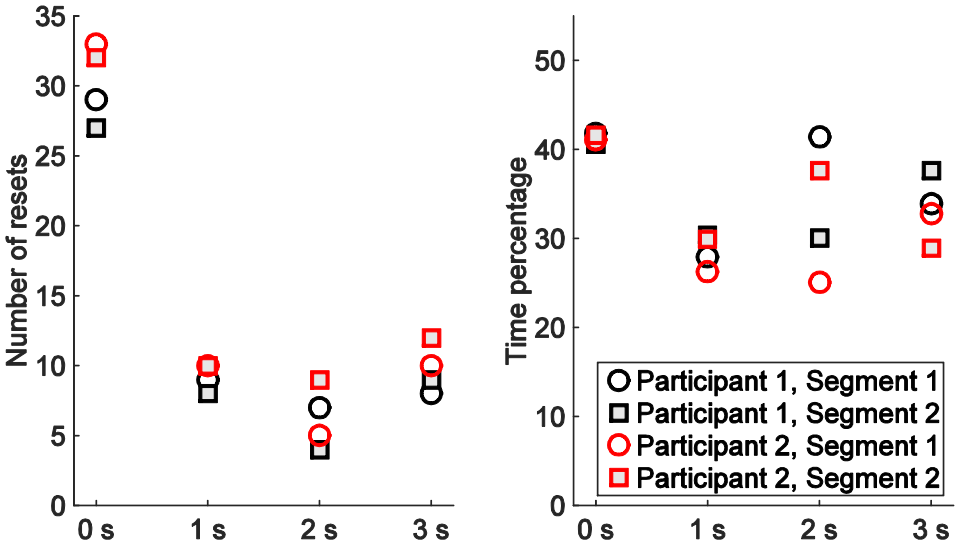


Figure 5. Results for Parameter set 1 (Predictor time). Linear volume feedback was used.

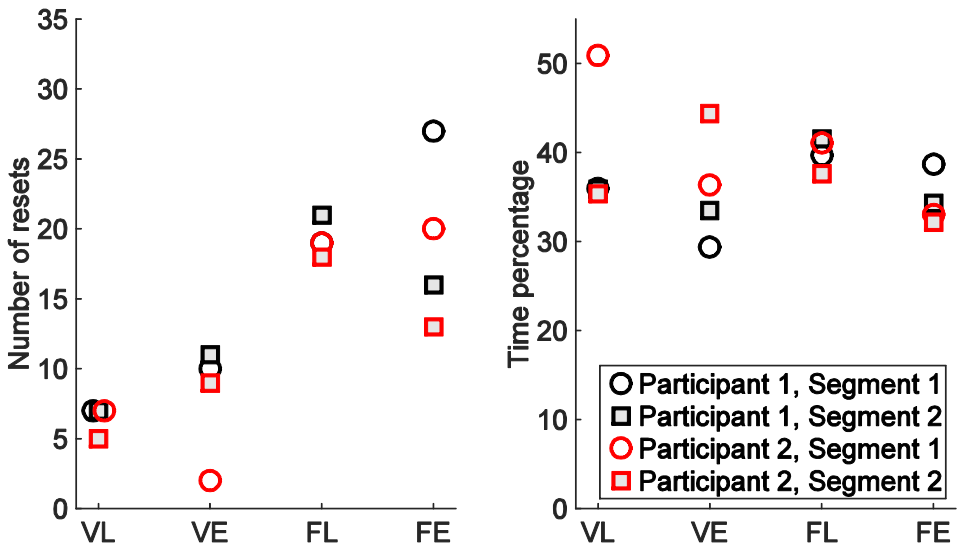


Figure 6. Results for Parameter set 2 (Feedback mode & mapping). VL = linear graded volume feedback, VE = exponential volume feedback, FL = linear graded beep-frequency feedback, FE = exponential beep-frequency feedback. t_{pred} was 2 s.

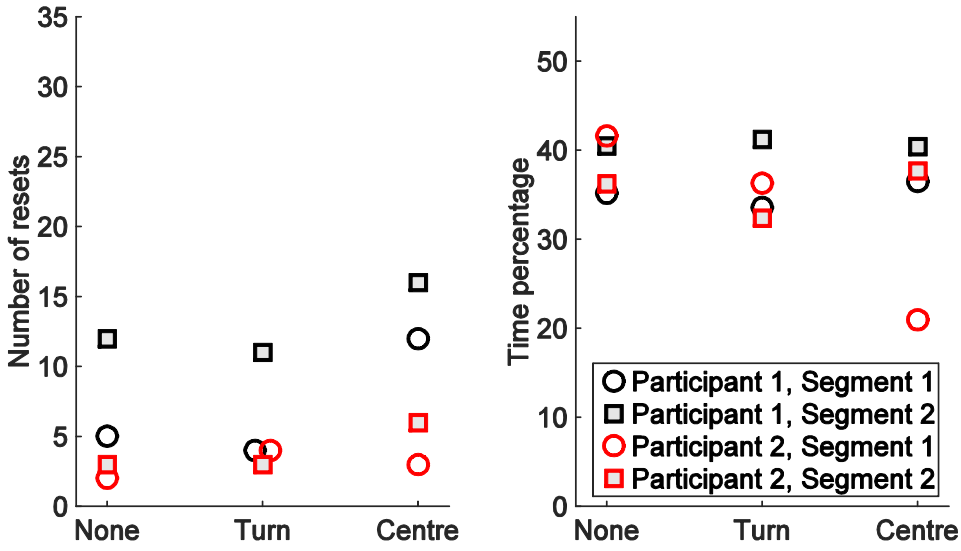


Figure 7. Results for Parameter set 3 (Corner support). None = no corner support, Turn = beep when entering and leaving a corner, Centre = beep when crossing road centreline in a corner. t_{pred} was 2 s, and linear volume feedback was used.

4.2.5 Discussion

This research sonified the predicted lateral error of the car in a driving simulator experiment involving two experienced drivers with good knowledge of the auditory feedback and the test track. Moreover, we evaluated volume versus beep-frequency feedback, both with a linearly and exponentially graded dependency on the predicted lateral error. The 'blind' drivers were also given support in corners in the form of beeps issued upon entering and exiting corners, or when crossing the centre of the right lane in corners.

The prediction time of 0 s resulted in a large number of road departures. With 0 s prediction, participants were often too late in compensating for errors

from the lane centre (cf. Fig. 8). The auditory feedback linked to a predicted lateral error effectively supported the blinded participants in performing a lane-keeping task. One of the participants drove very well with the volume feedback combined with t_{pred} = 2 s, resulting in 'only' two resets in 3 km of driving (see Fig. 6: VE condition & Fig. 7: None condition). In summary, substantial improvements were obtained compared to driving with t_{pred} = 0 s, a condition that resulted in 30 resets per drive (Fig. 5).

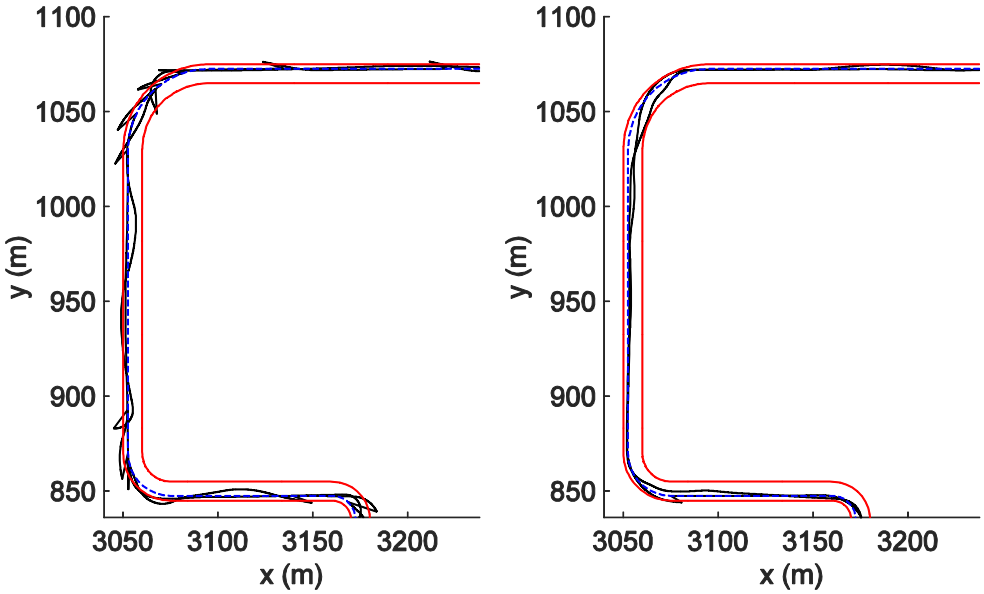


Figure 8. Paths driven by the participants in Parameter set 1 when using t_{pred} = 0 s (left) and when using t_{pred} = 2 s (right).

Panëels et al. (2013) found that continuous guidance of visually impaired during a walking task was more effective than intermittent guidance. Similarly, we found that the volume feedback was more effective than the beep-frequency feedback, possibly because the former provided continual feedback. The intermittent nature of the low frequency beeps may have made it difficult for the participants to perceive when they were entering or leaving the 1-m wide dead-zone.

There were no substantial differences between systems with or without corner support. It is noted that when approaching a corner, the participants could hear the engine slowing down due to the automated speed control. In other words, the drivers could already infer that they were approaching a corner even without the corner support.

In conclusion, our results show that appropriate auditory support can be effective in conditions where visual information is absent. There were no reset-free runs, which indicates that under the given conditions driving cannot be a purely auditory task. One possible reason for the overall high number of road departures (other than obviously the lack of visual feedback) may be that the driving simulator did not offer tactile or vestibular motion feedback.

Future studies may build on the methods presented in this paper, and focus on the development of a 'blind driving' system by means of multimodal auditory/vibrotactile feedback. Improvement of the system may be achieved by taking into account that most of the road departures occurred in corners. The design of a corner support system that more accurately predicts the future path (e.g., based on steering angle) may prove to be fruitful.

The test track did not feature any stationary or moving obstacles. The speed was not controlled by the driver, which reduces the comparability with real-life driving. Furthermore, the participants were two experienced drivers, and so the results do not reflect the entire driving population. A single-subject experiment design (Sidman, 1960; Horner et al., 2005) was chosen to promote an iterative design approach. The results in this paper represent the first iteration in a series of planned studies on the topic of 'blind driving'.

4.2.6 References

Baumberger, B., Flückiger, M., Paquette, M., Bergeron, J., & Delorme, A. (2005). Perception of relative distance in a driving simulator. *Japanese Psychological Research, 47*(3), 230–237.

Brookhuis, K. A., & De Waard, D. (1993). The use of psychophysiology to assess driver status. *Ergonomics, 36*(9), 1099–1110.

Colby, R. (2012). Considerations for the development of non-visual interfaces for driving applications. *Virginia Polytechnic Institute and State University*.

De Florez, L. (1936). True blind flight. *Journal of the Aeronautical Sciences, 3*, 168–170.

De Groot, S., Centino Ricote, F., & De Winter, J. C. F. (2012). The effect of tire grip on learning driving skill and driving style: A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour, 15*(4), 413–426.

De Groot, S., De Winter, J. C. F., López García, J. M., Mulder, M., & Wieringa, P. A. (2011). The effect of concurrent bandwidth feedback on learning the lane-keeping task in a driving simulator. *Human Factors, 53*(1), 50–62.

Donges, E. (1978). A two-level model of driver steering behavior. Human Factors, 20(6), 691–707.

Edwards, J. B. (1999). The relationship between road accident severity and recorded weather. *Journal of Safety Research, 29*(4), 249–262.

Godthelp, H., Milgram, P., & Blaauw, G. J. (1984). The development of a time-related measure to describe driving strategy. *Human Factors, 26*(3), 257–268.

Groeger, J. A. (2000). Understanding driving: Applying cognitive psychology to a complex everyday task. Hove, East Sussex: Psychology Press.

Hellström, E., Ivarsson, M., Åslund, J., & Nielsen, L. (2009). Look-ahead control for heavy trucks to minimize trip time and fuel consumption. *Control Engineering Practice, 17*(2), 245–254.

Hermann, T., Hunt, A., & Neuhoff, J. (2011). *The sonification handbook*. Berlin: Logos Publishing House.

Hingwe, P., & Tomizuka, M. (1998). A variable look-ahead controller for lateral guidance of four wheeled vehicles. Proceedings of *American Control Conference*, Philadelphia, PA, 31–35.

Hong, D., Kimmel, S., Boehling, R., Camoriano, N., Cardwell, W., Jannaman, G., … & Russel, E. (2008). Development of a semi-autonomous vehicle operable by the visually-impaired. Proceedings of *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems,* 539–544.

Horiguchi, Y., Miyajima, K., Nakanishi, H., Sawaragi, T. (2013). Parameter-mapping sonification for manual control task: Timbre and intensity

manupluration to sonify dynamic system state. Proceedings of *SICE Annual Conference*, 2341–2346.
Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165–179.
Janke, M. K. (1994). Age-related disabilities that may impair driving and their assessment. Sacramento, CA: NHTSA.
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* (Tech. Rep. No. DOT HS 810 594). Washington, DC: NHTSA.
Kline, D. W., & Fuchs, P. (1993). The visibility of symbolic highway signs can be increased among drivers of all ages. *Human Factors, 35*(1), 25–34.
Lee, J. D. (2008). Fifty years of driving safety research. Human Factors, 50(3), 521–528.
Lee, J. D., Hoffman, J. D., & Hayes, E. (2004). Collision warning design to mitigate driver distraction. Proceedings of *Conference on Human Factors in Computing Systems*, Vienna, Austria.
Macadam, C. C. (2003). Understanding and modeling the human driver. *Vehicle System Dynamics, 40*(1–3), 101–134.
North, R. V. (1985). The relationship between the extent of visual field and driving performance—a review. *Ophthalmic and Physiological Optics, 5*(2), 205–210.
Panëels, S. A., Varenne, D., Blum, J. R., & Cooperstock, J. R. (2013). The walking straight mobile application: Helping the visually impaired avoid veering. Proceedings of *the 2013 International Conference on Auditory Display*, Georgia Institute of Technology.
Petermeijer, S. M., Abbink, D. A., & De Winter, J. C. F. (2015). Should drivers be operating within an automation-free bandwidth? Evaluating haptic steering support systems with different levels of authority. *Human Factors*, *57*(1), 5–20.
Piccinini, G. F., Simões, A., & Rodrigues, C. M. (2012). Focusing on drivers' opinions and road safety impact of Blind Spot Information System (BLIS)'. In Stanton, N.A. ed. *Advances in Human Aspects of Road and Rail Transportation.* Boca Raton, FL: CRC Press, 57–66.
Prince, R. (2012). Legally blind man takes to wheel of self-driving car. *The Telegraph*. Available from http://www.telegraph.co.uk/technology/news/9176660/Legally-blind-man-takes-to-wheel-of-self-driving-car.html
Recarte, M. A., & Nunes, L. M. (1996). Perception of speed in an automobile: Estimation and production. *Journal of Experimental Psychology: Applied*, *2*(4), 291–304.
Salthouse, T. A. (2009). Decomposing age correlations on neuropsychological and cognitive variables. *Journal of the International Neuropsychological Society, 15*(5), 650–661.
Sanders, M. S., & McCormick, E. J. (1987). *Human Factors in Engineering and Design.* New York, NY: McGraw-Hill.
Sekuler, B., Bennett, P. J., & Mamelak, M. (2000). Effects of aging on the Useful Field of View. *Experimental Aging Research, 26*, 103–120.
Shladover, S. E. (2015). *Road vehicle automation history, opportunities and*

challenges [Presentation]. Mini-seminar 'developments selfdriving vehicles in USA,. 9 November, Delft, the Netherlands.

Simpson, B. D., Brungart, D. S., Dallman, R. C., Yasky, R. J., & Romigh, G.D. (2008). Flying by ear: Blind flight with a music-based artificial horizon. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52*(1), 6–10.

Sidman, M. (1960). Tactics of scientific research: Evaluating experimental data in psychology. *New York: Basic Books*.

Sivak, M. (1996). The information that drivers use: is it indeed 90% visual? *Perception, 25*(9), 1081–1090.

Smith, K. (1982). How seasonal and weather conditions influence road accidents in Glasgow. *The Scottish Geographical Magazine, 98*(2), 103–114.

Stanton, N. A., & Edworthy, J. (1999). Auditory warnings and displays: An overview. In Stanton, N.A. and Edworthy, J. eds. *Human factors in auditory warnings*. Aldershot: Ashgate Publishing Limited, 3–30.

Staubach, M. (2009). Factors correlated with traffic accidents as a basis for evaluating Advanced Driver Assistance Systems. *Accident Analysis and Prevention, 41*(5), 1025–1033.

Teghtsoonian, R., & Teghtsoonian, M. (1970). Scaling apparent distance in a natural outdoor setting. *Psychonomic Science, 21*(4), 215–216.

Van Leeuwen, P. M., Gómez i Subils, C., Ramon Jimenez, A., Happee, R., & De Winter, J. C. F. (2015). Effects of visual fidelity on curve negotiation, gaze behaviour, and simulator discomfort. *Ergonomics, 58*(8), 1347–1364.

Van Leeuwen, P. M., Happee, R., & De Winter, J. C. F. (2014). Vertical field of view restriction in driver training: A simulator-based evaluation. *Transportation Research Part F: Traffic Psychology and Behaviour, 24*, 169–182.

Verbist, K., Boer, E. R., Mulder, M., & Van Paassen, M. M. (2009). Car lane-keeping using auditory feedback. *Delft University of Technology*, the Netherlands.

Vinje, E. W., & Pitkin, E. T. (1972). Human operator dynamics for aural compensatory tracking. *IEEE Transactions on Systems, Man and Cybernetics, 4*, 504–512.

Vinod, S., Simpson, R., LoPresti, E., & Schmeler, M. (2010). Evaluation of semiautonomous navigation assistance system for power wheelchairs with blindfolded nondisabled individuals. *Journal of Rehabilitation Research & Development, 47*(9), 877–890.

Wickens, C. D. (1992). *Engineering psychology and human performance.* Glenview, IL: Scott, Foreman, and Co.

Wolf, K. E., & Nees, M. A. (2015). Soundscapes for in-vehicle technologies. Workshop on In-Vehicle Auditory Interactions at the 21st International Conference on Auditory Display (ICAD), 1–5.

World Health Organisation (2015). *Global status report on road safety 2015.* Available from http://www.who.int/violence_injury_prevention/road_safety_status/2015/en

4.3 Blind driving by means of a steering-based predictor algorithm

Corresponding article: Bazilinskyy, P., Beaumont, C. J. A. M., Van der Geest, X. O. S., De Jonge, R. F., Van der Kroft, K., & De Winter, J. C. F. (2017). Blind driving by means of a steering-based predictor algorithm. Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics (AHFE), Los Angeles, USA.

4.3.1 Abstract

The aim of this work was to develop and empirically test different algorithms of a lane-keeping assistance system that supports drivers by means of a tone when the car is about to deviate from its lane. These auditory assistance systems were tested in a driving simulator with its screens shut down, so that the participants used auditory feedback only. Five participants drove with a previously published algorithm that predicted the future position of the car based on the current velocity vector, and three new algorithms that predicted the future position based on the momentary speed and steering angle. Results of a total of 5 hours of driving across participants showed that, with extensive practice and knowledge of the system, it is possible to drive on a track with sharp curves for 5 minutes without leaving the road. Future research should aim to improve the intuitiveness of the auditory feedback.

4.3.2 Introduction

Road traffic crashes are a major public health problem. If no action is undertaken, road crashes will become the seventh leading cause of death by 2030 (World Health Organization, 2015). About 95% of crashes are caused by driver error, in particular inattention and distraction (Dingus et al., 2016).

One way to avoid crashes caused by poor driver behavior is automated driving (Jamson, Merat, Carsten & Lai, 2011). An important technology, which may be seen as one of the first steps towards autonomous driving, is automated lane keeping—as used in modern Tesla and Mercedes Benz cars, for example. Current automated driving systems cannot predict the behavior of other road users in all situations. Once the system fails to handle a traffic situation, the driver needs to take over the steering and keep the car on the road. These transitory situations are a safety concern if the driver fails to reclaim control properly. The median estimate among the general public is that autonomous driving will be widespread by 2030 (Kyriakidis, Happee & De Winter, 2015), whereas some experts argue that autonomous driving will not be feasible before than 2075 (Shladover, 2016). Until driving is autonomous, automated driving systems will require driver intervention at certain times.

The aim of this research was to develop and perform a preliminary evaluation of a lane-keeping assistance system in which drivers are supported by auditory feedback as a function of the position on the road (Verbist, Boer, Mulder & Van Paassen, 2009). Auditory feedback may be beneficial for regaining steering control from automated driving, especially if the driver is visually overloaded. Auditory feedback may also be useful in regular manual driving

when visual information is temporarily lacking, such as when driving in heavy fog or during a visual distraction.

Previous research concurs that the use of real-time feedback can enhance driving performance. Such performance gains were demonstrated for example by Powell & Lumsden (2015), who provided tonal cues to racing drivers based on the lateral G-force, and by Houtenbos et al. (2016), who provided auditory beeps as a function of the speed and direction of another car approaching an intersection. Furthermore, lane departure and forward collision warning systems are already commercially available. In these systems, the driver is alerted by audio when a problem occurs.

We tested auditory feedback concepts while drivers did not receive any visual information regarding the road environment. This 'blind driving' paradigm can be regarded as the ultimate condition for testing human-machine interfaces: if drivers are able to steer a vehicle by means of sound only, this may provide evidence that the sound cues are effective if visual information is compromised.

This paper presents a design iteration of a previous concept of blind driving by means of auditory feedback (hereafter called 'Blind Driving 1'; BD1; Bazilinskyy et al., 2016). The focus in BD1 was on investigating how far into the future the system has to 'look' to determine the predicted position of the car—the prediction time. In BD1, auditory feedback was based on the predicted location of the car 0 s, 1 s, 2 s or 3 s into the future. When the predicted location of the car deviated more than 0.5 m from the center of the lane, audio feedback (i.e., a tone) was issued to alert the drive to correct their trajectory. The tone became louder the farther the predicted position was from the lane center. Results of this project showed that without predictor feedback (i.e., 0 s prediction), participants were more likely to depart the road compared to with predictor feedback. In this paper, the algorithm presented in BD1 is enhanced with the aim to improve the accuracy of the prediction path.

4.3.3 Method

4.3.3.1 Apparatus

A fixed-base driving simulator was used (Fig. 1; Green Dino, the Netherlands). An interface was programmed in MATLAB/Simulink r2015a to retrieve location, speed, and steering data from the simulator and to generate audio output via Sennheiser CX-200 headphones. When wearing the headphones, the participants were still able to hear engine and tire sounds via loudspeakers mounted in the simulator. Similar to BD1, the participants had to steer away from the sound: sound on the left was produced when the predicted lateral error was left of the center of the right lane, and vice versa. During the experiment, the screens of the simulator were turned off.

Figure 1. The driving simulator used in this research. In all trials, the screens were turned off and participants wore headphones.

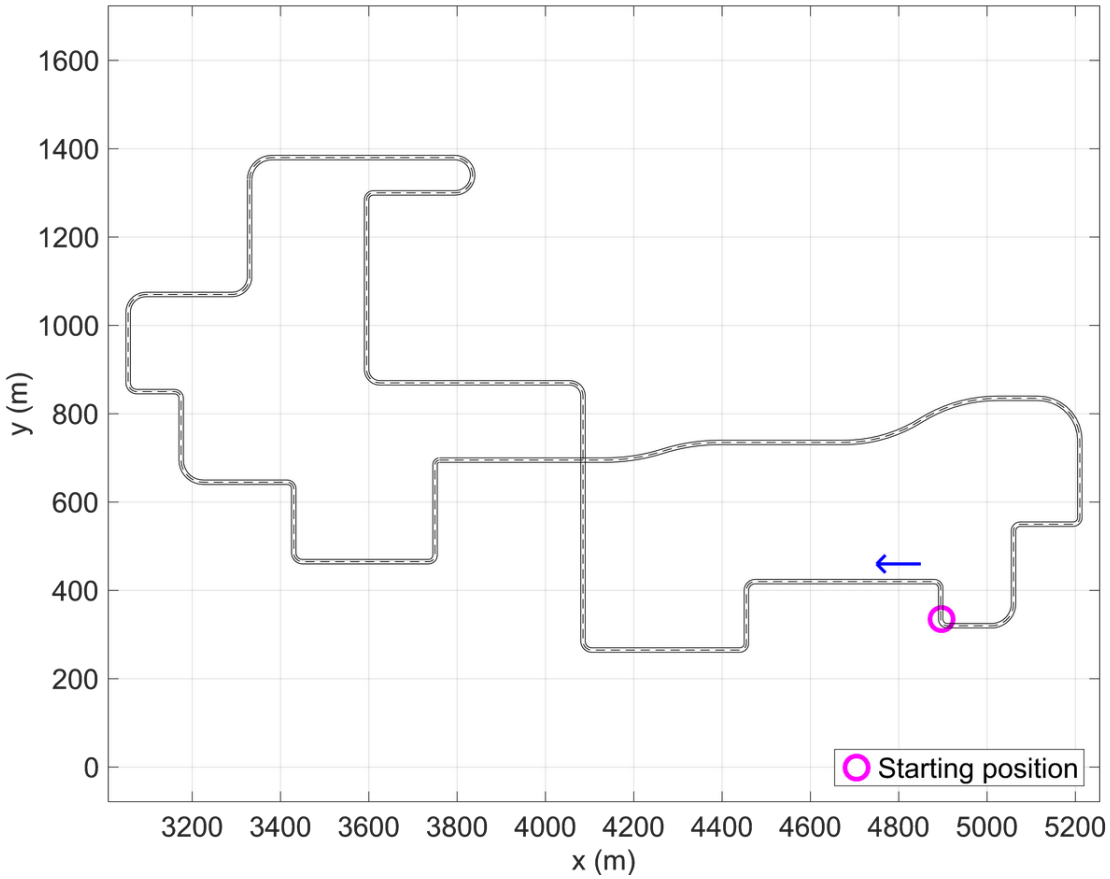


Figure 2. The test track.

4.3.3.2 Track

The track was a two-lane 7.5-km road without intersections and without other road users (Fig. 2). It contained straight segments and mostly 90-degree sharp curves with a radius of about 20 m (for research using the same track, see Bazilinskyy et al., 2016; De Groot, Centeno Ricote & De Winter, 2012; Van Leeuwen et al., 2015; Van Leeuwen, Happee & De Winter, 2014). The lane width was 5 m. The width of the simulated car was 1.76 m, and its length was 4.22 m. In each trial, participants drove 5 minutes. The driven distance per trial varied between 2,069 m and 4,206 m, depending on the number of times the car left the road and was reset on the center of the right lane with zero speed. If driving the full 4,206 m, participants encountered twelve 90-degree curves and one 180-degree curve.

4.3.3.3 Participants

In total, five males (mean age = 26.6 years, SD = 6.3 years) participated in the study. None of the participants had hearing impairments. Participants 1–4 (authors 2–5 of this paper) had been involved in the design, and therefore had detailed knowledge of the feedback concepts. Participant 5 was an expert racing driver who was new to the feedback designs and was not informed about their working mechanisms in any way. Participant 5 was invited in order to investigate how well a competent driver, who is naïve to the auditory systems, is able to keep the car on the road.

4.3.3.4 Speed and gearbox settings

An automatic gearbox was used. The speed of the car was predetermined; the participants did not use the pedals. The car automatically accelerated to a speed of around 80 km/h on straights, and decelerated to 20–40 km/h for the curves, depending on curve radius.

4.3.3.5 New algorithm for issuing feedback

In the BD1 concept, when driving through a curve, the predicted location of the car was mostly outside of the road boundaries because the prediction was based on the momentary velocity vector of the car. In the present study, the steering angle was used in the prediction, making it possible to create a more accurate prediction of the future position of the car (see also Godthelp, Milgram & Blaauw, 1984; Van Winsum, Brookhuis & De Waard, 2000). Figure 3 illustrates the predicted path in the BD1 and BD2 algorithms, both for a 2 s prediction.

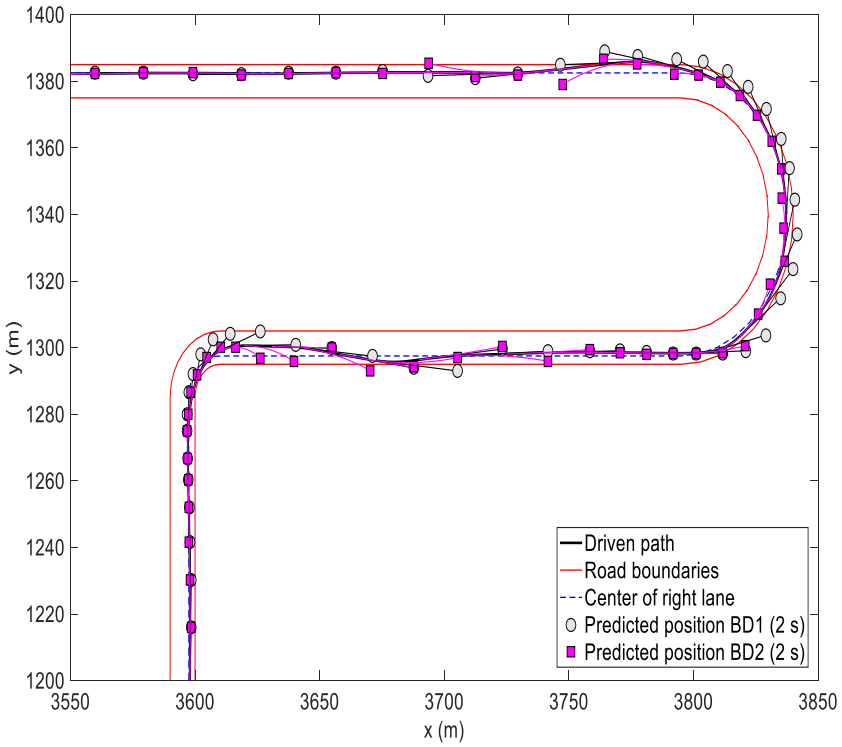


Figure 3. Working mechanism of the BD1 versus BD2 predictor feedback, both with a 2 s prediction. The figure shows the path driven by a participant through two curves. The circular markers represent the predicted position with the BD1 system (a straight line from the current location), whereas the square markers represent the predicted position with the BD2 system (a curved path from the current location). The markers are shown with 1 s intervals.

It was observed in preliminary tests that a shorter prediction time yielded better driving performance at high speeds and on straights, whereas a longer prediction time yielded better driving at lower speeds and in curves. Long prediction time on straights may lead to oscillatory steering behavior, because a small error is amplified by a long prediction path. A variable prediction time may solve these problems. This study included a condition with a prediction time that varied, from 3 s at 20 km/h (in curves) to 2 s at 80 km/h (on straights).

An overview of the tested concepts is provided in Table 1. In summary, there were two different kinds of feedback: feedback from iteration 1 (BD1), which linearly predicts the vehicle location, and feedback based on the algorithm presented in the current iteration (BD2), which takes steering into account.

In all concepts, volume feedback was provided when the predicted lateral position with respect to the center of the right lane exceeded 0.5 m. The larger the distance from the lane center, the louder the volume became. Further details about the pitch and volume are provided in (Bazilinskyy et al., 2016). The decision to select a 1 m wide tolerance zone was based on our earlier research in which the same threshold was used (Bazilinskyy et al., 2016; De Groot et al., 2011). De Groot et al. (2011) also indicated that off-target feedback (i.e., augmented feedback provided when deviating more than 0.5 m from the lane center) yielded better lane keeping performance than on-target feedback (i.e., augmented feedback provided when deviating less than 0.5 from the lane center).

Table 1. The four blind driving concepts that were tested by the five participants.

Feedback name	Prediction type
BD1 (2 s)	2 s prediction based on current car speed
BD2 (2 s)	2 s prediction based on current car speed and current steering angle
BD2 (3 s)	3 s prediction based on current car speed and current steering angle
BD2 (3 s)	Variable 2–3 s prediction (3 s when driving at 20 km/h, 2 s when driving at 80 km/h) based on current car speed and steering angle

4.3.3.6 Experiment design

Participants 1–4 tested each of the four algorithms three times, in counterbalanced order. Participant 5 tested each of the four algorithms two times, and also drove four times with visual information (twice with and twice without the BD2 variable algorithm), in counterbalanced order. In summary, each participant performed a total of 12 trials of 5 minutes each (i.e., 1 hour of driving per participant).

4.3.3.7 Dependent variables

Driving performance was assessed by means of the number of resets. A reset occurred when the car drove outside of the road boundaries with all its four corners. Secondly, the on-target percentage (OTP) was used as a measure of lane-keeping accuracy. OTP was defined as the percentage of time that the current absolute lateral position was less than 0.5 m. Data from 3 s prior to 10 s after each reset were excluded from the calculation of OTP.

4.3.4 Results

Figure 4 shows the number of resets that the five participants experienced with the four tested systems. The variable prediction time and the 3 s prediction time led to better performance than the other two algorithms. It is noteworthy that for the BD1 system, Participant 5 performed better than Participants 1–4. Participants 1–4 performed better with the BD2 systems (which they designed themselves) than with the BD1 system.

Most resets occurred around curves (see Fig. 5 for an illustration). The prediction time in curves for the variable model was mostly around 3 s (as there are few curves where the car reaches a speed higher than 20 km/h). This explains that the resets for the 3 s prediction time and variable prediction times are similar (Fig. 4). There was one trial without a single reset, for the BD2 system with a 3 s prediction time.

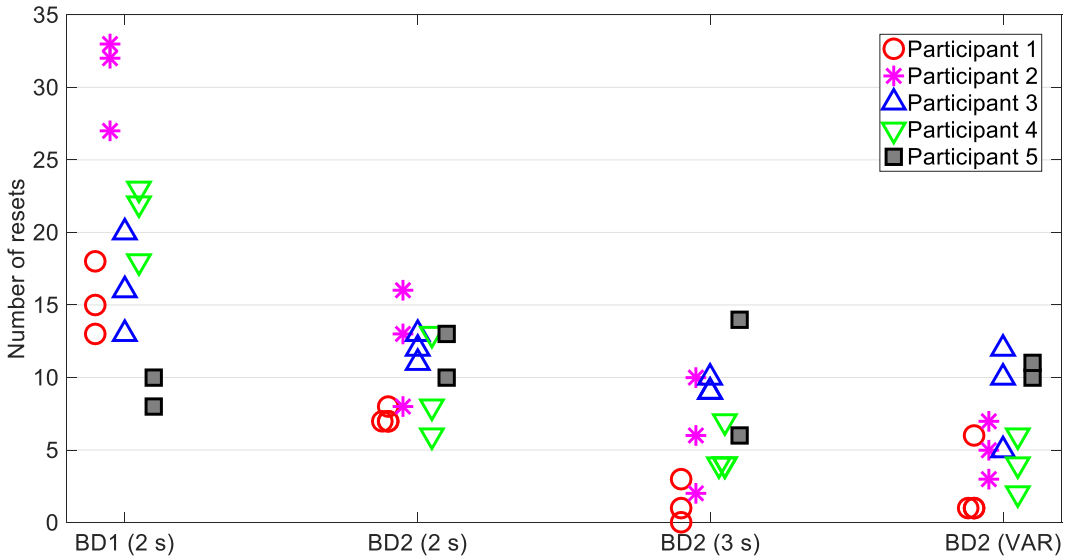


Figure 4. Numbers of resets.

The results for OTP (Fig. 6) mirror the results of the number of resets (Fig. 4), with the BD1 system yielding lower OTP values than the three BD2 systems, for Participants 1–4. It can also be seen that there were substantial individual differences, with some participants performing substantially better than others. There was no trial that had an OTP greater than 50%. As a reference, Participant 5 attained an OTP of 98% when driving with visual feedback in one of his trials.

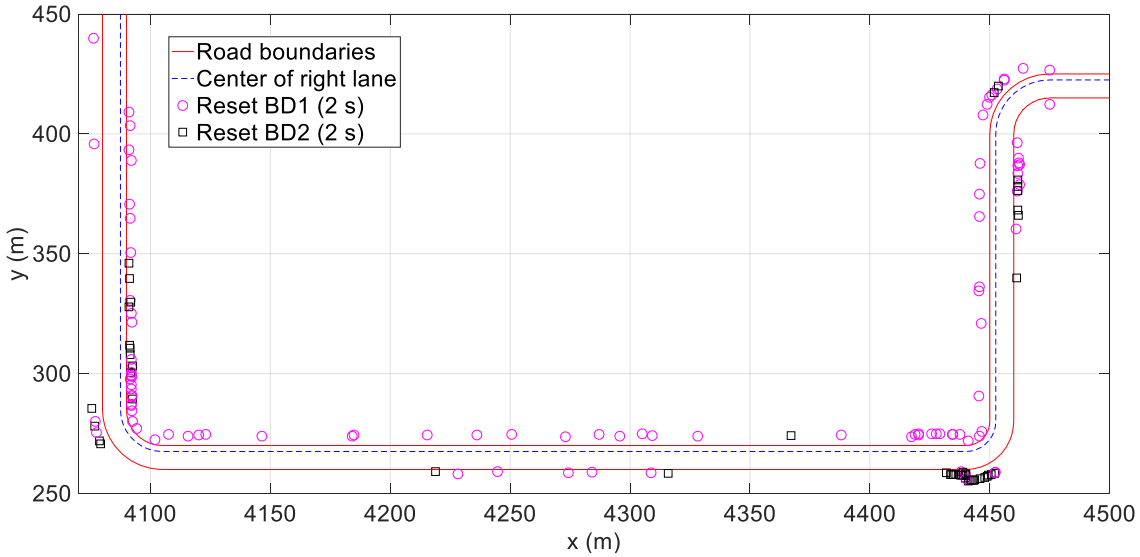


Figure 5. Locations of resets in part of the course for two of the four algorithms. It can be seen that participants crashed on distinct locations depending on the curve and the concept used.

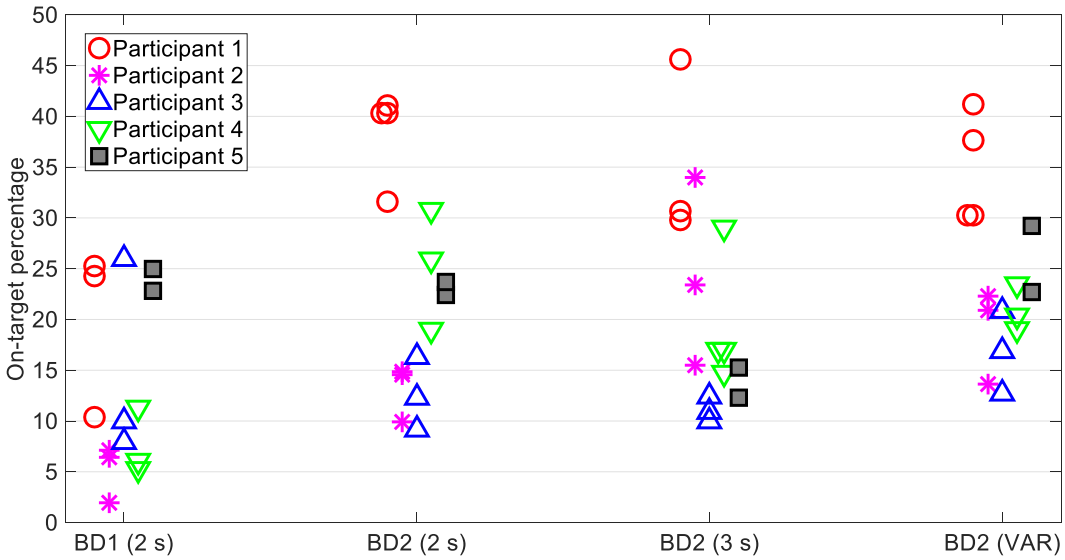


Figure 6. On-target percentage.

4.3.5 Discussion

In this design study, we implemented a predicted position based on momentary steering angle, for providing real-time auditory feedback to the driver.

The results of test drives with human subjects suggest that the proposed concept of using steering angle in the prediction of future position yielded a substantial improvement in driving performance as compared to a velocity-based predictor (Fig. 4), although the proposed concept still yielded resets during driving. The prediction time of 3 s featured one trial without a single reset. In other words, with sufficient practice and knowledge of the workings of the system, it is possible to drive a course for 5 minutes without leaving the road, purely based on auditory information.

The experiment was conducted with four young male engineering students as participants, who were also working on the project. Hence, these participants are not representative of the general population. Furthermore, because the sample size was small, we did not apply null hypothesis significance testing of any sort. Instead, the goal of this work was to examine whether it is possible to drive blindly by means of sound only.

Participant 5 drove without prior knowledge of the algorithms and performed better with the BD1 algorithm than Participants 1–4 did. Participant 5 commented afterwards that the auditory feedback was very hard to understand even after several trials of practice and after having driven with visual feedback and auditory feedback combined. The fact that the BD1 and BD2 concepts were not based on current lateral position, but on future lateral position, and the fact that BD1 and BD2 required different steering actions for a given audio input, may have been confusing. Participant 5 mentioned afterwards that he had not realized that the way the audio feedback has to be used is fundamentally different for these two algorithms. In BD1, no auditory feedback implies that the driver has to drive straight ahead; therefore, the steering wheel should be in the centered position. In the BD2 algorithms, feedback promotes a change of steering wheel position: no sound means the steering wheel is at the correct angle, and if there is audio feedback, the driver has to keep turning the steering wheel until the sound stops. These findings suggest that future research should

be directed towards the intuitiveness and stimulus-response compatibility of the auditory stimuli.

In the future studies we propose to investigate the effectiveness of multimodal feedback (e.g., a combination of vibrotactile and auditory feedback) in blind driving. A more realistic scenario in which participants have to control the speed of the car themselves may also yield insightful insights. Further development of techniques for issuing auditory feedback based on the type of road and type of curve may also be required. Additionally, future research should apply larger sample sizes as well as female participants. Finally, we point out that the idea of blind driving using headphones is not practical for real-life applications. Our experiment should be seen as a paradigm for investigating the value of auditory feedback under conditions where visual feedback is compromised. Future research could investigate spatial auditory feedback (e.g., via the car's speakers) in naturalistic conditions, such as driving when being visually distracted, when driving in rain or fog, or when visual information is otherwise unavailable or occluded.

4.3.6 References

Bazilinskyy, P., Geest, L. Van Der, Van Leeuwen, S., Numan, B., Pijnacker, J., & De Winter, J. C. F. (2016). Blind driving by means of auditory feedback. Proceedings of *13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems*, Kyoto, Japan.

De Groot, S., De Winter, J. C. F., López-García, J. M., Mulder, M., Wieringa, P. A. (2011). The Effect of Concurrent Bandwidth Feedback on Learning the Lane Keeping Task in a Driving Simulator. *Human Factors, 53*, 50–62.

De Groot, S., Centeno Ricote, F., De Winter, J. C. F. (2012). The effect of tire grip on learning driving skill and driving style: A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour, 15*(4), 413–426.

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. *PNAS, 113*, 2636–2641.

Godthelp, H., Milgram, P., Blaauw, G. J. (1984). The development of a time-related measure to describe driving strategy. *Human Factors, 26*(3), 257–268.

Houtenbos, M., De Winter, J. C. F., Hale, A. R., Wieringa, P. A., & Hagenzieker, M. P. (2016). Concurrent Audio-Visual Feedback for Supporting Drivers at Intersections: A Study Using Two Linked Driving Simulators. *Applied ergonomics*, *60*, 30–42.

Jamson, H., Merat, N., Carsten, O., & Lai, F. (2011). Fully-Automated Driving: The Road to Future Vehicles. Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, 2–9.

Kyriakidis, M., Happee, R., & De Winter, J. C. F. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour, 32*, 127–140.

Powell, N., & Lumsden, J. (2015). Exploring Novel Auditory Displays for Supporting Accelerated Skills Acquisition and Enhanced Performance in

Motorsport. Proceedings of *the 21st International Conference on Auditory Display*, 173–180.
Shladover, S. (2016). The truth about “self-driving” cars. *Scientific American*, *314*, 52–57.
Van Leeuwen, P. M., Happee, R., De Winter, J. C. F. (2014). Vertical field of view restriction in driver training: A simulator-based evaluation. *Transportation Research Part F: Traffic Psychology and Behaviour, 24*, 169–182.
Van Leeuwen, P. M., Gómez i Subils, C., Ramon Jimenez, A., Happee, R., De Winter, J. C. F. (2015). Effects of visual fidelity on curve negotiation, gaze behaviour, and simulator discomfort. *Ergonomics, 58*(8), 1347–1364.
Van Winsum, W., Brookhuis, K. A., De Waard, D. (2000). A Comparison of Different Ways to Approximate Time-to-Line Crossing (TLC) During Car Driving. *Accident Analysis & Prevention*, *32*(1), 47-56.
Verbist, K., Boer, E. R., Mulder, M., & Van Paassen, M. M. (2009). Car Lane-Keeping Using Auditory Feedback. Delft University of Technology, the Netherlands.
World Health Organization. (2015). Global Status Report on Road Safety 2015. World Health Organization, Geneva, Switzerland.

4.4 Blind driving by means of the predicted track angle error

Corresponding article: Bazilinskyy, P., Bijker, L., Dielissen, T., French, S., Mooijman, T., Peters, L., Happee, R., Dodou, D., & De Winter, J. C. F. (2018). Blind driving by means of the predicted track angle error. Under review.

4.4.1 Abstract

This study is the third iteration in a series of studies aimed to develop a system that allows driving blindfolded. We used a sonification approach, where the predicted angular error 2 seconds into the future was translated into spatialized beeping sounds. In a driving simulator experiment, we tested with 20 participants whether a directional surround-sound feedback system that uses four speakers yields better lane-keeping performance than binary directional feedback produced by two speakers. We also examined whether adding a corner support system to the binary system improves lane-keeping performance. Compared to the two previous iterations, this study presents a more realistic experimental setting, as drivers were unfamiliar with the feedback system and received the feedback without headphones. The results of the experiment show that drivers had poor lane-keeping performance and often left the road. Furthermore, the driving task was perceived as demanding, especially in the case of the additional corner support. Our findings from the blind driving projects suggest that drivers benefit from simple auditory feedback, and that additional auditory stimuli (e.g., corner support) add workload without improving performance.

4.4.2 Introduction

Driving is predominantly a visual task (Groeger, 2013). To drive safely, an accurate estimation of the position of the vehicle in relation to the road boundaries is indispensable (Land & Lee, 1994; Wann & Swapp, 2010). However, visual information from the environment may be compromised, for example in the case of darkness, fog, or rain (Edwards, 1998; Smith, 1982). Furthermore, studies have shown that the loss of one's visual field is a significant predictor of crash involvement, for older drivers in particular (Rubin et al., 2007). Moreover, even when visual information is present, drivers may not use it properly: it is estimated that 6% of fatal accidents are caused by driver distraction, including visual distraction (National Highway Traffic Safety Administration, 2018).

A substantial amount of research and development is happening in the domain of automated driving, both in industry and academia. Shladover (2016) argued that fully automated cars (i.e., SAE level 5 automation) will not be released to the public before 2075. Instead, it is likely that lower levels of automation will become available before the introduction of fully automated cars. In automation levels 3 and 4, the driver does not have to pay attention to the road, but it may sometimes be necessary to take back control, for example when hardware or software malfunctions or when the performance envelope of the automated driving system is exceeded. When taking back control, the driver needs to quickly establish awareness of the vehicle's position on the road. The use of sound could help in this process.

Auditory displays are effective for warning or supporting human operators, because sounds can be perceived regardless of the orientation of the eyes (Stanton & Edworthy, 1999). For example, Belz et al. (1999) found that auditory icons as collision warnings reduced brake response times in case of an imminent rear-end collision as compared to a dash-mounted visual display and no display. Similarly, in a driving simulator study where participants used an advanced traveller information display while driving, it was found that the use of an auditory-only or an audio-visual display yielded faster response times, more correct turns, and lower level of subjective workload than a visual-only display (Liu, 2001). Sound can also aid in the perception of speed and distance. For example, a study using videos of traffic scenarios found that participants who received a lower level of auditory feedback of the internal car noise chose a higher speed and were less accurate in estimating their speed (Horswill & McKenna, 1999). In Bazilinskyy, Van Haarlem et al. (2016) participants were able to estimate the distance to an object by means of artificial sounds; the mapping of distance to sounds is a process called 'sonification'.

Auditory feedback can also be applicable in situations when information needs to be transferred to a blind person. One example is the racing auditory display (RAD) by Smith and Nayar (2018), an audio-based user interface that allows blind players to play the same racing games as sighted players. The RAD used a spatialized soundscape to represent the driver's relative risk of hitting either edge of the track. Furthermore, a turn indicator system was used, which alerted drivers of the type of upcoming turn by means of spatialized sound. The results showed that participants subjectively appreciated the RAD concept and that they were able to drive competitive lap times. A real-world example of the use of auditory feedback for assisting blind individuals is a device used for para-biathlon, where blind athletes are guided where to shoot in a two-dimensional

space by sound (International Biathlon Union, 2017). Others have investigated the effectiveness of haptic feedback for visually impaired individuals while driving (Sucu & Folmer, 2014) or walking on a track (Rector, Bartlett & Mullan, 2018).

Summarizing, there is a need for assistive technologies that support the driver when it is not possible to use the available visual information (e.g., driver distraction), when there is not enough visual information (e.g., driving in fog), or when taking over vehicular control from the automation system. To gain knowledge on the feasibility of developing such a system, the extreme case is evaluated here, by eliminating all visual input.

This paper presents follow-up research aiming to improve the auditory feedback systems designed in Blind Driving 1 (BD1; Bazilinskyy, Van der Geest, et al., 2016) and Blind Driving 2 (BD2; Bazilinskyy, Beaumont, et al., 2017). Several findings stand out from these previous two studies, which are enumerated below:

1. In BD1, the authors found that a preview time of 2 seconds yielded the best lane keeping performance as compared to preview times of 0, 1, and 3 seconds; this finding corresponds to literature about preview in normal (i.e., non-blinded) driving and other tracking tasks (Land & Lee, 1994; Lehtonen, Lappi, Koirikivi & Summala, 2014).

2. In BD1, the feedback system was tested by two of the authors, with deep knowledge of the feedback system and the track. In BD2, the feedback system was tested by four of the authors/developers of the system, as well as by an expert racing driver. That is, six of the seven participants in the first two iterations were well acquainted with the rationale of the tested feedback systems. In the present study, novice participants were used instead, as a more realistic user sample.

3. In BD1 and BD2, headphones were used. In the present study, the feedback was provided without headphones. Accordingly, compared to BD1 and BD2, this study presents a more realistic experimental setting than that in the first two iterations.

4. In BD1 the future lateral position was based on the velocity vector only, whereas in BD2, the predictor algorithm of BD1 was improved by including the steering angle in the prediction. Here, we propose directional feedback based on the angular error of the vehicle with respect to the target track (i.e., the centre of the driving lane). We hypothesize that such feedback is more intuitive than that used in BD1 and BD2, as it represents the expected deviation of the car in the future instead of mere a velocity vector and/or steering angle.

5. In BD1 and BD2, sound was produced from the left side of the headphones when the predicted lateral error was left of the center of the right lane, and vice versa. Here, we tested whether a directional surround-sound feedback system that uses four speakers (hereafter referred to as the ‘beacon’) yields better lane-keeping performance than directional feedback presented with two speakers (called henceforth ‘binary feedback’). With the binary feedback, the sound came from speakers at either the right or the left side of the driver, which resembles the way that feedback was provided in the first two iterations, that is, from the left or right side of the headphones. The beacon feedback depicts the direction and the magnitude of the steering angle that must be applied to correct the trajectory of the car.

6. In BD1, the added value of two types of corner support (a beep when entering and leaving a corner vs. a beep when crossing the road centerline in a corner) to the feedback system was also tested. In BD1, corner support resulted in equivalent driving performance as compared to a reference condition without corner support. Herein, we re-examined whether adding a corner support system to the binary feedback (similar to the corner support tested in BD1) improves lane-keeping performance compared to the binary feedback system without corner support.

4.4.3 Method

4.4.3.1 Apparatus

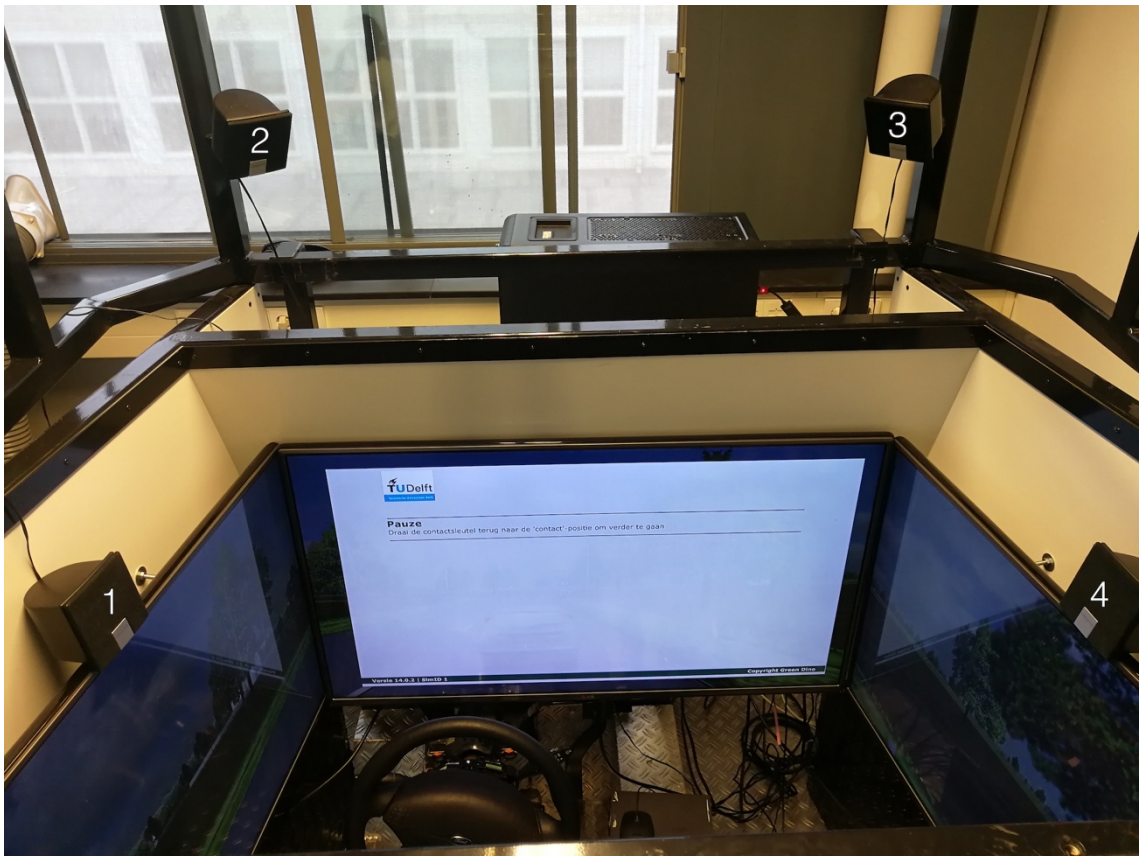


Figure 1. Fixed-base driving simulator 'Green Dino' used in the experiment.

For this research, a fixed-base driving simulator (Fig. 1, Green Dino, The Netherlands) was used (the same as in BD1 and BD2). An interface was programmed in MATLAB/Simulink r2016b to retrieve data from the simulator and to generate sounds from the 4.0 speaker system (Creative Inspire 4.1 4400 without the subwoofer) mounted in the driving simulator. The four speakers were placed in the simulator as shown in Fig. 1. The distance between the left and right speakers was about 1 m. The rear speakers were positioned at about the height of the ears of the human driver. The front speakers were positioned somewhat higher than the rear speakers (Fig. 1).

4.4.3.2 Track

The same track was used as in BD1 and BD2 (Fig. 2). This track was a two-lane 7.5-km road without intersections and without other road users. It contained

straight segments, 180-degree corners, and sharp 90-degree corners, most of which had a radius of about 20 m. The lane width was 5 m. Participants started at different points along the track. In each trial, the participant drove for 3 minutes. The speed of the car was controlled automatically. The speed in curves was 20 to 30 km/h, depending on the curve radius, and the speed on straights was about 70 km/h. Information about the current speed was not provided to the participants.

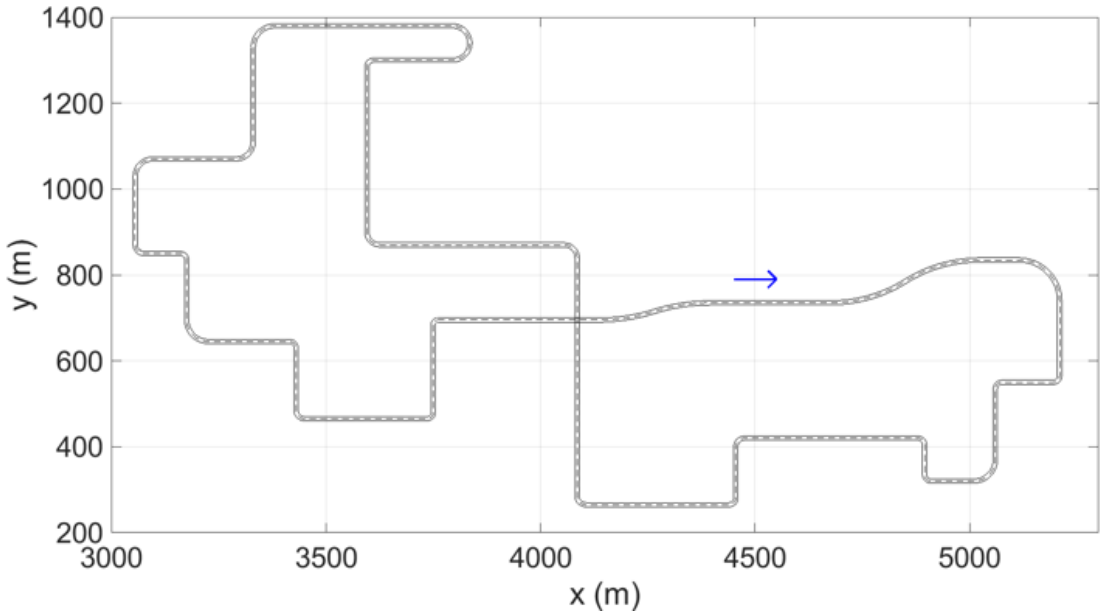


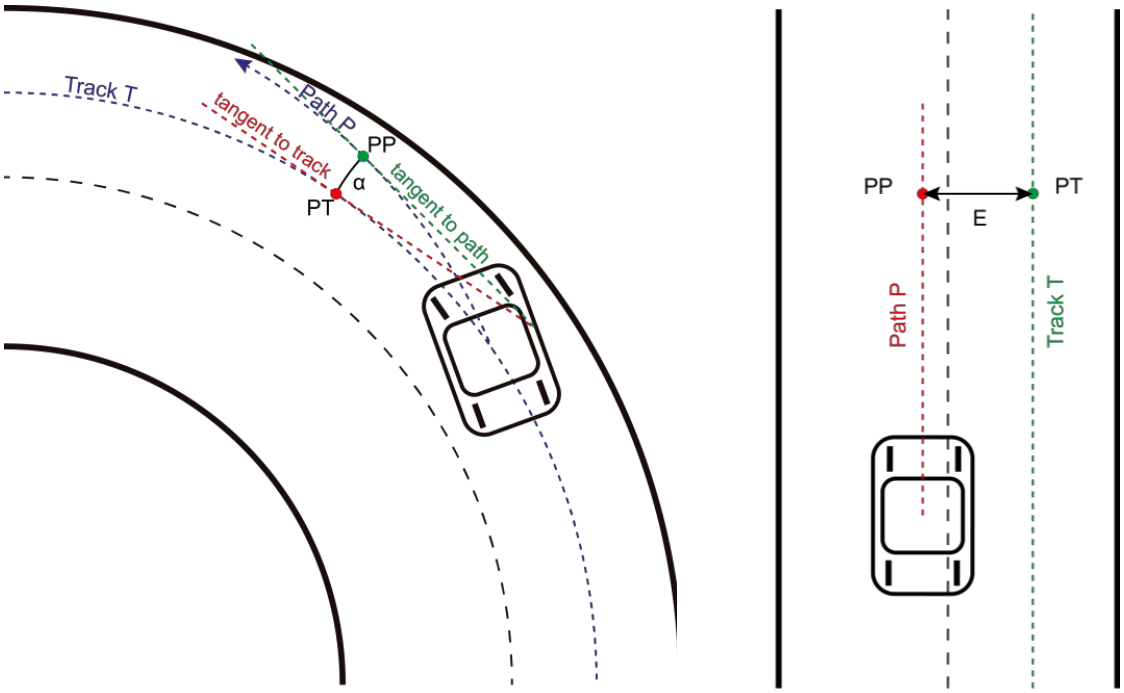
Figure 2. Top view of the test track. x and y are Cartesian coordinates in meters. The arrow indicates the driving direction.

4.4.3.3 Error prediction

A representation of the error-prediction algorithm is given in Fig. 3a. Using the steering angle and velocity, a prediction point (PP), representing the position of the car if it continues with the current constant steering input, is calculated 2 seconds into the future (as in BD2; Bazilinskyy, Beaumont, et al., 2017). A representation of the distance between the car and PP on the track (i.e., lane centre) determines a predicted point on the track (PT), the position of the car 2 seconds into the future if it remains on the track (i.e., the centre of the right line). Angle alpha is the angle between the tangent lines to PT and PP. The magnitude and sign of alpha are translated to auditory feedback.

A limitation of the proposed error prediction is the case shown in Fig. 3b, where angle alpha is undefined while the vehicle is not on the desired track. Accordingly, when alpha was between 3 degrees to the left and 3 degrees to the right of the track, the expected lateral error (E; i.e., the distance from PP to T) was used instead of the angle alpha. That is, when the angle was small, the algorithm switched from the sonification of angle information to the sonification of lateral position.

Figure 4 illustrates the principle of the prediction algorithm based on actual vehicle data. The figure presents the trajectory of the vehicle in a 180-degree turn. The magenta lines represent the predicted paths of the vehicle; these are identical to what would be generated by the predictor algorithm in BD2.



(a) Situation where alpha is defined

(b) Situation where alpha is undefined

Figure 3. Working principle of the error prediction algorithm. PP = Prediction point, PT = Prediction point on the track.

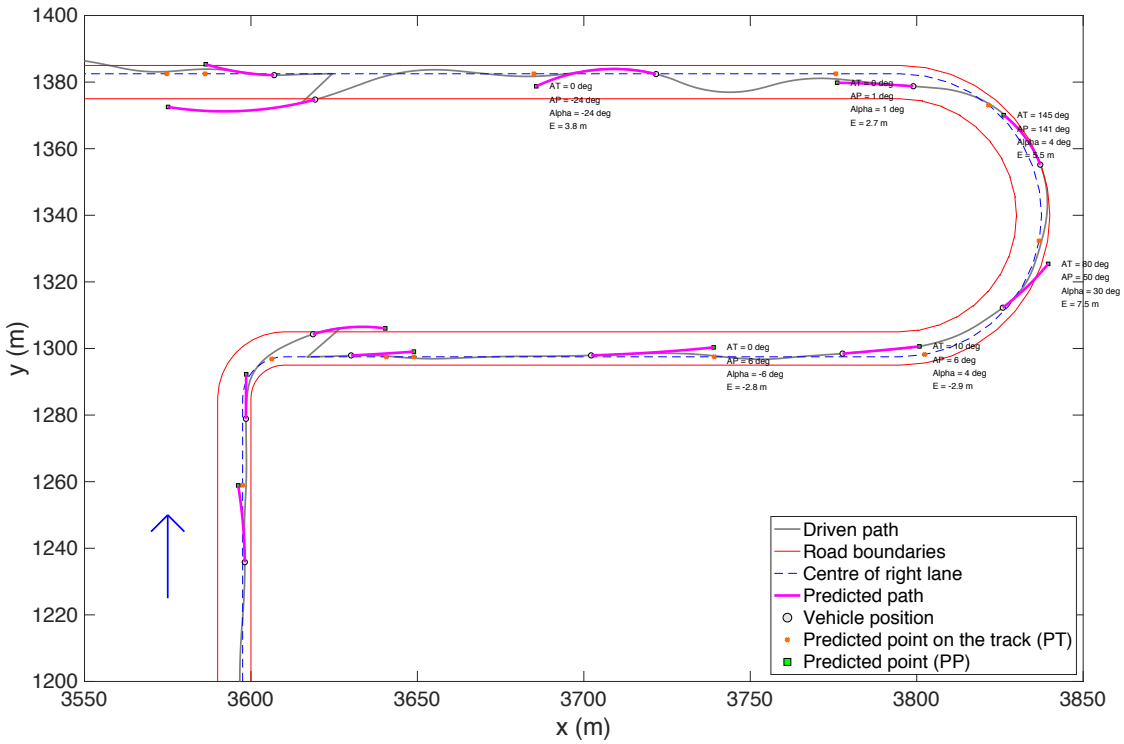


Figure 4. Working principle of the prediction algorithm in a 180-degree left turn. AT = Angle of the tangent line of the predicted point on the track (PT); AP = Angle of the tangent line of the predicted point (PP); E = Lateral error between PP and PT. The position of the vehicle is shown every 5 seconds.

The vehicle dynamics model represented a passenger car including realistic tire modelling. However, because the speed of the car was controlled automatically at a moderate speed, the nonlinear regions of the tires were never reached, so slip was never experienced.

4.4.3.4 Auditory feedback

We presented three types of auditory feedback (binary, beacon, and corner support). The auditory feedback was audible over engine sounds that were outputted by the driving simulator software.

1) With the **binary** feedback, the sound came from either the right rear speaker (if alpha was positive, i.e., greater than 3 degrees to the right of the track) or the left rear speaker (if alpha was negative, i.e., greater than 3 degrees to the left of the track). Feedback was also provided from the right or left rear speakers if alpha was smaller than 3 degrees while the lateral error E was larger than 0.5 m. The volume increased with increasing alpha, and the aim was to steer away from the sound to minimize the sound volume. No auditory feedback was provided when the participant was within an angle bandwidth of 3 degrees and a lateral position bandwidth of 0.5 m. The sound was beeps with a frequency of 464 Hz; the duration of each beep was 0.2 seconds, with an inter-beep time of 0.2 seconds.

2) With the **beacon** feedback, the predicted error was mapped to the four speakers using a division of volume between the speakers to mimic a shifting sound location in front of the driver. When angle alpha was between –20 and 20 degrees, the sound was linearly divided between the two front speakers. An alpha between 20 and 40 degrees was mapped between the two right speakers, and an alpha between –20 and –40 degrees was mapped between the two left speakers. An angle exceeding this bandwidth was represented by either the right rear speaker or the left rear speaker. As with the binary condition, the volume increased with increasing alpha. No feedback was provided if the participant was within an angular bandwidth of 3 degrees and a lateral position bandwidth of 0.5 m.

3) To clarify when a corner starts or ends, **corner support** was used *in addition* to binary feedback. At the start of a corner, loud beeps were generated from either the left rear speakers (for right curves) or the right rear speaker (for left curves). The beep was played once, twice, or three times, depending on how sharp the corner was. For a wide curve, with a required steering angle between 0 and 90 deg, the sound played three times. The tone played twice for a required steering wheel angle between 90 and 180 deg, and once for a required steering wheel angle between 180 and 270 deg, which corresponds to the sharpest type of curve. The beeps had a frequency of 928 Hz, a duration of 0.1 seconds, and inter-beep time of 0.2 seconds. When exiting the corner, the driver heard the same sound from both speakers for 0.9 seconds.

4.4.3.5 Participants

Twenty persons (15 males; 5 females), aged between 19 and 26 with a mean age of 22.9 years, participated in the experiment. All participants possessed a driver's license; the average amount of driving experience was 4.5 years. All participants provided written informed consent, and the research was approved by the Human Research Ethics Committee of the university.

4.4.3.6 Experiment design

Each participant drove three trials with a blindfold, receiving one type of auditory feedback per trial. Before each trial, a brief description was given about the particular system. In addition, before each of the three trials, participants drove 1 minute without blindfold to familiarise with the system. After each trial, the participants completed the NASA Task Load Index (TLX) questionnaire (Hart & Staveland, 1988). Half of the participants started with testing the binary feedback system, after which they tested the binary feedback system including corner support, and lastly the beacon feedback system (i.e., 1. binary, 2. binary+corner, 3. beacon). The other group tested the beacon feedback system first, followed by the binary feedback system and the binary feedback system with corner support (i.e.,1. beacon, 2. binary, 3. binary+corner). That is, the binary+corner condition was always preceded by the binary condition, because the corner support was combined with the binary feedback. After each trial, participants completed a technology-acceptance scale (Van der Laan et al., 1997). This was a five-point semantic-differential scale where participants were asked: "I find the auditory feedback in the last trial…" Participants had to select their response between two adjectives for nine items (1. useful–useless, 2. pleasant–unpleasant. 3. effective–superfluous, 4. irritating–likeable, 5. assisting–worthless, 6. undesirable–desirable, and 7. raising alertness–sleep-inducing). Note that two items (bad–good, nice–annoying) from the original scale were not used herein, due to a formatting error in the questionnaire.

4.4.3.7 Driving performance assessment

The number of resets and the on-target percentage (OTP) were used as performance measures, consistent with BD1 and BD2. A reset occurred when the car was entirely off the road, resulting in a restart in the middle of the lane. The on-target percentage was defined as the percentage of time the centre of the vehicle was within 0.5 m from the lane centre. Intervals from 3 seconds before each reset and 10 seconds after each reset were removed from the calculation of OTP. We used a repeated measures analysis of variance (ANOVA) and paired-samples *t* test to investigate whether the means of conditions significantly differed from each other.

4.4.4 Results

The results for the number of resets and the on-target percentage are shown in Figure 5. A repeated-measures ANOVA showed significant differences in the number of resets between the three conditions, $F(2, 38) = 4.27$, $p = 0.021$, $\eta_p = 0.18$. Post-hoc paired *t* tests showed that there was a statistically significant difference in the number of resets between the binary condition and the binary+corner condition ($t(19) = 3.37$, $p = 0.003$) as well as between the beacon condition and the binary+corner condition ($t(19) = 2.50$, $p = 0.022$). There were no statistically significant differences between the OTPs of the three conditions, $F(2, 38) = 0.85$, $p = 0.435$, $\eta_p = 0.04$. The number of resets was high, considering that there were only 3 minutes of driving per participant per condition. Because 3 seconds before each reset and 10 seconds after each reset were removed from the analysis, the OTP was calculated based on 40%, 42%, and 30% for the binary, beacon, and binary+corner conditions, respectively.

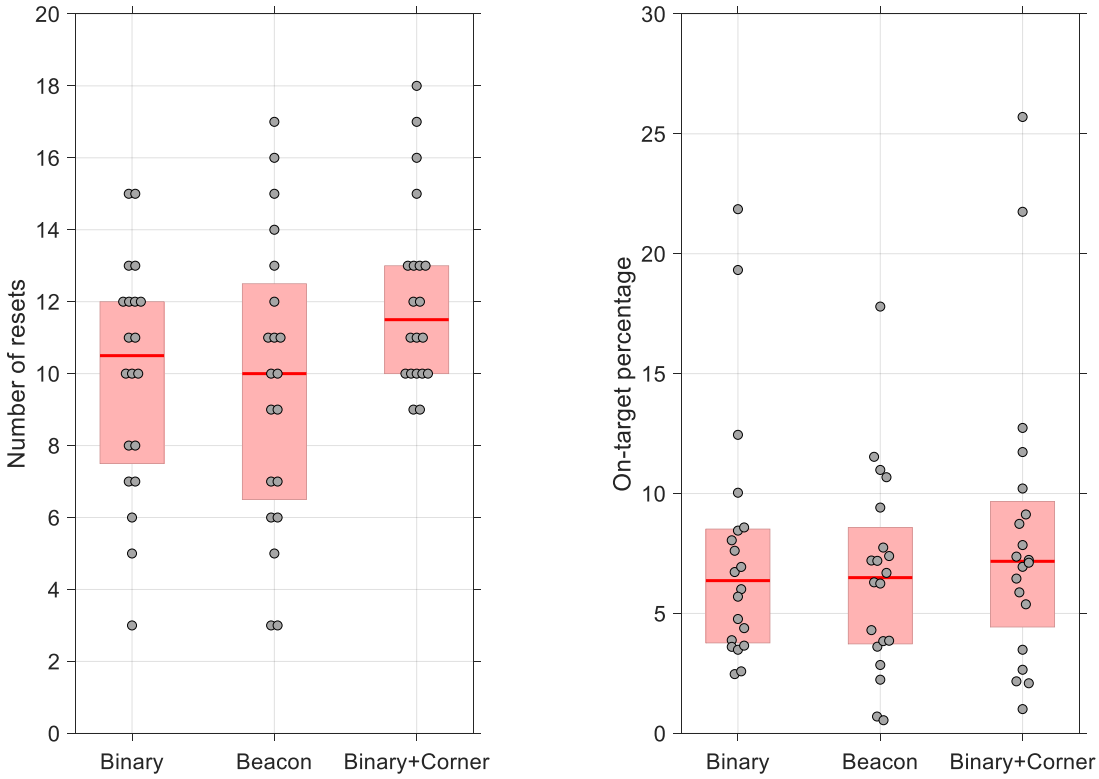


Figure 5. Boxplots of the number of resets (left) and on-target percentage (right). The boxplot shows the median, and 25th and 75th percentiles. Each marker represents one participant.

Figure 6 shows the locations of the resets on a selected part of the course. It is noticeable that more resets happen when the vehicle enters the corner.

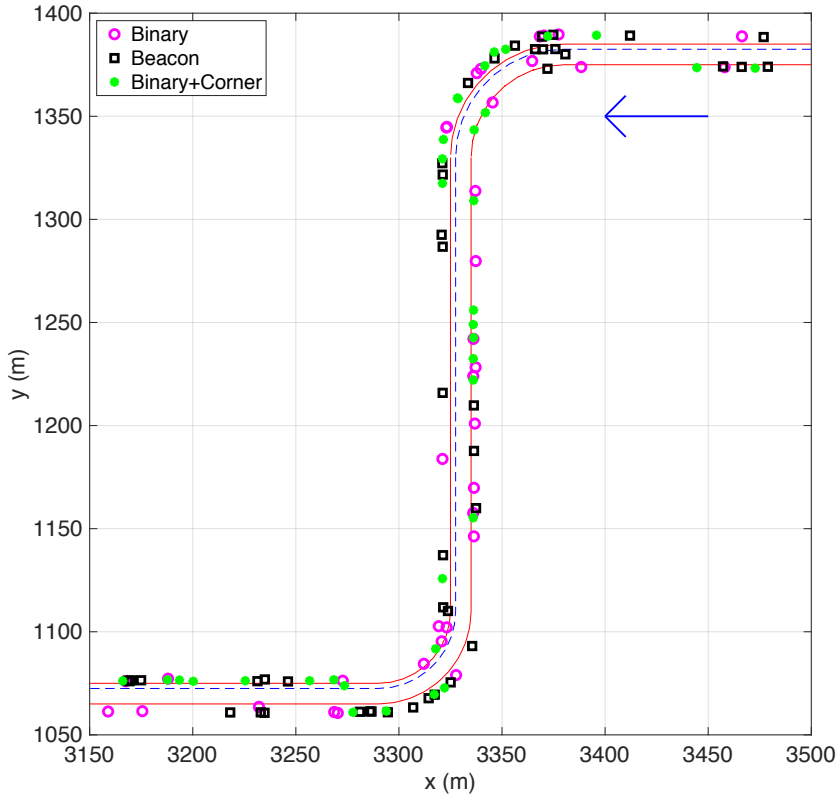


Figure 6. Locations of resets in a part of the course, for the three algorithms.

Figure 7 shows the results of the NASA-TLX. Overall, the scores were high (between 60% and 80%), with the exception of physical demand (with scores between 20% and 30%). These responses indicate that participants had difficulty with the driving task. Repeated measures ANOVAs indicated significant effects for mental demand ($F(2, 36) = 5.40$, $p = 0.009$, $\eta_p = 0.23$), physical demand ($F(2, 36) = 3.86$, $p = 0.030$, $\eta_p = 0.18$), performance ($F(2, 36) = 8.12$, $p = 0.001$, $\eta_p = 0.31$), effort ($F(2, 36) = 4.20$, $p = 0.023$, $\eta_p = 0.19$), and frustration ($F(2, 36) = 5.96$, $p = 0.006$, $\eta_p = 0.25$), but not for temporal demand ($F(2, 36) = 1.68$, $p = 0.200$, $\eta_p = 0.09$).

The highest workload ratings were observed for the binary+corner condition. For example, the performance item yielded higher ratings (i.e., more towards the 'failure' end) for the binary+corner condition than for both the binary condition ($t(19) = 3.36$, $p = 0.003$) and the beacon condition ($t(18) = 2.47$, $p = 0.024$). Similarly, the binary+corner condition yielded significantly higher ratings for the frustration item (i.e., more towards the 'very high' end) than both the binary condition ($t(19) = 3.21$, $p = 0.005$) and the beacon condition ($t(18) = 2.22$, $p = 0.039$).

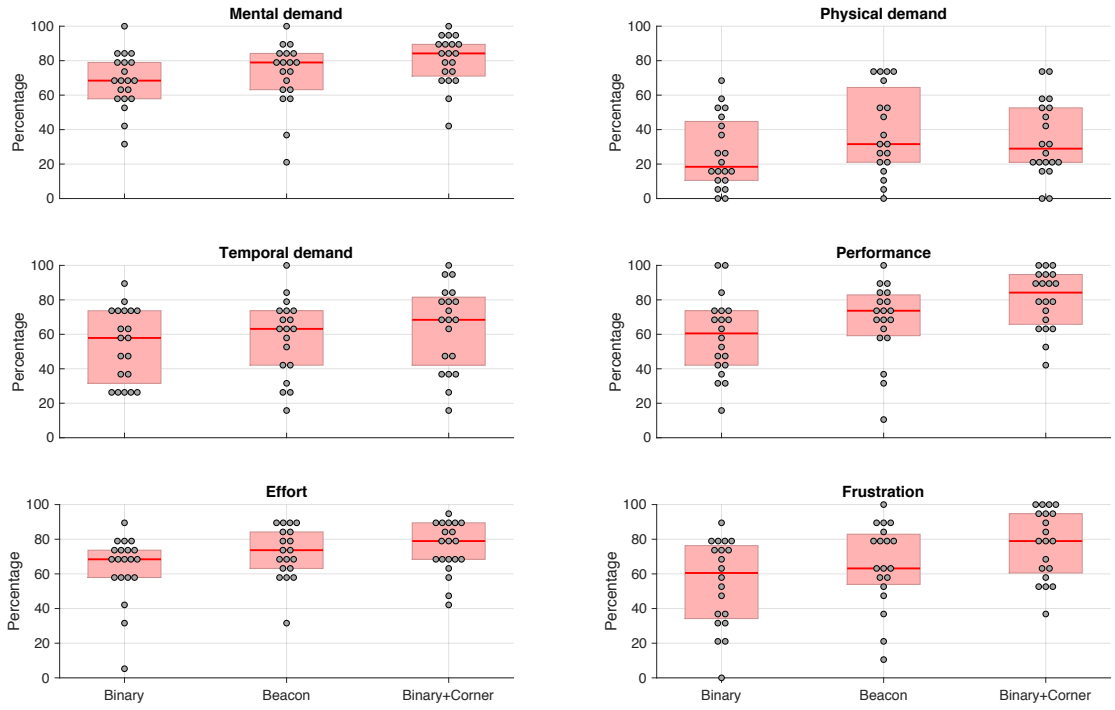


Figure 7. Boxplots with NASA Task Load Index (TLX) scores. Participants rated the items on a scale from very low (0%) to very high (100%), except for the performance item which was rated from perfect (0%) to failure (100%). Data for one participant were missing for the beacon condition (i.e., $n = 19$ instead of $n = 20$).

The results of the acceptance scale (Fig. 8) are consistent with the results of the NASA TLX. The binary and beacon systems were regarded as useful and effective. The binary+corner system was regarded as unpleasant and superfluous compared to the binary and beacon systems. However, the binary+corner system scored highly on 'raising alertness', which is probably because of the loud beeps provided upon entering and leaving each corner. All three systems were perceived as irritating, which might have been caused by the overall simplicity of the sounds and the specific frequency at which the sounds were delivered.

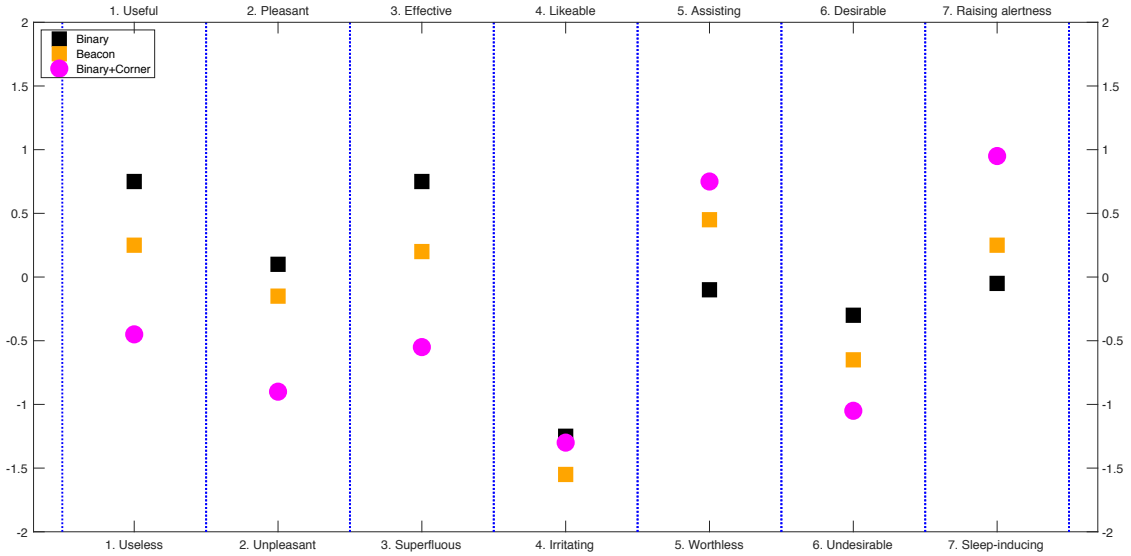


Figure 8. Mean scores on the acceptance scale.

4.4.5 Discussion

This study was aimed to use the benefits of auditory feedback to assist a driver when visual information is unavailable. We used a sonification approach, where the predicted angular error was translated into spatialized beeping sounds.

The results of the experiment in this study, performed with novice participants, yielded an overall poor driving performance (i.e., a high number of resets). The task was perceived as rather difficult by the participants, which is evident by the high self-reported workload. The high workload might have been caused by a lack of training: the participants in the present experiment were exposed to each feedback system only once. Since driving a car with visual feedback usually takes people months to master, it can also be expected that it will take some time to develop the skills to drive with auditory feedback only. In BD1 and BD2, measurements were performed with participants who had deep knowledge of the tested feedback systems and were exposed to the feedback for extensive periods of time prior to the experiment.

The binary+corner support feedback yielded OTPs that were similar to the OTPs of the beacon and the binary systems. This similarity may be due to the chosen experiment design, where the binary+corner condition was always tested *after* the binary condition, as a result of which the participants were not completely inexperienced anymore. However, the number of resets was significantly higher for the binary+corner condition compared to the other two conditions, which is consistent with the high self-reported workload of the participants for the binary+corner condition. It may be concluded that the augmenting the binary feedback system with corner support is confusing and distracting. Collectively, our findings suggest that drivers benefit from a simple system, such as the binary support, and that additional stimuli add workload without improving performance.

We recommend additional research to develop the auditory feedback further. The feedback in BD3 was implemented in such a way that when the angle was below 3 degrees, the lateral error determined the magnitude of the feedback. As soon as the driver corrected for this lateral error, the angle between the predicted path and track often became greater than 3 degrees, and the

participant was informed to steer in the opposite direction. This switching of feedback direction may have been confusing for the participants. An algorithm that combines the lateral error prediction with the angle between the predicted curve and track therefore deserves to be investigated (Griffiths & Gillespie, 2004).

As mentioned above, the feedback system was tested with novice participants. Considering that a feedback system could require some learning by the user, investigating the effect of experience on performance is a topic that deems further investigation.

In our method, we used an angle that was mapped to a directional beacon by distributing the volume across the four speakers. A proposed direction of further research is to examine whether the use of a head-related transfer function (HRTF) improves driving performance, as it provides a more realistic representation of the location of the sound source. Humans locate the direction of sound by using three sound characteristics: the relative volume of the sound, the phase difference between the sound waves in each ear, and the spatialization of the sound. The beacon feedback changes the relative volume between speakers. In the future, it will be interesting to investigate the effects sound change and phase in a blind driving paradigm.

Donges (1978) developed a now-classic manual control model of driving behaviour, consisting of an open-loop component (upcoming curves) and a compensatory component (correcting lateral and heading deviation errors). Developing a similar model of driving behaviour for auditory rather than visual feedback may be an interesting direction for future research.

4.4.6 Supplementary material

Example test drives of the three feedback systems:

- beacon: https://youtu.be/PyGILpMZ26U
- binary: https://youtu.be/T7kBxnoMQbU
- corner support: https://youtu.be/xFMQJa8nSmU

4.4.7 References

Bazilinskyy, P., Beaumont, C. J. A. M., Van der Geest, X. O. S., De Jonge, R. F., Van der Kroft, K., & De Winter, J. C. F. (2017). Blind driving by means of a steering-based predictor algorithm. Proceedings of *the 8th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Los Angeles, USA

Bazilinskyy, P., Geest, L. Van Der, Van Leeuwen, S., Numan, B., Pijnacker, J., & De Winter, J. C. F. (2016). Blind driving by means of auditory feedback. Proceedings of *13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems*, Kyoto, Japan.

Bazilinskyy, P., Van Haarlem, W., Quraishi, H., Berssenbrugge, C., Binda, J., & De Winter, J. C. F. (2016). Sonifying the location of an object: A comparison of three methods. Proceedings of *13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems*, Kyoto, Japan.

Belz, S. M., Robinson, G. S., & Casali, J. G. (1999). A new class of auditory warning signals for complex systems: Auditory icons. *Human Factors, 41*, 608–618.

Donges, E. (1978). A two-level model of driver steering behavior. *Human factors, 20*(6), 691–707.

Edwards, J. B. (1998). The relationship between road accident severity and recorded weather. *Journal of Safety Research, 29*, 249–262.

Griffiths, P., & Gillespie, R. B. (2004). Shared control between human and machine: Haptic display of automation during manual control of vehicle heading. Proceedings of *International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 358–366.

Groeger, J. A. (2013). Understanding driving: Applying cognitive psychology to a complex everyday task. Routledge.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183.

Horswill, M. S., & McKenna, F. P. (1999). The development, validation, and application of a video-based technique for measuring an everyday risk-taking behavior: Drivers' speed choice. *Journal of Applied Psychology*, 84, 977.

International Biathlon Union. (2017). Guided by sound: Blind biathlon shooting. Available at http://www.biathlonworld.com/news/detail/guided-by-sound-blind-biathlon-shooting

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature, 369*(6483), 742.

Lehtonen, E., Lappi, O., Koirikivi, I., & Summala, H. (2014). Effect of driving experience on anticipatory look-ahead fixations in real curve driving. *Accident Analysis & Prevention*, 70, 195–208.

Liu, Y.-C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics, 44*, 425–442.

National Highway Traffic Safety Administration (2018). *Distracted driving 2016* (Report No. DOT HS 812 517). Available at https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812517

Rector, K., Bartlett, R., & Mullan, S. (2018, October). Exploring aural and haptic feedback for visually impaired people on a track: A Wizard of Oz study. Proceedings of *the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, 295–306.

Rubin, G. S., Ng, E. S. W., Bandeen-Roche, K., Keyl, P. M., Freeman, E. E., & West, S. K. (2007). A prospective, population-based study of the role of visual impairment in motor vehicle crashes among older drivers: the SEE study. *Investigative Ophthalmology & Visual Science*, 48, 1483–1491.

Shladover, S. (2016). The truth about "self-driving" cars. *Scientific American*, *314*, 52–57.

Smith, B. A., & Nayar, S. K. (2018). The RAD: making racing games equivalently accessible to people who are blind. Proceedings of *the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC, Canada.

Smith, K. (1982). How seasonal and weather conditions influence road accidents in Glasgow. *Scottish Geographical Magazine*, 98, 103–114.

Stanton, N. A., & Edworthy, J. (1999). *Human factors in auditory warnings*. Ashgate.

Sucu, B., & Folmer, E. (2014, October). The blind driver challenge: steering

using haptic cues. Proceedings of *the 16th international ACM SIGACCESS conference on Computers & accessibility*, 3–10.

Van der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, 5, 1–10.

Wann, J. P., & Swapp, D. K. (2001). Where do we look when we steer and does it matter? *Journal of Vision, 1*(3), 185-185.

5 CONTINUOUS AUDITORY FEEDBACK FOR DISPLAYING AUTOMATION STATUS, LANE DEVIATION, AND HEADWAY

Corresponding article: Bazilinskyy, P., Larsson, P., Johansson, E., & De Winter, J. C. F. (2018). Continuous auditory feedback for the status of ACC, lane deviation, and time headway. Under review.

5.1 Abstract

Trucks that are equipped with driver assistance systems, such as adaptive cruise control (ACC), are emerging on the roads. Typically, these driver assistance systems offer binary auditory warnings or notifications upon lane departure, close headway, or automation (de)activation. Such binary sounds may annoy the driver if frequently presented. Truck drivers are well accustomed to the sound of the engine and wind in the cabin. Based on the premise that continuous sounds are more natural than binary warnings, we propose continuous auditory feedback on the status of ACC, lane offset, and headway, which blends with the engine and wind sounds that are already present in the cabin. An on-road study with 23 truck drivers was performed, where participants were presented with the additional sounds both in isolation from each other and in combination. Results showed that the sounds were easy to understand and that the lane-offset sound was regarded as somewhat useful. Systems with feedback on the status of ACC and headway were seen as not useful. Participants overall preferred a silent cabin and expressed displeasure with the idea of being presented with extra sounds on a continuous basis. Suggestions are provided for designing less intrusive continuous auditory feedback.

5.2 Introduction

5.2.1 Auditory interfaces for trucks

Trucks are increasingly deployed with advanced driver systems, such as adaptive cruise control (ACC). Such trucks typically provide binary auditory warnings based upon ACC deactivation, lane departure, and close headway (forward collision warning). Auditory signals are attractive as warnings because they are perceivable regardless of the driver's direction of visual attention (Stanton, 1994). People have a tendency to perceive a sequence of sounds as more than one auditory stream, each arising from a distinct source, also called 'stream segregation' (Bregman & Campbell, 1971), a phenomenon which may be useful for transmitting multiple types of warnings simultaneously.

The threshold settings of an auditory warning system must strike a balance between early detection of critical events and the avoidance of false alarms; false alarms are problematic, because they are annoying to the driver,

as a result of which the driver may disengage the warning system (Kidd, Cicchino, Reagan, & Kerfoot, 2017; Parasuraman, Hancock, & Olofinboba, 1997). To avoid annoyance, some car manufacturers have implemented visual warnings instead. For example, the status of the automation in the passenger car Volvo XC90 is shown by means of a green icon on the dashboard. When the automation has no clear picture of the environment, the icon becomes grey, and no auditory warning is provided. The reason for having no auditory warnings in such systems is that frequent auditory warnings are perceived as annoying. On the other hand, auditory warnings are typically used as imminent warnings, for example as the final stage of a two-stage or graded warning system (e.g., Bazilinskyy et al., 2018; Campbell et al., 2016)

5.2.2 Continuous auditory feedback

It has been argued that human interaction with the world (e.g., maintaining balance, applying forces, steering, aiming) is essentially continuous (Rath & Rocchesso, 2005). Although discrete triggers do occur in traffic (e.g., another road user suddenly appearing in sight), stimuli in normal driving (car following, lane-keeping) are of continuous nature. Furthermore, as stipulated by Newton's second law of motion, the physical movements of road users are necessarily continuous as well; road users cannot change their position, speed, or heading instantaneously. Therefore, continuous feedback may be perceived as more natural than discrete warnings.

Both continuous and binary warning sounds can be spatialized, giving information about the location of the source of the sound. For example, spatialized sound can be beneficial for providing information about surrounding traffic (Chen, Qvint, & Jarlengrip, 2007). However, the performance of spatialized sound is constrained by the resolution of the human auditory system (Crispien, Fellbaum, Savidis, & Stephanidis, 1996). The maximum capability to distinguish the origin of a sound in the frontal position is about 28 deg. The minimum resolution for lateral sources is about 108 deg, while a resolution for sounds coming from the back is about 68 deg. (Blauert, 1985). Chen et al. (2007) stated that a number of participants expressed disbelief in the feasibility of using spatialized auditory feedback on the road, despite ranking these stimuli highly during a driving simulator experiment on driver traffic awareness in trucks.

5.2.3 Aim of the paper

There is a need for designing concepts of auditory feedback that yield high acceptance ratings among drivers. It is postulated that truck drivers are sensitive to how their truck sounds like, and that they sometimes rely on engine noise to infer the state of the vehicle. Accordingly, we aimed to investigate a non-annoying functionality that provides continuous feedback, by creating a sound that resembles, and blends with, the natural engine and speed-dependent wind noise inside the truck cabin. By means of an on-road experiment, we tested whether continuous auditory feedback is a possible alternative to standard auditory warnings used in modern production trucks equipped with low-level automation. We hypothesized that continuous feedback about the system status, headway to the vehicle in front, and deviation of the vehicle from the centre of the lane will receive high acceptance ratings of drivers. The results of

the study are intended to be transferable to trucks with higher levels of automation.

5.3 Method

5.3.1 Participants

Twenty-three participants (18 male, 5 female) holding a truck driver's license participated in the experiment. The participants were employees of Volvo Trucks and were between 38 and 65 years old (M = 49.5 years; SD = 6.5 years). Their mean number of years of having a truck driver's license was 20.8 years (SD = 11.5). Thirteen of the participants reported a mileage of 1–1000 km, 7 participants reported 1,001–5,000 km, 2 participants reported 5,001–15,000 km, and 1 participants reported 15,001–20,000 km of driving in a truck in the past 12 months. One participant reported suffering from a hearing impairment (sensitivity to background sounds). All participants provided written informed consent, and the research complied with the American Psychological Association Code of Ethics.

5.3.2 Apparatus

The experiment took place on the E6 highway in Gothenburg, Sweden. An FH460 Volvo truck was used. The standard sound setup of the truck was used, where: 2 speakers are located in front, 2 speakers are in the doors, and 2 speakers are in the back. Spatialisation can be provided by outputting auditory feedback in the front/back or the right/left of the cabin of the truck.

5.3.3 Continuous auditory feedback

Three types of continuous auditory feedback were evaluated: 1) feedback based on the state (on vs. off) of the ACC (ACC-status sound), 2) feedback based on the deviation of the truck from the lane centre (lane-offset sound), 3) feedback based on the headway time to the vehicle in front (headway sound). All feedback was developed in Pure Data, a visual programming language for multimedia works.

Continuous auditory feedback on the state of ACC informed the driver whether the system was on or off. If ACC was on, feedback is generated. When ACC is off, no sound is produced. The sound was created by a white noise generator fed through a second-order bandpass filter. The bandwidth of the filter was 102 Hz. The centre frequency of the filter was adjusted based on the current speed of the truck as follows: speed x 30.05 +1054.6 (speed in km/h). The sound was designed to mimic the speed-dependent sound of the wind. The file 'ACC-status sound.mp3' in the supplementary material gives an example of this feedback when the speed of the truck is increased gradually from 0 km/h to 100 km/h.

Continuous auditory feedback on the deviation of the truck from the lane centre is a form of spatial auditory feedback where the driver is informed about the distance from the centre of the truck to the right or left edge of the lane. That is, if the truck was deviating right from the centre of the lane, the sound would appear from the right side, and vice versa. This sound level was linearly based on the deviation of the centre of the truck. The sound was generated with a cosine wave oscillator with speed as input for frequency multiplied by a

coefficient of 1.732. File 'Lane-offset sound.mp3' in the supplementary material gives an example of such feedback, when the location of the truck is gradually increased from left to right.

Continuous auditory feedback on the headway time to the vehicle in front informed the driver about the time headway to the vehicle in front. Feedback was given when the headway was smaller than 3.5 seconds. Similar to the feedback on the status of ACC, the base sound mimics the sound of wind. Similarly to the ACC status sound, it was created by a noise generator filtered by a 102 Hz wide, second-order bandpass filter, whose centre frequency depended on the truck's speed. The level of the sound was varied with the time headway to the lead vehicle so that the shorter the time headway, the louder the sound would be. File 'Headway sound.mp3' in supplementary material gives an example of such feedback, where THW is gradually decreased from 3.5 seconds to 0 seconds; the speed was set at 80 km/h.

5.3.4 Scenario

The participants drove four trials. They started in the garage of Volvo Group Trucks in Lindholmen, Gothenburg, Sweden. Then they drove towards St1 gas station near Kungälv on E6, see Figure 1.

(1) During the first trial, participants experienced the standard ACC and lane departure warning system available in the truck. The ACC was activated and deactivated a few times voluntarily by the participant. Upon de(activation) of the ACC, a standard sound was produced. After the first trial the participants stopped near a gas station Preem near Tuve.
(2) During the second trial, the ACC-status sound was played. After the second trial, the participants stopped at the St1 gas station near Kungälv. During the second trial, the ACC was active and the lane departure warning system was disabled.
(3) During the second trial, the ACC-status sound was played. After the second trial, the participants stopped at the St1 gas station near Kungälv. During the second trial, the ACC was active and the lane departure warning system was disabled.
(4) During the fourth trial, all three types of auditory feedback were played at the same time. The standard lane departure warning system was disabled whereas the ACC was enabled in about half of the trial.

The participants were asked to complete an introductory questionnaire before the start of the first trial. The questionnaire included questions on demographics, driving behaviour, and opinion on the types of auditory feedback that would be offered in trials 2–4 (i.e., prior to being exposed to the feedback). At different moments during the trials, an unstructured verbal interview on the sound systems was conducted with the driver. Participants were asked to give their general impression on the feedback, how the feedback could be improved, and whether they would like it in a future model of the truck. At the gas stations after trials 2–4, participants completed a questionnaire regarding the auditory feedback experienced in the preceding trial. The questionnaire asked whether the feedback was easy to hear and included a transport telematics acceptance

scale (Van der Laan et al., 1997) as well as the System Usability Scale (Brooke, 1986). The questionnaires can be found in the supplementary material.

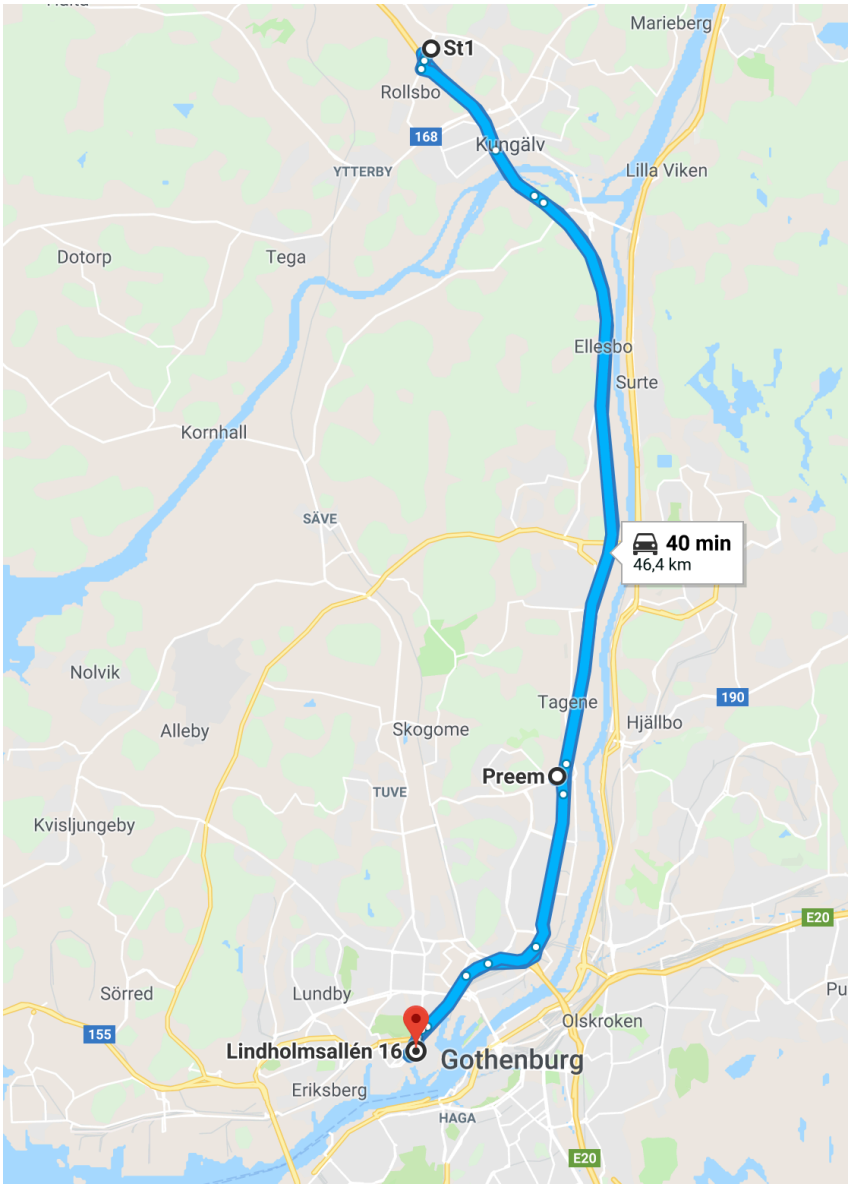


Figure 1. Route travelled during the experiment. Participants started at the Volvo Trucks garage in Lindholmen, drove towards Preem gas station for a stop after Trial 1, then drove towards St1 gas station in Kungälv for a stop after Trial 2. Trials 3 and 4 were conducted on the way back to the garage with a stop at the Preem gas station after Trial 3.

Figure 2 illustrates the driving speed of one of the participants. In Phase 2 (9.5–19.5 km) and in the last part of Phase 4 (38–40 km), the ACC was active as can be seen from the constant speed. The breaks between sessions can be distinguished by speeds of 0 km/h.

Figure 3 shows the lateral position for the same participant as shown in Figure 2. The lane width was about 3.5 m. Considering that the width of the truck is about 2.5 m, an absolute lateral position of 0.5 m or greater corresponds to driving on the lane markers.

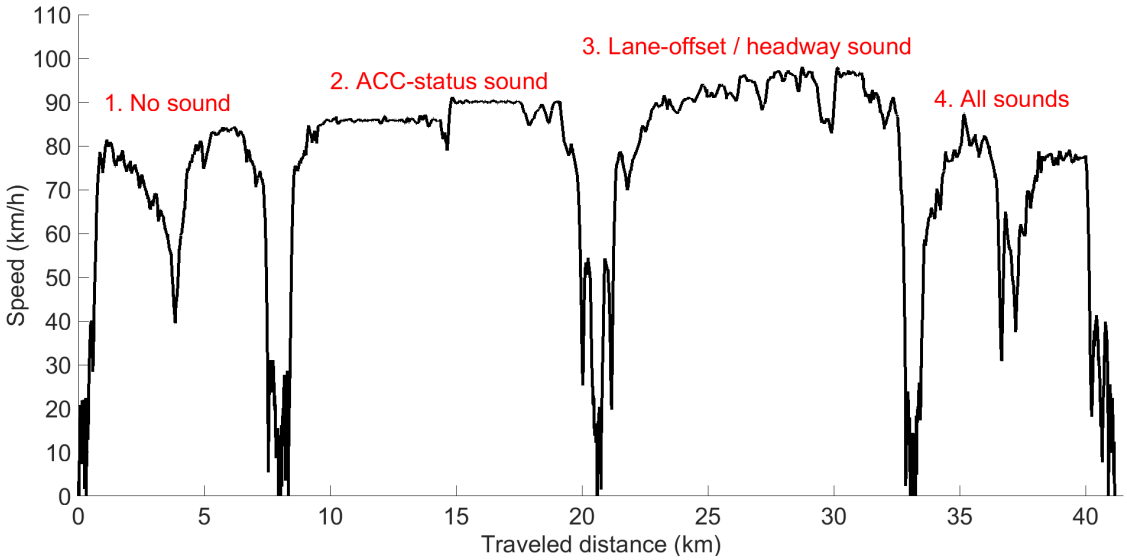


Figure 2. Driving speed for a selected participant.

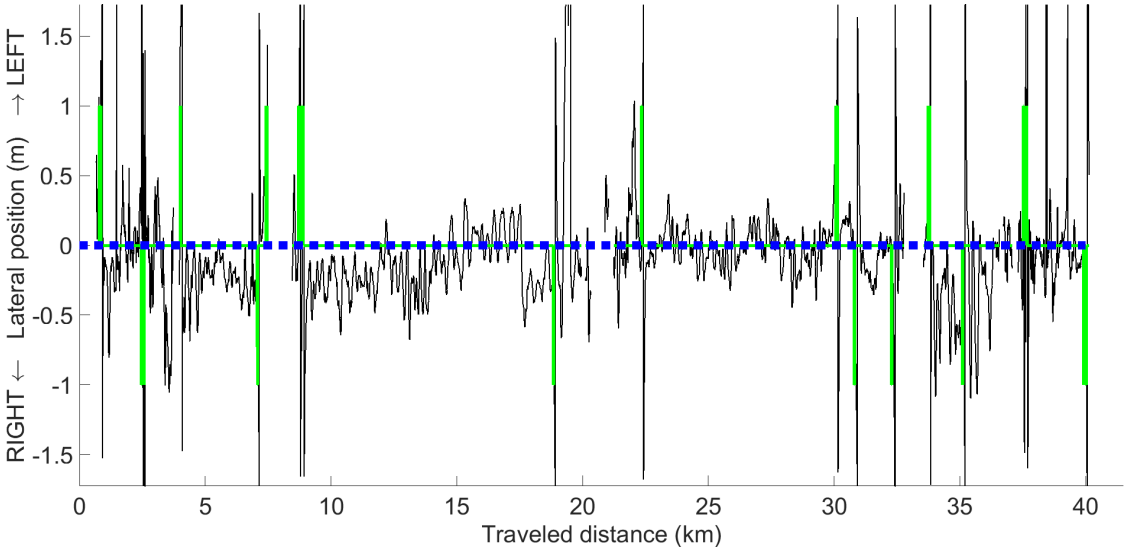


Figure 3. Lateral position (black) and the use of the turn indicator (green; 1 = left, –1 = right) for the same participant as in Figure 2. Data are shown only when the driving speed was greater than 50 km/h.

Figure 4 shows the sound volume during a trial. Sound was produced when the ACC was active (Trial 2: 8.5–19 km), except for four brief moments where the participant disengaged the ACC or the experimenter disabled the sound feedback. When lane-offset sound was produced (Trial 3: 21–27.5 km), the sound was not equal from the left and right speakers. A lane change to the left can be distinguished around 22.5 km. In Trial 3, between 28.5 and 32.5 km, the participant experienced headway sound. Finally, in Trial 4 (from 33.5 km onward), the participants experienced all sounds together.

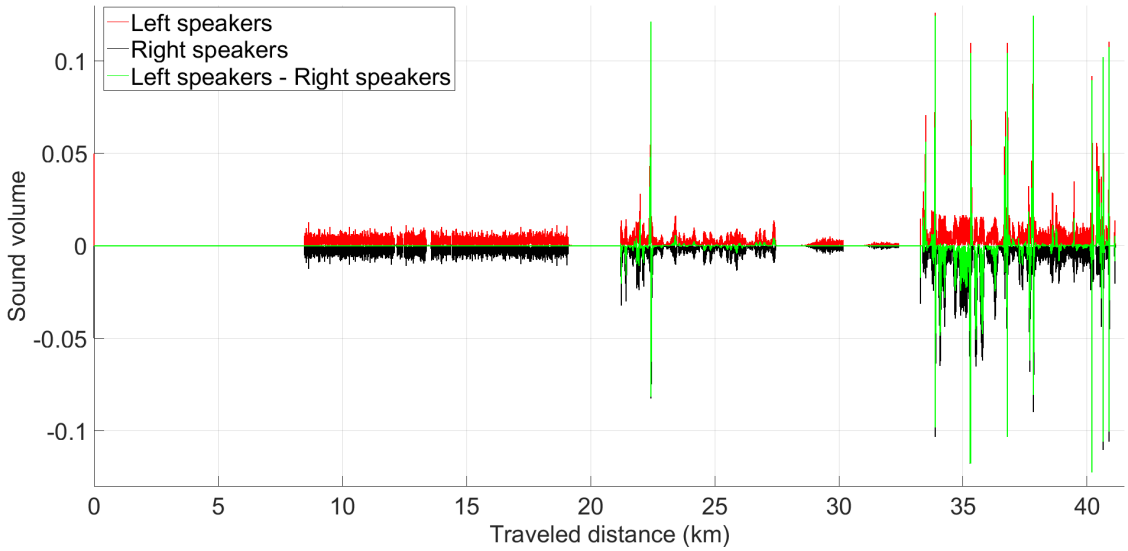


Figure 4. Sound produced during a trial for the same participant as in Figure 2. The signal 'left speakers – right speakers' indicates the difference in volume between the left and right speakers, that is, whether the sound was dominant on the left side (positive values) or on the right side (negative values).

5.4 Results

5.4.1 Interview responses and responses to open-ended questions

When responding to questions in the verbal interviews, participants expressed that they did not appreciate the idea of having additional sounds in the cabin. They indicated that significant research funds are actually directed to the reduction of noise in the cabin. During the second trial, multiple participants said that they would rather have auditory feedback when the ACC is turned off, instead of having it when the system is on. The idea of adding auditory feedback when no action was needed was not well accepted. On multiple occasions it was stated that a driver should be able to change the type of sound, its volume, and frequency to have personalized feedback.

In each post-trial questionnaire, the participants were given an open question '*What did you think of the feedback in the last trial? Is it useful and satisfactory?*' After Trial 1, 13 participants provided a response. It was mentioned by 9 participants that the standard lane keeping support warnings were annoying or could be improved, and 5 mentioned that the feedback was useful.

After Trial 2, 18 of 23 participants noted something down. For example, one participant expressed his opinion that auditory feedback when ACC was turned on is not needed as 'You should get 'rewarded' when using for example ACC so there should be sound when it is off instead'. Another participant reported that the ACC-status feedback sounded like a 'malfunctioning fan'.

After Trial 3, 19 of 23 participants provided their feedback on the lane-offset sound. One person said 'Much better than standard function. It supports me instead of dismissing my capability'. A number of people said that such feedback is helpful and can be used especially by novice drivers. However, many participants were displeased with the system, or found it hard to distinguish.

After Trial 3, 18 of 23 participants provided feedback on the headway sound. This sound was given mixed reviews. It was reported that such feedback was difficult to hear and may not be useful in dense traffic. One participant stated

that both the lane-offset sound and the headway sound were not useful when the driver is focused but would be useful during automated driving. Feedback on the combination of all three systems received after Trial 4 was mostly negative, where multiple participants reported not being able to distinguish between sounds from the different types of feedback.

5.4.2 Responses to closed-ended questions

The mean values given to the question of whether it was easy to hear feedback given during the trial are shown in Figure 5. The existing lane keeping system was easy to hear, with unanimous agreement among participants. All three wind-based sounds were hard to perceive, especially the headway sound. This may be due to the fact that short headways were not often experienced, resulting in low overall volume.

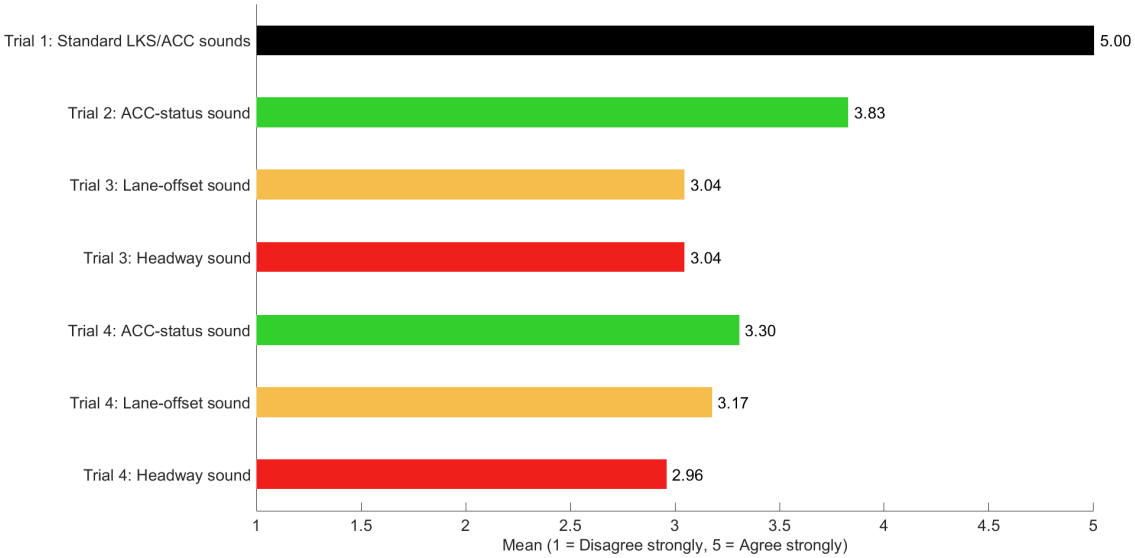


Figure 5. Mean responses regarding whether the sound was easy to hear.

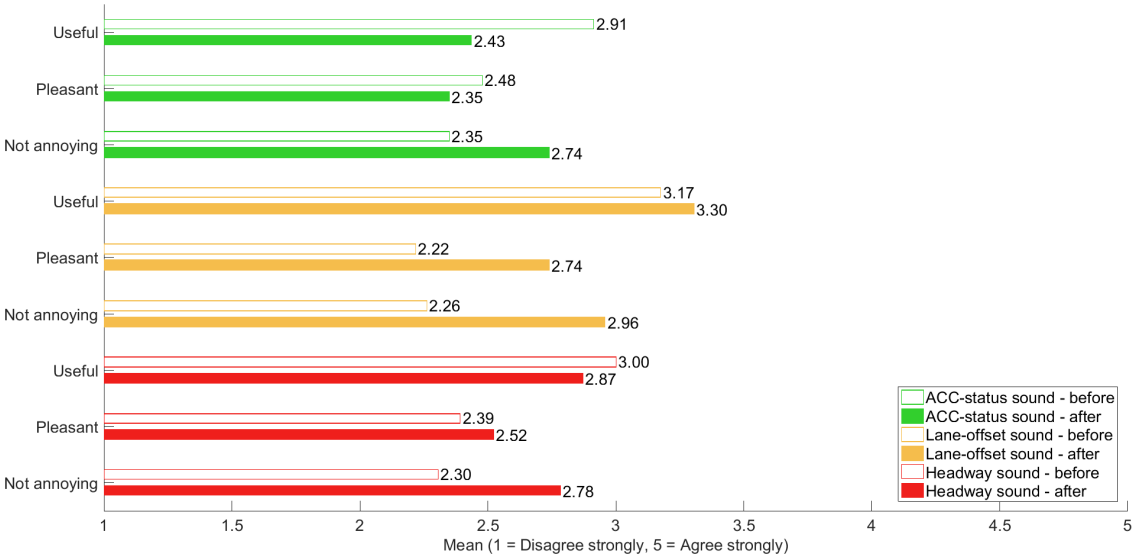


Figure 6. Acceptance of feedback before the experiment (i.e., before Trial 1) and after the experiment (i.e., after Trial 4).

Figure 6 shows the self-reported acceptance of the feedback before the experiment as was reported in the introductory questionnaire, and after the

experiment. Participants saw some merit in the ACC-status sound, with a mean usefulness score of 3.30 on the scale from 1 to 5. However, the ACC-status sound was seen as unpleasant and irritating.

Figure 7 shows the results of the Van der Laan acceptance questionnaire. The results confirm the above observations, where the lane-offset sound was regarded as somewhat useful, whereas the ACC-status sound and all sounds in combination received low acceptance ratings. The mean usefulness ratings were -0.30, 0.14, 0.00, and 0.04 for the ACC-status sound, lane-offset sound, headway sound, and all sounds, respectively. A repeated-measures ANOVA showed no significant difference, $F(3,66) = 1.24$, $p = 0.303$. The mean satisfaction ratings were -0.76, -0.28, -0.08, and -0.49 for the ACC-status sound, lane-offset sound, headway sound, and all sounds, respectively. A repeated-measures ANOVA showed a significant difference, $F(3,66) = 3.51$, $p = 0.020$.

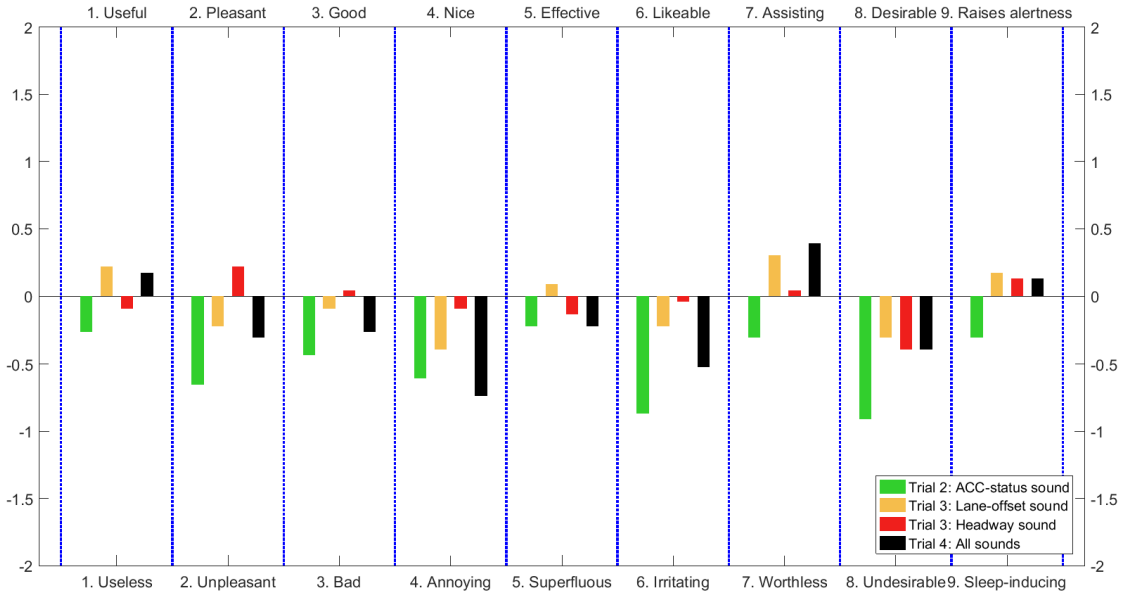


Figure 7. Mean scores on the Van der Laan acceptance questionnaire.

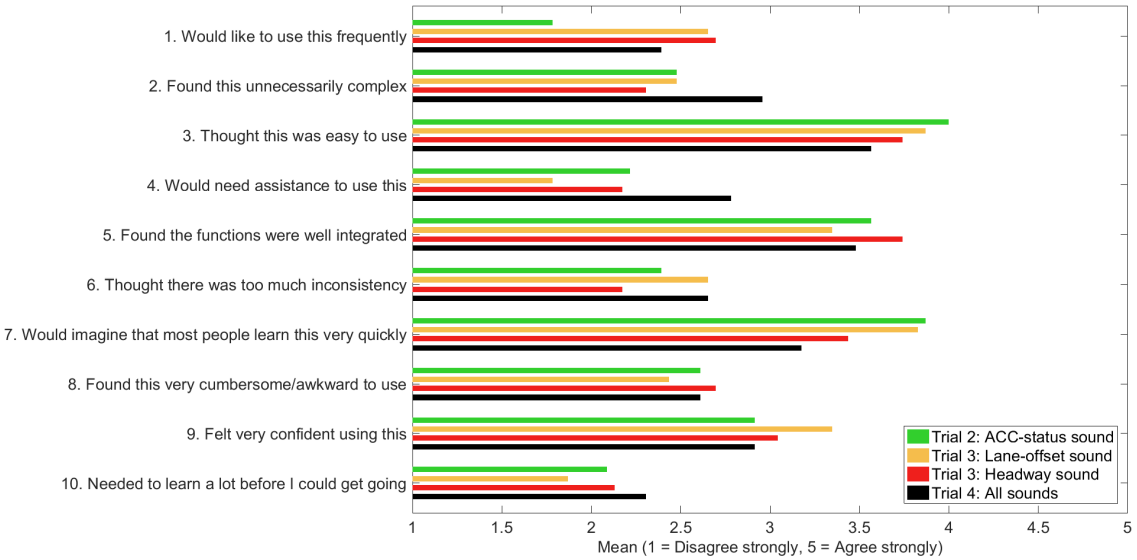


Figure 8. Mean scores on the System Usability Scale as reported after Trials 2–4.

Figure 8 shows the results of the System Usability Scale (SUS). The results indicate that all systems were regarded as easy to use and learn (Items

3 & 7). Furthermore, on average, participants indicated that they would not like using the systems frequently; the ACC-status sound received particularly low ratings (see Item 1). The mean usability scores were 60.9%, 64.6%, 62.9%, and 55.5% for the ACC-status sound, lane-offset sound, headway sound, and all sounds, respectively. A repeated-measures ANOVA showed no significant effect, $F(3,66) = 1.29$, $p = 0.286$.

5.5 Discussion

5.5.1 Main findings and interpretation

Before the experiment, we hypothesized that because our world is essentially continuous and discrete triggers are rare in nature, continuous feedback would be perceived as pleasant and natural. In this study we investigated whether the use of continuous auditory feedback on the status of a truck equipped with ADAS is beneficial for the user experience. The presented experiment showed that our hypothesis may not be true, since truck drivers were not favourable to adding in-vehicle auditory feedback which was intended to blends with the natural engine and wind noise inside the cabin.

All presented concepts were easy to understand for most of the participants. The lane-offset sound was the most accepted type of feedback presented in the study; a number of participants said that such feedback was helpful. The volume level of the headway sound was reported to be too low. The ACC-status sound was not accepted well. Most participants would prefer to have it turned off, instead of having it when the system is on. The presentation of all three sounds together was also regarded as annoying. The combination of all three concepts yielded simultaneous presentation of sounds of different frequencies, which was not tolerated well.

Our results can be explained by the fact that truck drivers are usually confined to their cabin for extensive periods (Roetting et al., 2003) and therefore may not tolerate extra sounds. In fact, much research has been conducted on the cancellation of noise in the cabin (Behar, 1981; Borello, 1999; Mohanty, Pierre, & Suruli-Narayanasami, 2000; Sarigül & Kiral, 1999). Truck drivers have a risk of hearing loss due to the noise in the working environment (Karimi, Nasiri, Kazerooni & Oliaei, 2010). Annoyance due to environmental noise has been shown to be largely determined by overall loudness (Berglund, Berglund, & Lindvall, 1975; Dornic & Laaksonen, 1989). Low-frequency noise with a dominating frequency spectrum of up to 200 Hz was shown to be perceived as more annoying than noises with higher frequency (Persson & Björkman, 1988).

It is our impression that the truck drivers were markedly open and critical; they expressed no social desirability but provided honest feedback on both existing systems and new concepts. Our findings confirm the importance of conducting on-road experiments when developing in-vehicle feedback systems. The present results serve as a useful reminder that theoretically interesting ideas (e.g., the use of continuous auditory feedback) that are proposed in the academic realm may be rejected by end users.

5.5.2 Implications and recommendations

The offered concepts, as they were presented in this study, are not ready to be alternatives to basic auditory warnings used in modern production trucks

equipped with advanced driver-assistance systems. However, they may have future, especially the lane-offset sound, as it received positive comments from some of the participants in the experiment. Accordingly, we recommend testing continuous lane-offset sound in future experiments. As reported by the participants, such feedback is promising. However, it was provided from the left and right speakers at the same time with a weighting factor depending on the deviation from the centre of the lane (see Fig. 4) and in future experiments the effects of issuing such feedback solely from the side of the deviation from the trajectory could be investigated. Further, an improvement may be to disable the sounds when the turn indicators are enabled by the driver. The continuous feedback on the status of ACC or automation of a vehicle may be tested further with reversed feedback, where the sound is on, when the system is turned on. Headway sound should be tested further in a more controlled environment with well-managed headway to the vehicle in front.

Are the present results generalizable to higher levels of automated truck driving? The truck industry is one of the early adopters of automated driving. In 2017, MIT's Technology Review considered automated trucks as one of top 10 Breakthrough Technologies of the year and speculated that the introduction of such trucks would happen in the next 5 to 10 years (Freedman, 2017). Automation in trucks could bring substantial benefits, because drivers may be able to use the periods when the truck is driving automatically to have their mandatory breaks. More revolutionarily, trucks may drive without any drivers on the highway. Truck platooning could be the first commercially successful application of automated driving, where one or multiple trucks within a platoon are automated (Bergenhem et al., 2012; Janssen, Zwijnenberg, Blankers, & De Kruijff, 2015). It is important for drivers of automated trucks are aware of the automation mode because, as with other applications, mode confusions are an important contributor to accidents (Sarter & Woods, 1995). Drivers need to be informed about upcoming mode changes, as well as situations where the system limits are reached, such as when a collision or lane departure is about to occur. It may be interesting to use continuous auditory feedback for presenting the automation mode of the automated truck in a continuous manner, akin to how we presented the ACC mode in the present study. However, we showed that many truck drivers believe that adding additional in-cabin auditory feedback is not beneficial.

It is still possible that other types of continuous auditory feedback may be less annoying and, if designed with special care, more useful than standard discrete sounds. The topic of continuous auditory feedback for in-vehicle interfaces needs further attention.

5.6 Supplementary material

Samples of sounds, questionnaires used during the experiment and MATLAB code for analysis are available at https://www.dropbox.com/sh/d88u0z6al8rl7c1/AADLpKg_MjWL30QpLoCVy0Tva

5.7 References

Bazilinskyy, P., Stapel, J., De Koning, C., Lingmont, H., De Lint, T., Van der

Sijs, T., … De Winter, J. C. F. (2017). Real-time auditory feedback systems based on headway. Proceedings of *HFES Europe Annual Meeting*, Rome, Italy.

Behar, A. (1981). Measurement of noise inside truck cabins. *Applied Acoustics*, *14*, 215–223.

Bergenhem, C., Pettersson, H., Coelingh, E., Englund, C., Shladover, S., & Tsugawa, S. (2012). Overview of platooning systems. Proceedings of the *19th ITS World Congress, Vienna, Austria*, 1–7.

Berglund, B., Berglund, U., & Lindvall, T. (1975). Scaling loudness, noisiness, and annoyance of aircraft noise. *The Journal of the Acoustical Society of America*, *57*, 930–934. https://doi.org/10.1121/1.381212

Blauert, J. (1985). Spatial hearing - The psychophysics of human of sound. *Journal of the Acoustical Society of America*, *77*, 334–335. https://dx.doi.org/10.1121/1.392109

Borello, G. (1999). Predicting noise transmission in a truck cabin using the Statistical Energy Analysis approach. Proceedings of *IUTAM symposium on statistical energy analysis*, 281–287.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89*, 244–249.

Brooke, J. (1986). System usability scale (SUS): a quick-and-dirty method of system evaluation user information. Reading, UK: Digital Equipment Co Ltd, 43.

Campbell, J. L., Brown, J. L., Graving, J. S., Richard, C. M., Lichty, M. G., Sanquist, T., ... & Morgan, J. F. (2016). Human Factors Design Guidance for Driver-Vehicle Interfaces (DVI). *National Highway Traffic Safety Administration*.

Chen, F., Qvint, G., & Jarlengrip, J. (2007). Listen! there are other road users close to you–improve the traffic awareness of truck drivers. Proceedings of the *International Conference on Universal Access in Human-Computer Interaction*, 323–329. https://doi.org/10.1007/978-3-540-73281-5_3

Crispien, K., Fellbaum, K., Savidis, A., & Stephanidis, C. (1996). A 3D-auditory environment for hierarchical navigation in non-visual interaction. Proceedings of *the Third International Conference on Audio Display (ICAD)*, 18–21. https://doi.org/10.1145/228347.228366

Dornic, S., & Laaksonen, T. (1989). Continuous noise, intermittent noise, and annoyance. *Perceptual and Motor Skills*, *68*, 11–18. https://doi.org/10.2466/pms.1989.68.1.11

Freedman, D. H. (2017). Self-driving trucks: 10 breakthrough technologies 2017. Available at https://www.technologyreview.com/s/603493/10-breakthrough-technologies-2017-self-driving-trucks

Janssen, R., Zwijnenberg, H., Blankers, I., & de Kruijff, J. (2015). *Truck platooning: Driving the future of transportation*.

Karimi, A., Nasiri, S., Kazerooni, F. K., Oliaei, M. (2010). Noise induced hearing loss risk assessment in truck drivers. *Noise and Health*, *12*, 49–55. https://doi.org/10.4103/1463-1741.59999

Kidd, D. G., Cicchino, J. B., Reagan, I. J., & Kerfoot, L. B. (2017). Driver trust in five driver assistance technologies following real-world use in four production vehicles. *Traffic Injury Prevention, 18*, S44–S50.

Mohanty, A. R., Pierre, B. D. S., & Suruli-Narayanasami, P. (2000). Structure-

borne noise reduction in a truck cab interior using numerical techniques. *Applied Acoustics*, *59*, 1–17.

Parasuraman, R., Hancock, P. A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centred collision-warning systems. *Ergonomics*, *40*(3), 390–399.

Persson, K., & Björkman, M. (1988). Annoyance due to low frequency noise and the use of the dB (A) scale. *Journal of Sound and Vibration*, *127*, 491–497.

Rath, M., & Rocchesso, D. (2005). Continuous sonic feedback from a rolling ball. *IEEE Multimedia*, *12*, 60–69. https://doi.org/10.1109/MMUL.2005.24

Roetting, M., Huang, Y. H., McDevitt, J. R., & Melton, D. (2003). When technology tells you how you drive - Truck drivers' attitudes towards feedback by technology. *Transportation Research Part F: Traffic Psychology and Behaviour*, *6*(4), 275–287. https://doi.org/10.1016/j.trf.2003.09.001

Sarigül, A. S., & Kiral, Z. (1999). Interior acoustics of a truck cabin with hard and impedance surfaces. *Engineering Analysis with Boundary Elements*, *23*, 769–775.

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors, 37*, 5–19.

Stanton, N. A. (Ed.). (1994). Human factors in alarm design. CRC Press.

Van der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C, 5*, 1–10. https://doi.org/10.1016/S0968-090X(96)00025-3

6 When will most cars be able to drive fully automatically?

Corresponding article: Bazilinskyy, P., Kyriakidis, M., & De Winter, J. C. F. (2018). When will most cars drive fully automatically? An analysis of international surveys. Under review.

6.1 Abstract

When fully automated cars will be widespread is a question that has attracted considerable attention from futurists, car manufacturers, and academics. The aim of this paper is to poll the public's expectations regarding the deployment of fully automated cars. In 14 crowdsourcing surveys conducted between June 2014 and October 2018, we obtained answers from 18,271 people from 128 countries regarding when they think that most cars will be able to drive fully automatically in their country of residence. The median reported year was 2030. We found that the later the survey date, the smaller the percentage of respondents who reported that most cars will be able to drive fully automatically by 2020, with 15–22% of the respondents providing this estimate in the surveys conducted between 2014 and 2016 versus 4–5% in the 2018 surveys. Respondents who completed multiple surveys were more likely to revise their estimate upward (40.0%) than downward (34.8%). Correlations at the individual level and national level show that people from more affluent countries and people who have heard of the Google Driverless Car (Waymo) or the Tesla Autopilot reported a significantly earlier year. Finally, we made a comparison between the crowdsourced respondents and respondents from a technical university who answered the same question; the median year reported by the latter respondents' group was 2040, that is, later than the median estimate given by the crowdsourced respondents. We conclude that over the course of four years respondents have moderated their expectations regarding the penetration of fully automated cars, but nevertheless public's expectations remain exceedingly optimistic compared to what experts currently believe.

6.2 Introduction

Fully automated driving is expected to improve road safety and traffic flow efficiency and may have a large influence on transportation businesses (e.g., car insurance) and the shape of road infrastructure (Fagnant & Kockelman, 2015). Parking spaces within cities may soon no longer be needed, and road networks will likely change. In order to develop appropriate transport policies, it is important to predict when fully automated driving will be commonplace.

Futurists have long been concerned with making predictions about the introduction of automated vehicles. As early as 1940, Geddes outlined detailed predictions of automated highway systems to be deployed in the United States (Geddes, 1940). In the late 1980s, Kurzweil predicted that by the end of the

1990s/early 2000s "the cybernetic chauffeur, installed in one's car, communicates with other cars and sensors on the roads. In this way it successfully drives and navigates from one point to another" (Kurzweil, 1990). In 2012, Kurzweil admitted that his prediction was wrong, yet noted that it was "not all wrong", considering the achievements in the Google self-driving car project (Kurzweil, 2012).

Today, in light of recent technological developments and on-road tests of automated driving systems, predictions on the advent of fully automated driving have moved from futurism into mainstream science. Automotive manufacturers are already testing their automated vehicles on public roads, with Google/Waymo having recently the milestone of 10 million self-driven miles across 25 cities in various states (Waymo, 2018) and being responsible for 97% of the total travelled distance in California in 2016 and for 72% in 2017, by logging 1,023,330 km and 567,365 km, respectively (Department of Motor Vehicles, 2017, 2018). However, these vehicles are not commercially viable yet, and do not formally fulfil the definition of *fully* automated driving, because a human driver occasionally has to take over control (Dixit, Chand, & Nair, 2016).

In August 2013, Nissan revealed plans for fully automated vehicles in 2020 (NissanNews.com, 2013), an estimate that was revised to 2022 in December 2017 (Nissan Motor Corporation, 2017) and repeated in April 2018 (Nissan Motor Corporation, 2018). In July 2016, BMW predicted that their first fully automated cars would be in production by 2021 (BMW News, 2016). In September 2018, the company presented the iNext model to be put in production in 2021; this is not a fully autonomous car but a highly automated one with a steering wheel that "retracts slightly" when in automated mode (BMW Group, 2018). Similarly, in 2016 Ford announced that they expect their first fully automated cars for commercial ride sharing in 2021, although according to the chief technical officer of the company, fully automated cars with no steering wheel or pedals are unlikely to be available to customers before 2025 (Sage & Lienert, 2016); the company's website in November 2018 refers to 2021 as the year when "the car will operate without a steering wheel, gas pedal or brake pedal within geo-fenced areas…. By doing this, the vehicle will be classified as a SAE Level 4 capable-vehicle" (Ford Motor Company, 2018). Also in 2016, Continental stated that they would be ready for production of fully automated cars by 2025 (Continental AG, 2016), an estimate persisting in September 2018 (Continental AG, 2018). On the one hand, automotive manufacturers should be able to make accurate predictions regarding the deployment of fully automated cars, because it is the car manufacturers that together with OEMs and ICT companies develop and will sell those vehicles. On the other hand, the predictions by automotive manufacturers presented in the media may not be the most reliable source of information, because of potential conflicts of interest in the market uptake.

Shladover, one of the pioneers of automated driving research in the US, argued that it is unlikely for fully automated cars to arrive any time soon: "fully automated vehicles capable of driving in every situation will not be here until 2075. Could it happen sooner than that? Certainly. But not by much." (Shladover, 2016). In a survey among 217 attendees of an automated vehicle conference (31% of whom were employed in academia, 24% in the automotive industry, and 9% in government positions), Underwood observed a median of 2030 regarding the estimate when fully automated driving will be introduced to the market in the United States (Underwood, 2014). In a scenario-construction

study, 20 transport experts in the Netherlands predicted that between 7% and 61% of the vehicle fleet would be fully automated by 2050, depending on how fast the technology will develop and on how restrictive or supportive the associated policies will be (Milakis, Snelder, Van Arem, Van Wee, & Correia, 2017).

Besides polling the vision of automotive manufacturers and scientists, it is also important to poll what the public thinks regarding the deployment of fully automated cars. It is the public who should eventually buy and use such vehicles and who will ultimately determine their future success. Currently, a large number of cars with features of (partially) automated driving are purchased, because they are seen as an interesting new technology. Previous surveys indicate that people appreciate automated driving, with a reduction in traffic accidents, emissions, and energy consumption being reported as important benefits (Bansal, Kockelman, & Singh, 2016; Piao, McDonald, Hounsell, Graindorge, Graindorge, & Malhene, 2016; Schoettle & Sivak, 2014). However, survey research has also revealed concerns about the security, privacy, legal liability, and ethical decisions of automated vehicles (Bonnefon, Shariff, & Rahwan, 2016; Kyriakidis, Happee, & De Winter, 2015; Schoettle & Sivak, 2014). There is currently no clear insight into when the public expects autonomous driving to be ubiquitous. This study aims to poll the public's expectation regarding the moment when fully automated cars will be ubiquitous.

6.3 Methods

6.3.1 Surveys

Between June 2014 and October 2018, we performed 14 surveys via crowdsourcing, to poll people's opinion on various aspects of automated driving, such as user's acceptance, worries, and willingness to buy automated vehicles, preferences for warning systems of different modalities, etc. In each survey, the following question was included: "*In which year do you think that most cars will be able to drive fully automatically in your country of residence?*" Here, we analyze the responses of the combined sample of respondents to this question across the 14 surveys. Table 1 shows the characteristics of the 14 surveys. In all surveys, 'level 1' crowdsourcing (defined by CrowdFlower as "All qualified contributors") was selected.

All data were collected anonymously. The research was approved by the Human Research Ethics Committee (HREC) of the Delft University of Technology. In all surveys, informed consent was obtained via a dedicated survey item asking whether the respondent had read and understood the survey instructions.

6.3.2 Data filtering

Per survey, we excluded respondents who did not indicate 'yes' to a question whether they had read the survey instructions, who indicated they were under 18 years old, who said they were older than 110 years, who did not respond to the question about their age or gender, or for whom no country information was provided by CrowdFlower. In some of the surveys, it was possible to generate multiple responses from different worker IDs with the same IP address. In these cases, we kept only the results from the first completion. The fastest 5% of the

respondents were also removed from the analyses (as in De Winter & Dodou, 2016).

Table 1. Overview of the 13 surveys

Survey	Period of completion	Subject
S1 (De Winter, Kyriakidis, Dodou, & Happee, 2015)	Jun 16, 2014–Jun 17, 2014	Knowledge of automated driving systems and cross-national differences in traffic violations as measured with the Manchester Driver Behaviour Questionnaire (DBQ).
S2 (Kyriakidis et al., 2015)	Jul 4, 2014–Jul 7, 2014	User acceptance, worries, and willingness to buy partially, highly, and fully automated vehicles; cross-national differences and correlations with personal variables, such as age, gender, and personality traits as measured with a short version of the Big Five Inventory.
S3 (Bazilinskyy & De Winter, 2015)	Sep 2, 2014	User acceptance of auditory interfaces in modern cars and their willingness to be exposed to auditory feedback in highly and fully automated driving. A 7-item DBQ was also completed.
S4 (De Winter & Hancock, 2015)	Nov 29, 2014–Nov 30, 2014	Opinion on whether humans surpass machines or machines surpass humans.
S5 (Bazilinskyy, Petermeijer, Petrovych, Dodou, & De Winter, 2017)	Mar 31, 2015–Apr 1, 2015	Preferences for auditory, visual, and vibrotactile take-over requests in highly automated driving; the survey included recordings of auditory messages and illustrations of visual and vibrational messages. A 7-item DBQ was also completed.
S6 (De Winter & Dodou, 2016)	Dec 24, 2015–Dec 27, 2015	Relationships between traffic violations measured with a 7-item DBQ and traffic accident involvement.
S7 (Bazilinskyy & De Winter, 2017a)	May 30, 2016–Jun 5, 2016	Effects of speech-based take-over requests on perceived urgency, commandingness, pleasantness, and ease of understanding; respondents listened to a random 10 out of 140 take-over requests and rated each take-over request in terms of the four aforementioned criteria. A 7-item DBQ was also completed.
S8 (Kovácsová, De Winter, & Hagenzieker, 2017)	Feb 27, 2017–Feb 28, 2017	Investigation of cyclists' behavior when approaching an intersection. The survey consisted of a questionnaire regarding cycling behavior, skills, and experience. Moreover, respondents watched videos from real traffic and answered questions about their predictions of what will happen next.
S9 (Bazilinskyy & De Winter, 2017b)	Mar 3, 2017–Mar 4, 2017	Determination of reaction times for different types of visual and auditory signals. Respondents participated in a reaction-time measurement task and filled in the DBQ.
S10 (Kovácsová, De Winter, & Hagenzieker, 2017)	Mar 4, 2017–Mar 7, 2017	Same as Survey 8, but now repeated among 15 selected Western high-income countries.
S11	Jun 16, 2017 –Jun 18, 2017	Cross-national differences in traffic violations as measured with the DBQ.
S12 (Rodríguez Palmeiro, Van der Kint, Hagenzieker, Van Schagen, & De Winter, 2018)	Jul 7, 2017 –Jul 12, 2017	Cyclist's behaviour when interacting with automated vehicles. Conducted among the same 15 selected Western high-income countries as S10.

S13	Apr 19, 2018–Apr 23, 2018	Cross-national differences in traffic violations as measured with the DBQ.
S14	Oct 3, 2018–Oct 29, 2018	External human-machine interfaces for automated driving.

Note. In S2, only numeric entries were permitted, whereas in the rest of the surveys textual responses were also allowed. In S10 and S12, we only permitted respondents from 15 targeted Western high-income countries (Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Sweden, Switzerland, United Kingdom, United States).

Responses reporting the year 2013 or earlier were excluded. If a respondent's answer equaled 'never' (i.e., single-word answer, case-insensitive), the answer was coded as 9999. Other textual responses were excluded from the analysis.

6.3.3 Analysis at the individual level

Analyses were conducted both at the individual level of respondents and at the national level. For the former, the distribution of the reported year when most cars are expected to be able to drive fully automatically was calculated. For respondents who participated in more than one survey, only the response from their first survey was included. Moreover, we calculated Spearman's rank-order correlations between the reported year and the following variables per respondent:

- the respondent's age;
- the respondent's gender;
- the respondent's self-reported violations. The self-reported violations were computed from Surveys 1, 3, 5, 6, 7, 9, 11, 13, and 14, which included a 7-item Driver Behaviour Questionnaire (DBQ; De Winter, 2013). We calculated a non-speeding violations score based on the following items: 1. using a mobile phone without a hands free kit, 2. driving so close to the car in front that it would be difficult to stop in an emergency, 3. sounding the horn to indicate annoyance with another road user, 4. becoming angered by a particular type of driver, and indicate hostility by whatever means one can, and 5. racing away from traffic lights with the intention of beating the driver next to own vehicle. Additionally, we calculated a speeding violation score from the following items: 1. disregarding the speed limit on a residential road, and 2. disregarding the speed limit on a motorway;
- the respondent's familiarity with automated driving. For this, we relied on Surveys 1, 6, 11, and 13, in which we asked respondents whether they had heard of the Google Driverless Car (Waymo), and Surveys 11 and 13 asked whether respondents had heard of the Tesla Autopilot.

A longitudinal analysis was also carried out to investigate whether respondents who participated in more than one of the surveys adjusted their expectations between their first and last survey.

6.3.4 Analysis at the national level

The analysis at the national level examined the relationships between the median years when most cars will be able to drive fully automatically and national developmental indexes. Specifically, we used the following variables per country:

- road traffic death rate per 100,000 population in 2013 (World Health Organization, 2015);
- gross domestic product (GDP) per capita in 2013 (World Bank, 2015);
- national performance in educational tests (Rindermann, 2007);
- average life expectancy in 2013 (World Bank, 2015);
- self-reported speeding violations and non-speeding violations (from Surveys 1, 3, 5, 6, 7, 9, 11, 13, and 14);
- motor vehicle density (cars, buses, and freight vehicles, but not two-wheelers, per 1,000 people) averaged over the years 2003–2010 (World Bank, 2015);
- median age in 2014 (Central Intelligence Agency, 2015).

In the national analysis, to reduce sampling error, we selected only those countries with 25 or more respondents having provided a numeric response or 'never'. If a respondent had completed 2 or more of the 14 surveys, the reported years were averaged across the completed surveys. We calculated a Spearman correlation matrix of the median year of introduction of fully automated cars as collected from the surveys, respondents' gender (percentage of male respondents in each country), respondents' mean age, and the above national variables.

6.4 Results

6.4.1 Results at the individual level

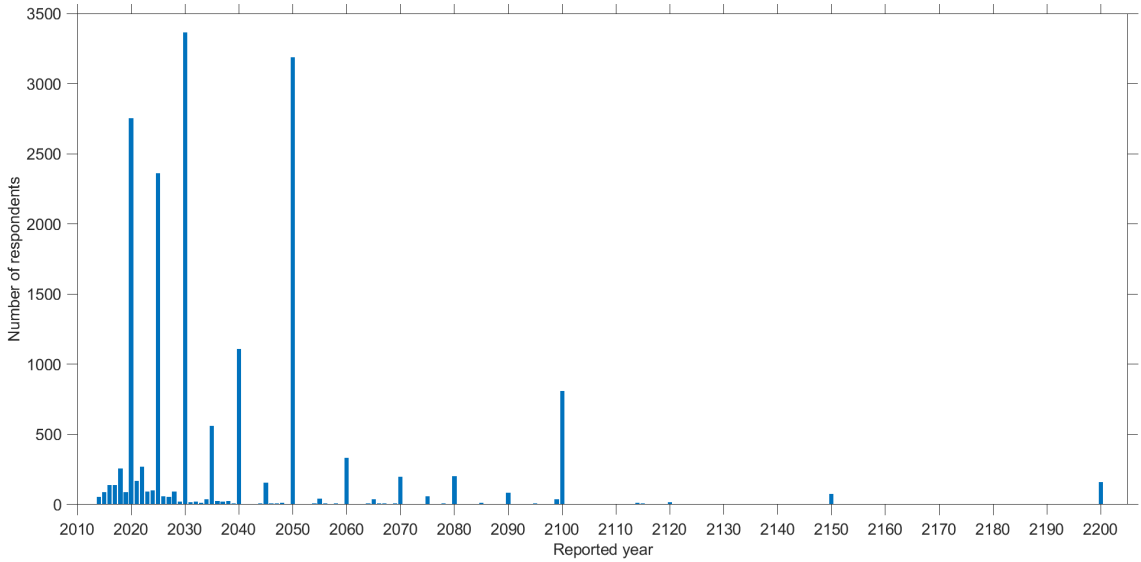


Figure 1. Distribution of the reported year across surveys.

Table 2 provides descriptive statistics of the respondents per study. There were 20,251 respondents from 130 countries, of whom 18,271 respondents from 128

countries provided a numeric response to the question of interest or answered 'never'. Across these 18,271 respondents, the reported year exhibited a skewed distribution, with a clear zero end-digit preference (Figure 1).

Table 2. Respondents' characteristics

Survey	# respondents	# respondents included	# unique countries	# respondents reporting a numeric year	# respondents reporting 'never'	% males	mean age
Jul 2014	1854	1711	91	1520	44	66.8	32.7
Jul 2014	5000	4365	105	3709	0	68.9	32.8
Sep 2014	2000	1656	95	1481	13	74.6	31.6
Nov 2014	2999	2800	104	2625	22	71.9	31.8
Mar 2015	3000	2794	101	2581	9	73.5	32.4
Dec 2015	3250	2935	95	2654	34	69.8	33.8
May 2016	3061	2842	98	2616	20	66.7	33.8
Feb 2017	700	633	60	550	5	75.1	32.6
Mar 2017	2000	1848	84	1702	14	70.6	34.0
Mar 2017	700	638	15	593	10	48.8	38.0
Jun 2017	2500	2249	92	2069	22	69.0	33.1
Jul 2017	700	630	15	597	4	47.1	38.6
Apr 2018	3000	2627	84	2427	22	64.4	33.7
Oct 2018	1770	1586	73	1441	7	63.3	34.6
Total		20251	130	18271	156	69.3	31.6

*Note***.** The percentage of male respondents and the respondents' mean age were calculated for the respondents who reported a numeric year or 'never'.

Table 3 shows that across the 14 surveys, 23–49% of the respondents reported a year between 2017 and 2029. The median predicted year across all surveys was 2030. Respondents in the more recent surveys were less likely to report that most cars will drive fully automatically by 2020 (Figure 2), with 15–22% of the respondents providing this estimate in the surveys conducted between 2014 and 2016 versus 4–5% in the 2018 surveys (Table 3).

Table 3. Distribution of the reported year per survey

Survey	Median year (P25, P75)	Percentage of respondents			
		2020	2030	2017–2029	2075+ (including 'never')
Jul 2014	2030 (2022, 2050)	18	16	34	16
Jul 2014	2030 (2021, 2050)	19	17	38	11
Sep 2014	2030 (2020, 2050)	22	16	42	11
Nov 2014	2030 (2025, 2050)	16	16	35	14
Mar 2015	2030 (2020, 2045)	22	19	47	7
Dec 2015	2030 (2025, 2050)	16	17	37	12
May 2016	2030 (2025, 2050)	15	20	38	10
Feb 2017	2035 (2025, 2050)	9	18	30	17
Mar 2017	2030 (2025, 2050)	10	19	30	13
Mar 2017	2030 (2025, 2040)	14	21	43	9
Jun 2017	2030 (2025, 2050)	9	22	30	12
Jul 2017	2030 (2025, 2035)	13	21	49	4
Apr 2018	2035 (2029, 2050)	5	22	25	14
Oct 2018	2035 (2030, 2050)	4	24	23	14
Total	2030 (2025, 2050)	15	18	35	12

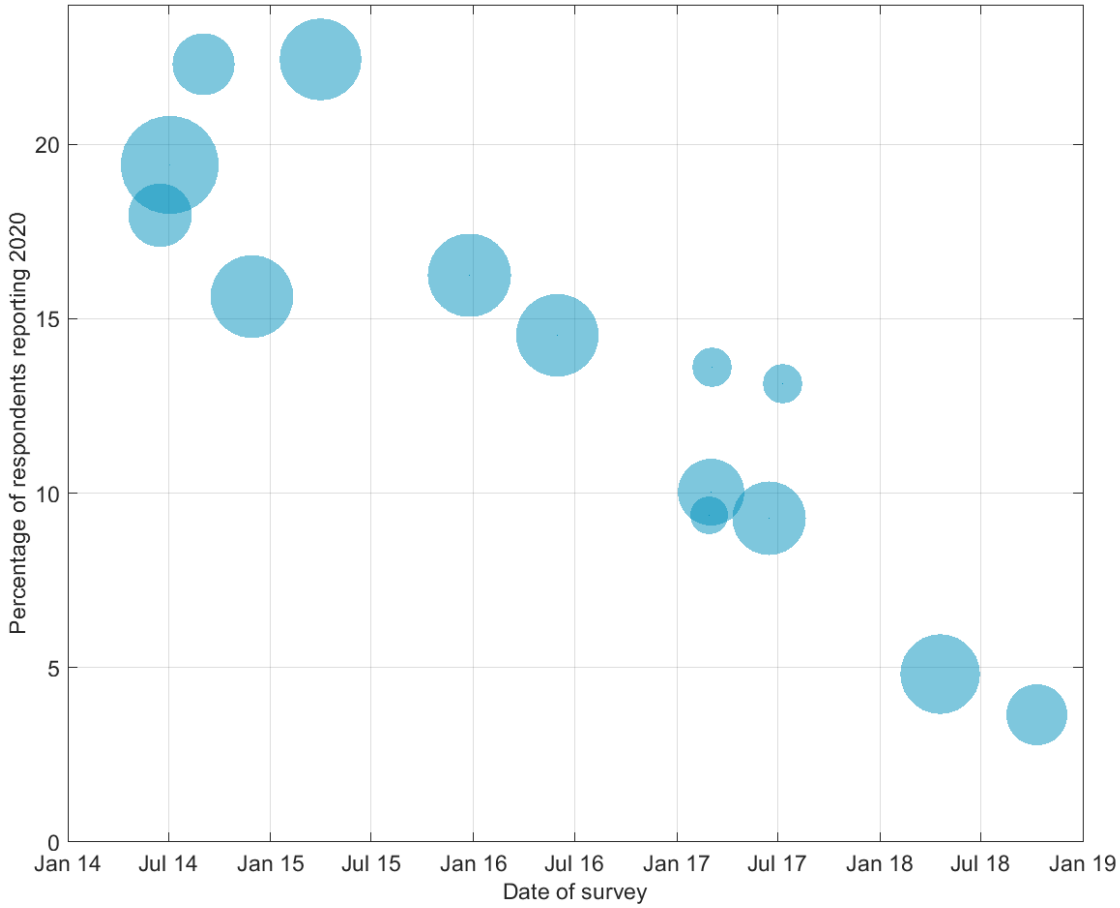


Figure 2. Percentage of respondents reporting '2020', as a function of the survey date. The area of each circle linearly corresponds to the number of respondents who provided a numeric response or reported 'never'.

Figure 3 shows correlations between individual characteristics and the reported year when most cars will drive fully automatically. Males reported a significantly higher year than females (p = 0.003), although the correlation between the reported year and gender was small (ρ = 0.02). There were no

significant correlations of the reported year with age, nor with self-reported traffic violations. However, people who were more familiar with automated driving technology (i.e., who had heard of the Google Driverless Car (Waymo) or the Tesla Autopilot) reported a significantly earlier year than participants who answered 'no' to these questions ($p < 0.001$). It is worth noting that the percentage of respondents who answered 'yes' to the question of whether they had heard of the Google Driverless Car was low (48%, 57%, 56%, and 45%, for Surveys 1, 6, 11, and 13, respectively). Similarly, the percentage of respondents who answered 'yes' to the question of whether they had heard of the Tesla Autopilot was 55% and 60% for Surveys 11 and 13, respectively.

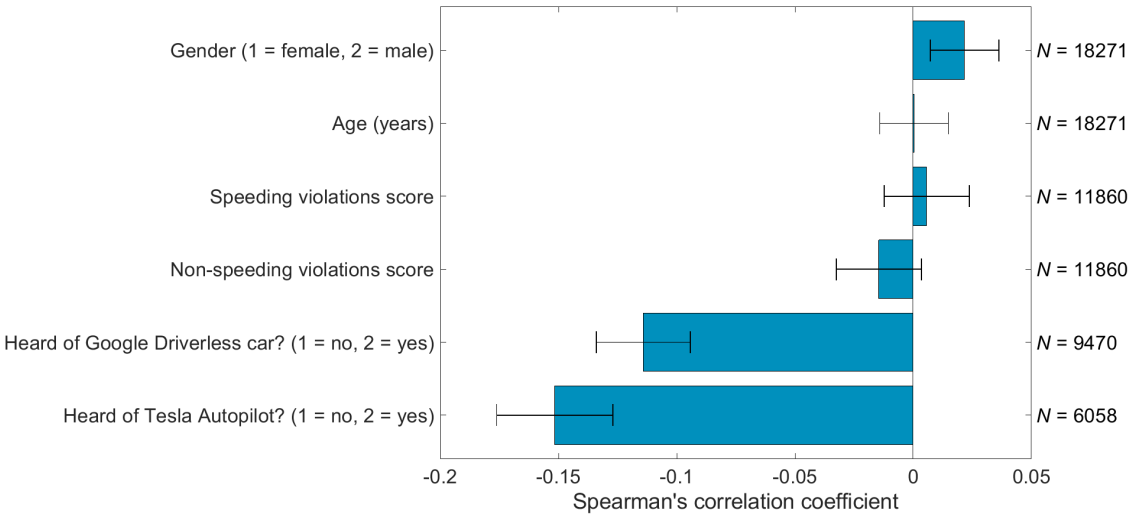


Figure 3. Spearman's correlation coefficients (equivalent to Pearson correlations after rank-transforming the variables) between the reported year and various individual characteristics. The error bars represent 95% confidence intervals.

5,258 respondents completed 2 or more of the 14 surveys, and 4,747 of them reported a year in at least two surveys. Among these 4,747 respondents, 25.3% indicated the same year in their first and last survey, 40.0% revisited their estimate upward, and 34.8% revisited their estimate downward. The year reported in the returning respondents' first and last surveys was significantly different (Wilcoxon signed-rank test: $p < 0.001$, sign statistic = 1651, z value = −4.11, Spearman ρ between the respondents' first and last survey = 0.49).

6.4.2 Results at the national level

Table 4 shows cross-national correlations for the 63 countries with 25 or more respondents. There was a tendency of people in more highly developed countries (in terms variables 6–11) to report an earlier year ($|\rho| < 0.33$). However, these correlations were small compared correlations among the national variables themselves (i.e., variables 6–11 exhibit correlations of $|\rho| > 0.60$). Figures 4 and 5 illustrate that the country's GDP was moderately correlated with the median year, and strongly correlated with self-reported non-speeding violations. Table S1 in the supplementary material presents results for each country separately.

Table 4. Spearman correlation matrix at the national level (N = 63, N = 58 for speeding and non-speeding violations)

	1	2	3	4	5	6	7	8	9	10
1 R: median year										
2 R: gender (% males)	0.16									
3 R: mean age	-0.12	-0.71								
4 R: speeding violations	0.21	-0.08	0.18							
5 R: non-speeding violations	0.17	0.54	-0.55	0.18						
6 S: road traffic death rate	0.19	0.51	-0.60	-0.07	0.73					
7 S: GDP per capita	-0.33	-0.54	0.66	0.09	-0.66	-0.73				
8 S: educational performance	-0.10	-0.63	0.69	0.17	-0.73	-0.79	0.78			
9 S: life expectancy	-0.25	-0.49	0.54	0.14	-0.61	-0.78	0.87	0.77		
10 S: motor vehicle density	-0.16	-0.64	0.77	0.28	-0.62	-0.66	0.83	0.77	0.76	
11 S: median age	0.10	-0.57	0.65	0.32	-0.60	-0.69	0.62	0.75	0.64	0.75

Note. 'R' indicates that data that were obtained from the CrowdFlower respondents. 'S' indicates that data that were obtained from previously published national statistics.

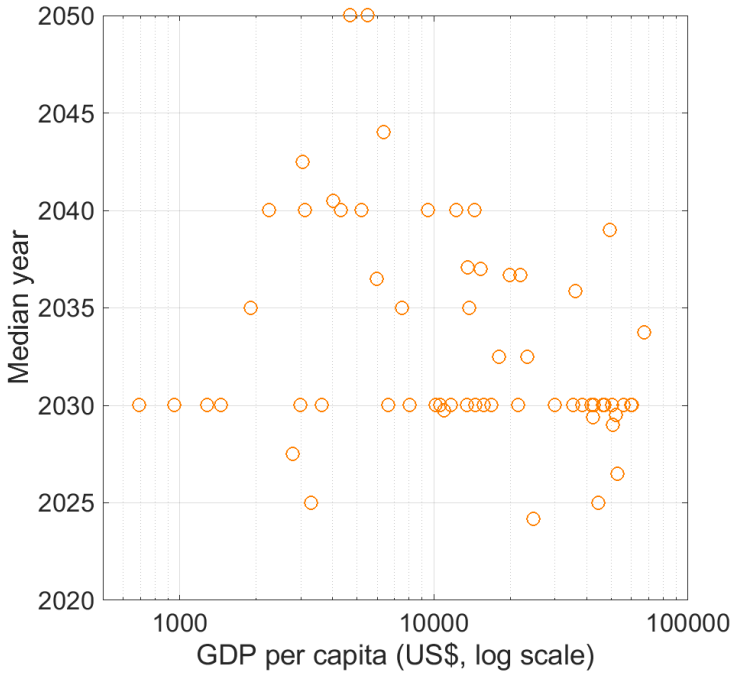


Figure 4. Median predicted year versus gross-domestic product (GDP) per capita (ρ = −0.33). Each marker represents a country.

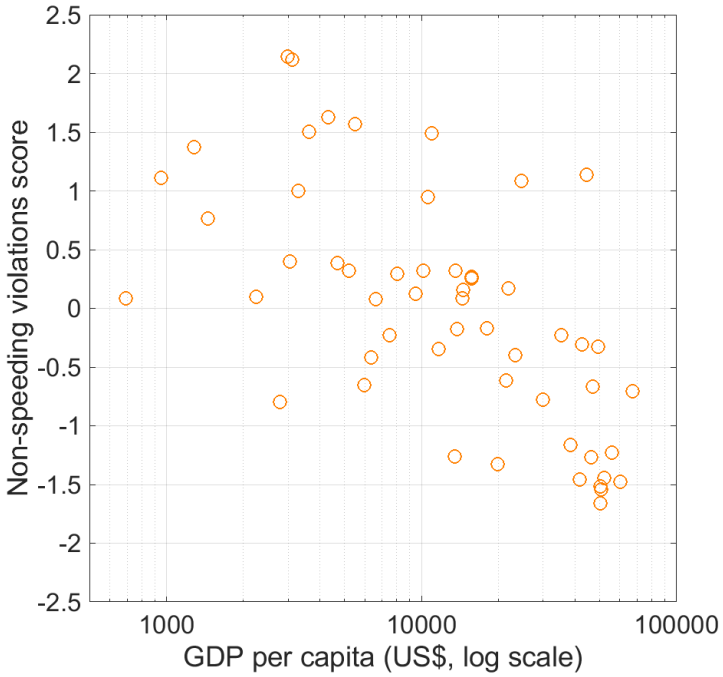


Figure 5. Non-speeding violations score versus gross-domestic product (GDP) per capita (ρ = –0.66). Each marker represents a country.

6.4.3 Control study with participants from a technical university

In addition to the crowdsourced surveys, we conducted a control study in March 2017 using 38 participants (31 males, 7 females, mean age = 26.6 years, *SD* age = 6.5 years). Participants were students and staff members of the Faculty of Mechanical, Maritime and Materials Engineering at the Delft University of Technology. The participants responded to the same questions, including the question on the year when most cars will drive fully automatically in their country of residence. The control sample reported a median year of 2040 (P25 = 2027, P75 = 2050). The minimum reported year was 2022. In comparison, across the 14 online surveys, there were 121 respondents from the Netherlands, and their median reported year was 2028.

6.5 Discussion

Over the course of four years, we conducted 14 online surveys in which we asked respondents when most cars will be able to drive fully automatically in their country of residence. The median reported year across all surveys was 2030, which is more optimistic than previously published expert estimates (Milakis et al., 2017; Litman, 2018; Shladover, 2016; Underwood, 2014). Underwood (2014) reported 2030 as median estimate of when fully automated driving will be *introduced* to the market (where fully automated vehicles were defined as “Vehicle is in control from beginning to end of trip, both on highway and surface streets, urban and rural, without human intervention”), whereas in our surveys, we polled the respondents’ opinion about the year when *most cars* will be able to drive fully automatically in their country of residence.

Returning respondents on average revised their initial estimate to a later year. In our first surveys launched in 2014–2016 between 15 and 22% of respondents reported 2020 as the predicted year and this had reduced to 4–5% in the surveys deployed in 2018. This calibration of predictions can be explained

by the fact that in 2014–2016, 2020 still appeared to be subjectively 'far away', making it plausible that most cars could drive fully automatically by then. Now that 2020 is less than two years away, it has become unlikely that fully automated cars will be ubiquitous by that time.

There are several reasons why 2030 can be regarded as a too optimistic prediction of when most cars will be able to drive fully automatically. First, there may be a large temporal lag between the introduction of fully automated vehicles and their widespread adoption. For Electronic Stability Control (ESC), for example, the lag was 20 years: ESC was introduced in 1995 and is included in most registered vehicles in the US since 2015 (Zuby, 2016). Kröger, Kuhnimhof, and Trommer (in press) estimated penetrated rates of fully automated vehicles by 2035 between 10% and 38% in Germany and between 8% and 29% in the United States. By taking into account the turnover rate of modern cars, Litman (2018) forecasted that 40% of the vehicle fleet would consist of fully automated vehicles by 2040. The introduction of fully automated cars may be accompanied by a shift in the organization of road transport. Examples are dedicated lanes for automated driving, and vehicle sharing via dynamic trip-vehicle assignment (Alonso-Mora, Samaranayake, Wallar, Frazzoli, & Rus, 2017). Such innovations, together with governmental mandates, accelerating technological change, and growing public acceptance, may make it possible that the lag between the introduction of fully automated cars and their widespread use will be shorter than the aforementioned 20 years. Second, the computer intelligence required for fully automated driving is high (Geiger, Lauer, Wojek, Stiller, & Urtasun, 2014; Ohn-Bar & Trivedi, 2016). Sierhuis pointed out that fully automated cars will need to anticipate whether a pedestrian will cross the road based on the body language of that pedestrian: "Can you imagine our autonomous vehicles figuring out that they [pedestrians] are not going to cross? That is a very very complex problem to solve." (Sierhuis, 2016, 49:38–49:45).

Crowdsourcing respondents may not be representative of the general population, and their expectations may have been influenced by the media. Shladover (2016) noted: "My concern is that the public's expectations have been raised to unreasonable levels because of the hype out there on the Internet". As evidenced by the dot-com bubble between 1995 and 2001 (Ofek & Richardson, 2003), overestimations regarding technology can have large socio-economic consequences. It should be noted that while the respondents' opinion is probably biased, experts may also suffer from professional deformation (Menton, 1957).

It is possible that respondents gave a fast and intuitive answer and did not deliberatively reflect on the future of automated driving. The fact that people gave more optimistic predictions than experts may imply that respondents were not knowledgeable or that the notion of fully automated driving was unclear to them. Some respondents may have been thinking about technology that is formally known as highly, conditionally, or partially automated driving systems (such as the Tesla Autopilot). Future research could be conducted using multiple-item surveys and explicit definitions or multimedia illustrations of fully automated driving.

Our results indicated that respondents in more developed countries provided more optimistic estimates regarding when most cars will be able to drive fully automatically in their country of residence. An explanation is that high-income countries have high-quality road infrastructure on which automated

vehicles can be deployed. A second explanation is that most companies developing fully automated vehicles are located in the high-income countries. Third, in high-income countries, more people are able to afford luxury goods such as automated cars. It may also be that these answers have been confounded, as respondents from higher income countries were more likely to be female and older (Table 4), and to exhibit a more law-abiding driving style than respondents in lower-income countries (Figure 5). Furthermore, we caution that correlations at the national level are not generalizable to the individual level (Pollet, Tybur, Frankenhuis, & Rickard, 2014). In Survey 2, we asked respondents about their educational level and yearly income via 7- and 14-point items, respectively. The median within-country Spearman correlation between the reported year versus education and income were close to zero (0.03 and −0.01, respectively, N = 37 countries). In other words, the correlation between the predicted year and income is observed between countries, not within countries.

We conclude that the crowdsourced public gave more optimistic predictions about automated driving than experts. Monitoring how the public's expectations change over time can offer important insight into the evolution of knowledge and trust in automated driving.

6.6 Supplementary material

Data from all surveys used in the analysis is available at https://www.dropbox.com/sh/4w5os9lh602896h/AADpvfPoA3Klc5V11pQ_VDvZa?dl=0

6.6.1 Table S1. Data at the national level

Country	Number of respondents who reported a year or ' never'	Respondents' median year	Respondents' gender (% males)	Respondents' mean age	Respondents' speeding violations	Respondents' non-speeding violations	Road traffic death rate per population	GDP per capita	Country' s educational performance	Country' s life expectancy	Motor vehicle density per population	Country' s median age
ARE	34	2025.0	67.6	29.7	-0.60	1.14	10.9	44507	78	77.1	285.5	30.3
ARG	326	2030.0	81.5	29.5	-0.97	0.16	13.6	14623	89	76.2	314.0	31.2
AUS	66	2033.8	58.0	36.3	-1.11	-0.70	5.4	67473	101	82.2	672.6	38.3
AUT	58	2030.0	57.1	32.7	-0.64	-1.66	5.4	50513	101	80.9	570.1	44.3
BEL	63	2030.0	70.2	34.4	0.38	-0.67	6.7	46927	100	80.4	541.2	43.1
BGD	122	2030.0	90.1	27.1	0.00	1.11	13.6	954	75	70.7	2.3	24.3
BGR	393	2035.0	64.7	33.6	-0.17	-0.23	8.3	7499	96	74.5	345.6	42.6
BIH	405	2050.0	75.5	30.2	1.20	0.38	17.7	4669	84	76.3	146.2	40.8
BRA	480	2030.0	78.1	29.1	-0.33	-0.35	23.4	11711	84	73.9	184.2	30.7
CAN	585	2029.5	39.6	36.8	0.32	-1.45	6.0	52305	102	81.4	595.4	41.7
CHL	89	2030.0	79.6	27.4	-0.73	0.27	12.4	15742	89	79.8	158.2	33.3
COL	181	2030.0	80.0	27.9	-0.49	0.29	16.8	8028	80	74.0	62.3	28.9
CZE	77	2036.7	73.9	28.4	0.74	-1.33	6.1	19858	100	78.3	459.7	40.9

DEU	335	2030.0	76.3	34.2	1.03	-1.27	4.3	46255	99	81.0	556.3	46.1
DNK	33	2030.0	65.1	35.8			3.5	59819	99	80.3	459.7	41.6
DOM	36	2036.5	84.2	27.8	-1.55	-0.66	29.3	5952	81	73.5	113.7	27.1
DZA	126	2050.0	92.9	29.2	0.54	1.57	23.8	5504	77	71.0	99.9	27.3
EGY	401	2040.0	84.0	27.0	0.33	2.12	12.8	3104	84	71.1	40.0	25.1
ESP	654	2030.0	72.1	32.9	-0.47	-0.78	3.7	29881	98	82.4	585.5	41.6
FIN	55	2039.0	80.2	31.9	0.84	-0.33	4.8	49310	103	80.8	538.8	43.2
FRA	172	2030.0	71.3	33.5	-0.05	-0.31	5.1	42628	100	82.0	594.6	40.9
GBR	611	2030.0	48.6	38.1	-0.32	-1.46	2.9	41777	102	81.0	517.3	40.4
GRC	293	2036.7	72.1	33.8	0.71	0.17	9.1	21966	97	80.6	537.8	43.5
HKG	47	2030.0	56.6	29.7	-1.99	-1.17	1.8	38353	106	83.8	72.9	43.2
HRV	204	2037.1	69.2	31.9	1.81	0.32	9.2	13598	90	77.1	366.7	42.1
HUN	160	2030.0	76.8	31.2	-1.15	-1.26	7.7	13487	100	75.3	354.8	41.1
IDN	443	2030.0	79.9	30.7	-0.14	1.50	15.3	3624	86	70.8	72.6	29.2
IND	1381	2030.0	80.9	29.3	-0.16	0.76	16.6	1455	81	66.5	14.9	27.0
IRL	39	2030.0	52.5	33.9	-1.63	-1.52	4.1	50470	98	81.0	491.6	35.7
ISR	34	2035.9	77.1	32.5			3.6	36051	96	82.1	302.1	29.9
ITA	524	2030.0	58.8	35.3	-0.04	-0.23	6.1	35477	101	82.3	671.2	44.5
LKA	47	2025.0	89.1	28.8	-0.22	1.00	17.4	3281	78	74.2	43.9	31.8
LTU	46	2030.0	66.0	29.2	0.30	0.26	10.6	15689	94	74.2	497.7	41.2
LVA	29	2037.0	67.3	32.9			10.0	15357	98	74.0	397.7	41.4
MAR	106	2042.5	91.7	28.0	-1.37	0.40	20.8	3056	77	70.9	62.8	28.1
MDA	44	2040.0	73.9	31.7	-0.05	0.10	12.5	2244	94	68.8	119.7	35.7
MEX	373	2030.0	74.0	28.5	0.07	0.32	12.3	10201	85	77.4	238.0	27.3
MKD	156	2040.0	69.2	31.8	1.03	0.32	9.4	5195	88	75.2	140.5	36.8
MYS	162	2030.0	65.7	32.3	1.10	0.95	24.0	10628	97	75.0	312.4	27.7
NGA	37	2030.0	88.1	28.7	0.74	2.14	20.5	2980	75	52.5	31.0	18.2
NLD	115	2029.0	67.1	37.4	0.83	-1.54	3.4	50793	102	81.1	510.7	42.1
NPL	39	2030.0	93.3	24.3	-0.87	0.08	17.0	691	77	68.4	5.0	22.9
NZL	28	2029.4	40.0	36.3			6.0	42409	101	81.4	713.4	37.6
PAK	216	2030.0	87.6	28.4	-0.93	1.37	14.2	1282	83	66.6	14.3	22.6
PER	128	2030.0	85.1	27.2	-0.94	0.08	13.9	6621	81	74.8	62.3	27.0
PHL	460	2027.5	54.0	30.6	-2.29	-0.80	10.5	2788	85	68.7	32.1	23.5
POL	261	2035.0	74.4	30.2	1.48	-0.17	10.3	13829	99	76.8	441.0	39.5
PRT	440	2030.0	72.2	31.0	1.40	-0.61	7.8	21508	95	80.4	508.0	41.1
ROU	412	2040.0	73.4	32.6	-0.01	0.13	8.7	9490	93	74.5	198.9	39.8
RUS	465	2040.0	65.4	35.9	-0.14	0.08	18.9	14487	99	71.1	236.2	38.9
SAU	42	2024.2	82.6	33.3	-0.30	1.08	27.4	24646	82	75.7	192.0	26.4
SGP	35	2030.0	87.8	29.3	-0.82	-1.23	3.6	55980	107	82.3	146.7	33.8
SRB	647	2044.0	65.4	32.8	0.08	-0.42	7.7	6354	91	75.1	238.3	41.9
SVK	47	2032.5	73.6	29.4	1.16	-0.17	6.6	18050	99	76.3	309.0	39.2
SVN	34	2032.5	66.7	33.6	3.53	-0.40	6.4	23297	99	80.3	536.1	43.5
SWE	54	2030.0	72.4	35.8	0.24	-1.48	2.8	60365	101	81.7	517.2	41.2
TUN	98	2040.0	86.8	29.1	0.32	1.63	24.4	4317	85	73.6	110.8	31.4
TUR	372	2029.8	75.9	30.9	0.13	1.49	8.9	10975	88	75.2	126.6	29.6
UKR	348	2040.5	69.3	34.3	-0.97	-0.54	13.5	4030	92	71.2	151.3	40.6

URY	29	2030.0	62.5	29.1			16.6	16879	92	77.1	193.2	34.3
USA	1348	2026.5	43.3	36.3	0.84	-0.46	10.6	52980	100	78.8	809.3	37.6
VEN	2956	2040.0	72.7	29.8	-0.33	0.13	37.2	12265	85	74.6	147.0	26.9
VNM	174	2035.0	74.0	25.9	0.62	1.82	24.5	1909	95	75.8	13.0	29.2

Note. Country abbreviations are according to ISO 3166-1 488 alpha-3.

6.7 References

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., & Rus, D. (2017). On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proceedings of *the National Academy of Sciences*, *114*, 462–467. https://doi.org/10.1073/pnas.1611675114

Bansal, P., Kockelman, K. M., & Singh, A. (2016). Assessing public opinions of and interest in new vehicle technologies: An Austin perspective. *Transportation Research Part C: Emerging Technologies*, *67*, 1–14. https://doi.org/10.1016/j.trc.2016.01.019

Bazilinskyy, P., & De Winter, J. (2015). Auditory interfaces in automated driving: an international survey. *PeerJ Computer Science*, *1*, e13. https://doi.org/10.7717/peerj-cs.13

Bazilinskyy, P., & De Winter, J. C. F. (2017). Analyzing crowdsourced ratings of speech-based take-over requests for automated driving. *Applied Ergonomics*, *64*, 56–64. https://doi.org/10.1016/j.apergo.2017.05.001

Bazilinskyy, P., & De Winter, J. C. F. (in press). Crowdsourced measurement of reaction times to audiovisual stimuli with various degrees of asynchrony. *Human Factors*. https://doi.org/10.1177/0018720818787126

Bazilinskyy, P., Petermeijer, S. M., Petrovych, V., Dodou, D., & De Winter, J. C. F. (2018). Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays. *Transportation Research Part F: Traffic Psychology and Behaviour*, *56*, 82–98. https://doi.org/10.1016/j.trf.2018.04.001

BMW News (2016, July 6). Autonomous cars to be in production by 2021. Available at https://news.bmw.co.uk/article/autonomous-cars-to-be-in-production-by-2021

BMW Group (2018). The BMW vision iNext. Future focused. Available from https://www.bmwgroup.com/BMW-Vision-iNEXT

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*, 1573–1576. https://doi.org/10.1126/science.aaf2654

Central Intelligence Agency (2015). The world factbook. Available from https://www.cia.gov/library/publications/the-world-factbook/fields/2177.html

Continental AG (2016). Facts and figures 2016. Available from http://www.continental-automotive.com/www/download/automotive_de_en/general/chassis/download/download_daten_fakten_2015_en.pdf

Continental AG (2018, September 11). Controlled by electronics: Continental launched its first driverless vehicle 50 years ago. Available from https://www.continental-corporation.com/en/press/press-releases/controlled-by-electronics-144932

De Winter, J. C. F. (2013). Predicting self-reported violations among novice

license drivers using pre-license simulator measures. *Accident Analysis & Prevention*, *52*, 71–79. https://doi.org/10.1016/j.aap.2012.12.018

De Winter, J. C. F., & Dodou, D. (2016). National correlates of self-reported traffic violations across 41 countries. *Personality and Individual Differences*, *98*, 145–152. https://doi.org/10.1016/j.paid.2016.03.091

De Winter, J. C. F., & Hancock, P. A. (2015). Reflections on the 1951 Fitts list: Do humans believe now that machines surpass them? In T. Ahram, W. Karwowski, & D. Schmorrow (Eds.), Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Las Vegas, NV, 5334–5341. https://doi.org/10.1016/j.promfg.2015.07.641

De Winter, J. C. F., Kyriakidis, M., Dodou, D., & Happee, R. (2015). Using CrowdFlower to study the relationship between self-reported violations and traffic accidents? In T. Ahram, W. Karwowski, & D. Schmorrow (Eds.), Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Las Vegas, NV, 2518–2525. https://doi.org/10.1016/j.promfg.2015.07.514

Department of Motor Vehicles (2017). Autonomous vehicle disengagement reports 2016. Available at https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/disengagement_report_2016

Department of Motor Vehicles (2018). Autonomous vehicle disengagement reports 2017. Available at https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/disengagement_report_2017

Dixit, V. V., Chand, S., & Nair, D. J. (2016). Autonomous vehicles: disengagements, accidents and reaction times. *PLOS ONE*, *11*, e0168054. https://doi.org/10.1371/journal.pone.0168054

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, *77*, 167–181. https://doi.org/10.1016/j.tra.2015.04.003

Ford Motor Company (2018). Looking further. Ford will have a fully autonomous vehicle in operation by 2021. Available at https://corporate.ford.com/innovation/autonomous-2021.html

Geddes, N. B. (1940). *Magic motorways.* Random House.

Geiger, A., Lauer, M., Wojek, C., Stiller, C., & Urtasun, R. (2014). 3d traffic scene understanding from movable platforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*, 1012–1025.

Kovácsová, N., De Winter, J. C. F., & Hagenzieker, M. P. (2017). Cyclists' predictions of what a car driver will do next at intersections. In L. Aultman-Hall, R. Buehler, & J. Chapman (Eds.), Proceedings of *the 6th Annual International Cycling Safety Conference (ICSC)*, Davis, CA.

Kröger, L., Kuhnimhof, T., & Trommer, S. (in press). Does context matter? A comparative study modelling autonomous vehicle impact on travel behaviour for Germany and the USA. *Transportation Research Part A: Policy and Practice*. https://doi.org/10.1016/j.tra.2018.03.033

Kurzweil, R. (1990). *The age of intelligent machines.* Cambridge: MIT Press.

Kurzweil, R. (2012). How to create a mind: The secret of human thought revealed. Penguin.

Kyriakidis, M., Happee, R., & De Winter, J. C. F. (2015). Public opinion on

automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour*, *32*, 127–140. https://doi.org/10.1016/j.trf.2015.04.014

Litman, T. (2018). *Autonomous vehicle implementation predictions.* Victoria Transport Policy Institute. Available at https://www.vtpi.org/avip.pdf

Menton, R. K. (1957). *Social theory and social structure.* Glencoe, IL: Free Press.

Milakis, D., Snelder, M., Van Arem, B., Van Wee, B., & Correia, G. (2017). Development of automated vehicles in the Netherlands: scenarios for 2030 and 2050. *European Journal of Transport and Infrastructure Research, 17,* 63–85.

Nissan Motor Corporation (2017, December). Autonomous drive technology roadmap. Available at https://www.nissan-global.com/EN/TECHNOLOGY/OVERVIEW/autonomous_drive.html

Nissan Motor Corporation (2018, March 23). Nissan aims to sell 1 million electrified vehicles a year by FY2022. Available at https://newsroom.nissan-global.com/releases/release-487297034c80023008bd9722aa05f858-180323-01-e

NissanNews.com (2013, August 27). Nissan announces unprecedented autonomous drive benchmarks. Available at http://nissannews.com/en-US/nissan/usa/releases/nissan-announces-unprecedented-autonomous-drive-benchmarks

Ofek, E., & Richardson, M. (2003). Dotcom mania: The rise and fall of internet stock prices. *The Journal of Finance*, *58*, 1113–1137.

Ohn-Bar, E., & Trivedi, M. M. (2016). Looking at humans in the age of self-driving and highly automated vehicles. *IEEE Transactions on Intelligent Vehicles*, *1*, 90–104. https://doi.org/10.1109/TIV.2016.2571067

Piao, J., McDonald, M., Hounsell, N., Graindorge, M., Graindorge, T., & Malhene, N. (2016). Public views towards implementation of automated vehicles in urban areas. *Transportation Research Procedia*, *14*, 2168–2177. https://doi.org/10.1016/j.trpro.2016.05.232

Pollet, T. V., Tybur, J. M., Frankenhuis, W. E., & Rickard, I. J. (2014). What can cross-cultural correlations teach us about human nature? *Human Nature*, *25*, 410–429. https://doi.org/10.1007/s12110-014-9206-3

Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality: Published for the European Association of Personality Psychology*, *21*, 667–706.

Rodríguez Palmeiro, A., Van der Kint, S., Hagenzieker, M., Van Schagen, I. N. L. G., & De Winter, J. C. F. (2018). Cyclists' expectations when encountering automated vehicles: results of an international photo-based questionnaire. Paper presented at *the 7th International Cycling Safety Conference*, Barcelona, Spain.

Sage, A., & Lienert, P. (2016, August 16). Ford plans self-driving car for ride share fleets in 2021. Available at http://www.reuters.com/article/us-ford-autonomous-idUSKCN10R1G1

Schoettle, B., & Sivak, M. (2014). A survey of public opinion about autonomous and selfdriving vehicles in the US, the UK, and Australia. Michigan, USA. University of Michigan, Transportation Research Institute (UMTRI).

Shladover, S. E. (2016). The truth about "self-driving" cars. *Scientific*

American, *314*, 52–57. https://doi.org/10.1038/scientificamerican0616-52
Sierhuis, M. (2016, November 3). The future is now: Self-driving vehicles are a reality. Available at https://livestream.com/lindahall/Self-Driving-Cars
Underwood, S. (2014). Automated, connected, and electric vehicle systems: Expert forecast and roadmap for sustainable transportation. *Ann Arbor, MI: University of Michigan, Graham institute for Sustainability.*
Waymo (2018). On the road. Available at https://waymo.com/ontheroad/
World Bank (2015). Indicators [Data]. Available at http://data.worldbank.org/indicator
World Health Organization (2015). WHO global status report on road safety 2015. Geneva, Switzerland: World Health Organization. Available at http://apps.who.int/iris/bitstream/10665/189242/1/9789241565066_eng.pdf?ua=1
Zuby, D. (2016). Crash avoidance technologies: Assessing the building blocks for tomorrow's driverless vehicles. Presentation at *the I-95 Corridor Coalition Connected & Automated Vehicles Conference: What States Need to Know*, Linthicum, MD. Available at http://i95coalition.org/wp-content/uploads/2016/03/I95-Corridor-Conf-Zuby_June-21_-2106.pdf?fe2c99

7 CONCLUSIONS AND RECOMMENDATIONS

The aim of this thesis was to discover how the auditory modality should be used in highly and fully automated cars, and to make a contribution towards the development of design guidelines. The main focus of this work was on the use of discrete auditory and multimodal displays for supporting TORs during highly automated driving. Additionally, the use of continuous sound for situation and mode awareness was examined.

This chapter summarises the main conclusions and recommendations of the thesis. The text is organised along Chapters 2–6 of the thesis, and in Section 7.5 special attention is given to the role of crowdsourcing in Human Factors research, whereas in Section 7.6 the results in the thesis are viewed in the scope of automated driving in general.

7.1 State of the art

The creation of auditory feedback for in-vehicle interfaces is challenging because of strict design criteria and the difficulty of communicating product requirements. There is a lack of a structured process for supporting the creation of such assets. Section 2.1 showed the benefits of documenting the process and outlining clear design criteria. The study was hosted at one of the major suppliers of the industry (Continental AG) and demonstrated a practical solution (i.e., a newly developed software tool) to address the problem of the lack of a structured sound design process.

Already 20 years ago, researchers emphasized the importance of communicating the status of the automation via in-vehicle feedback (Stanton, Young, & McCaulder, 1997). Still, even carefully designed feedback may be perceived as annoying or useless by the public. The opinion of people who actually drive cars must be taken into account when designing displays for cars, including cars that will be driving on the streets in the future. I learned that the public expresses a mixed opinion on the use of auditory feedback both in modern and future automated cars (see Sections 2.2 and 2.3).

From the surveys, it became clear that the overall most preferred way to support a take-over request (TOR) is an auditory instruction in the form of a female voice. It is beneficial to have a message spoken in an accent that corresponds to the participants' country. Moreover, peoples' preference depends on the urgency of the situation: (1) for high-urgency situations, multimodal warnings are the most preferred option; (2) for low-urgency situations, auditory messages are the most preferred option.

Furthermore, by means of an online survey I showed that the public has a split opinion about the idea of the fully automated driving. Section 2.4 showed that a large percentage of the population is positive and a significant number of people are negative towards the idea. A large percentage of the population would not trust an automated car and a significant number of people, if given a chance to drive such a vehicle, would still prefer to have manual control. These findings may have important implications for the design of human-machine

interfaces in cars, as it is possible that users may reject support/feedback even if it is carefully designed.

7.2 Auditory feedback for supporting takeover requests during highly automated driving

The above studies on auditory feedback (Sections 2.2 and 2.3) demonstrated the potential of sound for supporting TORs. Studies have shown that reaction times are fastest when a visual and auditory stimulus are generated at once instead of being presented with a temporal asynchrony (Diederich & Colonius, 2004). These findings were confirmed in this thesis using an online reaction-time study (Section 3.1). An increase of beep rate yields an increase of self-reported urgency.

Section 3.2 shows that the use of multimodal TORs is promising. Multimodal feedback increases the redundancy of the warning and consequently reduces the probability of misses. The use of multimodal TORs consisting of auditory non-speech feedback (beeps) and vibrotactile feedback (tactile seat) resulted in faster steer-touch times than the unimodal vibrotactile take-over request. No statistically significant differences could be found for lane change times and brake times. The usefulness of directional multimodal feedback (i.e., feedback provided on the left or right side of the driver, to indicate the direction of a hazard) to uninstructed drivers is questionable and needs to be researched further. Furthermore, I showed that auditory TORs in the form of beeps-only are regarded as useful but not satisfactory.

As mentioned in the Introduction, speech-based feedback is an alternative for designing a TOR. By means of an online study described in Section 3.2, it was shown that an increase in the speech rate results in an increase of perceived urgency and commandingness of such a TOR, which links well to the findings on the urgency of beep-based feedback presented in Section 3.1. In this study, I disputed the finding in Hellier et al. (2002) and showed that the female voice is preferred over a male voice when there is a high level of background noise.

Chapters 2 and 3 demonstrated that discrete speech-based and artificial auditory feedback is useful for conveying information during highly automated driving. In Section 2.3 it was suggested that auditory-only information should be used in non-urgent situations, when the feedback should be most of all pleasant and serve a notification purpose. The studies in Chapters 2 and 3 showcase the power of multimodal feedback: the use of such feedback gives a possibility to unite the advantages of its components (i.e. auditory, visual and/or vibrotactile signals) and avoid their drawbacks. Multimodal feedback is especially valid in urgent situations.

7.3 Auditory feedback for situation awareness during highly automated driving

It is crucial for people in an automated vehicle to be aware of the situation around the vehicle. Driving mostly relies on vision, but visual information from the environment is not always accessible or it is degraded. In such cases, for example during lane-keeping, the auditory spectrum may be used to provide the required information. People have the ability to accurately interpret artificial sounds on an artificial distance scale, if the sounds are presented by one of three

sonification methods: Beep Repetition Rate, Sound Intensity, and Sound Fundamental Frequency (Section 4.1).

Auditory feedback is not always discrete. In fact, many cues in the world around us are of continuous nature. Hence, continuous feedback may be perceived by people as a natural way to receive information. The studies in Sections 4.2–4.4 demonstrated that, if presented with a continuous tone representing a linear function of the predicted lateral error of the car from the lane centre 1 s, 2 s, or 3 s into the future, with sufficient training and regulated speed of the vehicle, it is possible to keep the car within the lane boundaries. It implies the potential of the use of sound in situations when visual information is unavailable during both manual and automated driving. It is still to be investigated if people can drive a car without any visual feedback when the speed is controlled manually.

7.4 Continuous auditory feedback for displaying automation status, lane deviation, and headway

While driving in an automated vehicle, one needs to be aware of the mode of the vehicle at all times. In Chapter 5 I described the use of continuous auditory on the status of adaptive cruise control (ACC), lane offset, and headway. The concept, which was tested in a real truck, was designed in such a way that the sounds blended with the sound of engine and wind noise that are already present in the cabin. The use of ACC allowed to mimic a highly automated truck. The results showed that the truck drivers are not favourable towards adding additional continuous feedback to the cabin. Even though truck drivers are normally quite sensitive to the way their trucks sound, the idea of adding any additional noise, even such that conveys useful information, was not appreciated. However, the continuous feedback on the lane-offset received positive feedback on its usefulness, and I recommend testing it in future experiments.

Chapters 4 and 5 demonstrated the use of continuous feedback during manual and assisted driving. Such feedback is promising, but it requires considerable attention during the design stage, in comparison to the amount of development required for discrete auditory feedback. And, not all versions may be appreciated by the end user, as was shown in Chapter 5. However, the intuitiveness (i.e., ease of understanding) of continuous feedback, combined with the possibility to blend such feedback into an already existing environment, allows conveying information in an effective (though potentially unpleasant) manner.

7.5 When will most cars be able to drive fully automatically?

Both academia and industry are spending considerable efforts on research of various aspects of automated driving. It is important to understand approximately when these efforts will be applied in consumer-ready vehicles that we or our (grand)children will be able to buy. In Chapter 6, the crowdsourced public expressed optimism towards the amount of time required for automated cars to become widespread. When 17,360 people from 129 countries were asked when they think that most cars will be able to drive fully automatically ('level 5 automation'; SAE International, 2014) in their country of residence, the median reported year was 2030. This opinion differs from what experts currently

believe, where most mention years that are much further from us. It may cause false expectations and frustration. It is important that industry and academia acknowledge the possibility that the public holds unrealistically optimistic predictions towards automated vehicles. Here, important lessons can be learned from the 'dot-com bubble' in the period 1997–2001, which caused a great number of jobs and billions of losses. One can make a clear parallel with Internet: the Internet is now ubiquitous, but it was overhyped in the beginning. Fully automated driving will come eventually, but society should not overhype it.

7.6 Crowdsourcing as a tool for Human Factors research

Most Human Factors research is conducted with small sample sizes comprised of participants belonging to the same age group and having a similar background. Crowdsourcing is promising for gaining statistically credible findings with large sample sizes (Fortenbaugh et al., 2015). Crowdsourcing may be particularly useful when no special equipment is required to conduct an experiment, which is mostly the case in usability research. Such method for conducting research requires an order of magnitude smaller financial contribution, compared to a traditional laboratory study. However, care is needed in designing a crowdsourced experiment. A limitation of crowdsourcing is that the participant pool is restricted by the people that are registered for a certain service, where crowdsourced tasks can be executed. It is argued that the benefits of the method outweigh its limitations, and most future Human factors research should be conducted by means of crowdsourcing.

7.7 Significance of present thesis for automated driving

Automated driving will change our world forever. Inevitably such cars will be introduced. However, the current thesis addressed only a small fraction of a larger problem. The focus of this thesis was on developing HMIs for supporting the human inside of the automated car. Future research should examine how humans outside of automated cars (pedestrians, cyclists) should be supported, to ensure that the entire traffic system will benefit. Society is entering an era where the human is central: the future human road user needs to control, cooperate with, and sometimes listen to the machine. In the end, humans should be relieved from the necessity of controlling a vehicle manually. If we remove the human factor (error) from the task of driving, we should be able to save the lives of millions of people and make our world safer and more predictable.

7.8 References

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*, 1388–1404. https://doi.org/10.3758/BF03195006

Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman, M. (2015). Sustained attention across the life span in a sample of 10,000: dissociating ability and strategy. *Psychological Science*, *26*, 1497–1510. https://doi.org/10.1177/0956797615594896

Hellier, E., Edworthy, J., Weedon, B., Walters, K., & Adams, A. (2002). The perceived urgency of speech warnings: Semantics versus acoustics. *Human Factors*, *44*, 1–17.

SAE International. (2014). *Taxonomy and definitions for terms related to onroad motor vehicle automated driving systems (Standard No. J3016)*. Available at http://standards.sae.org/j3016_201401

Stanton, N. A, Young, M. S., McCaulder, B. (1997). Drive-by-wire: The case of driver workload and reclaiming control with adaptive cruise control. *Safety Science,* 27, 149–159. https://doi.org/10.1016/S0925-7535(97)00054-4

8 PROPOSITIONS

These propositions are regarded as opposable and defendable, and have been approved as such by the promotors Dr.ir. J. C. F. De Winter and Prof.dr. F. C. T. Van der Helm.

1. The urgency of the traffic situation is the most important parameter in the design of the take-over request.
2. If fast reaction time is required, a warning should be multimodal.
3. Auditory feedback needs to be so good that a blindfolded person could keep the car on a curved road using such feedback only.
4. Directional auditory and vibrotactile take-over requests are not useful.
5. Continuous in-vehicle auditory feedback is less annoying than discrete auditory feedback.
6. People have more optimistic expectations of fully automated driving than the industry and academia.
7. Results from crowdsourcing Human Factors research are more valid than from lab-based research.
8. In future car driving the use of sound will be purely entertainment and notification.
9. Given the complexity of automated driving, sharing data is an absolute necessity.
10. The lack of structured sound design process is responsible for many casualties in daily traffic.
11. Love is life. Everything that can be understood, is understood only because of love (Tolstoy, 1873).
12. This world is learned through sound, not through visuals (Bhaktivedanta, 1968).
13. The world is a book and those who do not travel only read one page (Augustine of Hippo, c. 380).
14. If our brains were simple enough for us to understand them, we'd be so simple that we couldn't (Stewart & Cohen, 2000).

References

Augustine of Hippo (c. 380).

Bhaktivedanta, A. C., & Prabhupada, S. (1972). Bhagavad-Gita as it is. Bhaktivedanta Book Trust.

Stewart, I., & Cohen, J. (2000). The collapse of chaos: Discovering simplicity in a complex world.

Tolstoy, L., Pevear, R., & Volokhonsky, L. (2007). War and peace. London: Vintage.

ACKNOWLEDGEMENTS

I would like to thank a number of people for their contributions to this thesis.

Major credit for his contribution to this thesis is directed to Joost De Winter. I was very lucky to have him as my supervisor. His impressive availability and eagerness to support me with all of my new ideas allowed me to open my hands and feel the taste of true science, where each step taken is aimed to advance the society just a little further.

It has been a pleasure working with Riender Happee. Thank you very much for your valuable feedback and contributions. I hope we will be able to cooperate more in the future. Of course, a very special mention goes to Dimitra Dodou. Thank you very much for all your time that you spent on reviewing my drafts. You have a talent to view my research at a new angle, and I am deeply grateful for all the insights that you have given me. Prof. Frans van der Helm, my promoter, thank you very much for your time that you spent on my work. Even though some of the studies that I have worked on do not directly link to your own research, I was always impressed by the usefulness of your feedback and your critical views on my ideas. Big thank you to the secretaries of the BioMechanical Engineering Department and Cognitive Robotics Department, Nancy Kouters, Sabrina Ramos Rodrigues, Hanneke Hustinx, and Karin van Tongeren for providing great support. Also, my appreciation goes to Mascha Toppenberg for being the head of the Research council, that I was happy to be a part of.

The work in this thesis was conducted within the HFauto project. I am grateful to all my colleagues in the project, especially members of my work package, Bastiaan Petermeijer, Alexander Eriksson, and Veronika Petrovych, and Christopher Cabrall, Zhenji Lu, and Silvia Varotto, who were located in Delft together with me. My work could not be possible with direct or indirect contributions from the members of the team hosted in different corners of Europe: Daniel Heikoop, Joel Gonçalves, Alberto Morando, Bo Zhang, Matt Sassman, and Ignacio Solis. Thank you, Miltos Kyriakidis for being with us for the first part of the project and sharing the office with me at the start of my time in Delft. Even though we were all over the place and would meet once or twice per year, HFauto felt like a family to me. I would also like to express my gratitude to Klaus Bengler, who supported my secondment at TU München. Thank you to Jan Andersson, Thierry Bellet, and Marco Dozza, with whom we exchanged impressive ideas, which sadly never materialized in secondment opportunities. I am sure we'll find ways to stay in touch and collaborate in the future!

I had synergistic cooperation with my colleagues at TU Delft. Thank you, Prof. Dariu Gavrila for your leadership and organization of research in the group. My appreciation goes to Julian Kooij, who was open to collaborate and exchange ideas. I wish to say 'thank you' to Oliver Lee, Natalia Kovacsova, Jork Stapel, Felix Dreger, Ronald Ensing, Frank Everdij, Joris Domhof, Thomas Hehn, Tugrul Irmak, András Pálffy, and Ewoud Pool. Also, my great appreciation goes to Peter van Leeuwen for helping me decipher the documentation of the driving

simulator. My special gratitude goes to George Dialynas and Sreekar Reddy for your great support and understanding during the final stage.

Towards the end of my PhD, I was given a very special opportunity by Barys Shyrokau to visit Virginia Tech Transportation Institute. I would like to express my gratitude to Jeffrey Hickman, who introduced me to two great projects at VTTI and spent his precious time on supervising me during my short stay in Blacksburg. I hope to continue our collaboration during my post-PhD life!

I was fortunate enough to be given possibilities to look at the other side – the industry. My stay at Volvo Trucks could not be possible without hours and hours of negotiations between the company represented by Emma Johansson and Ann-Sofi Karlsson and TU Delft. Thank you, Pontus Larsson, for being my guru in Göteborg, you are a rock star! Big thank you to Bernhard Westerhof, Jeroen Roozendaal, Wouter van Unen (and his navy green camping van), and Niels Berger for sharing coffee breaks with me and making me feel just like at home. The time in Sweden escalated quickly, thanks to you! Then, I had great three months at Continental AG, which could not have become possible without dedication and support of Stephan Cieler. Thank you very much for your guidance and the great lunch runs we had! Big thank you to Jörg Witthaus for supervision and Markus Groh for very professional technical support. It was my great pleasure to be hosted both in Volvo Trucks and Continental AG. Also, I wish to express my enormous gratitude to Guido, Frank, and Thom from SD-insights for your openness to work together. I really look forward to our cooperation in the future!

The funding for my position came from the Marie Skłodowska-Curie Actions programme. I became an active members of the Marie Curie Alumni Association (MCAA), where I had a pleasure to work closely with Marco Masia, Yaroslav Vverkh, Murat Güneş, Natalia Balcázar, Oliver Diaz, Rohan Soman, Maria Górna, Brian Cahill, Esther Volz, Sajid Mohamed and many other extremely talented people. Thank you for appreciating my efforts to promote the cooperation between science and academia and help with lighting sparkles in the eyes of the talented youth of Europe.

I had the privilege of advising two MSc students: Tom Driessen, who focuses on the human-machine cooperation, and Mounir El Hassnaoui, who is looking at the situation assessment. They are both busy with finishing their studies. I wish them the best of luck, when I worked with them, they have shown great potential and dedication. And I hope to continue to collaborate with them.

It was my great pleasure to work with a number of BSc students during the last four years. I would like to express my deep gratitude to Wessel van Haarlem, Hashim Quraishi, Coen Berssenbrugge, Jasper Binda for the contribution to the work presented in Section 4.1. Then, I am grateful to Lars van der Geest, Stefan van Leeuwen, Bart Numan, Joris Pijnacker for their belief in the idea of blind driving, their work is presented in Section 4.2. Four enthusiastic students Charles Beaumont, Xander van der Geest, Reinier de Jonge, and Koen van der Kroft then continued the work, and their impressive results are depicted in Section 4.3. The last iteration of the line of studies on the idea of blind driving could not have happened without hard work of Laurens Bijker, Tim Dielissen, Shin French, Tom Mooijman, and Luka Peters, to whom I express my deepest gratitude; their work contributed to Section 4.4. Additionally, I had a great pleasure to work with Arne van der Aa, Michael Schoustra, John Spruit, Laurens Staats, Klaas Jan van der Vlist, Niels Heisterkamp, Philine Luik, Stijn Klevering,

Assia Haddou, Michiel Zult, Coert de Koning, Hidde Lingmont, Tjebbe de Lint, Twan van der Sijs, Florian van den Ouden, Amir Al Jawahiri, Pieter Kapel, Joppe Mulckhuyse, and Sjors Wagenaar.

My journey through the PhD study would not have been the same without my dearest friend Roberto Merino, who I met during a week of carving marble on a Tuscan hill. Thank you for pointing me to great opportunities, joining my endeavours and sharing board game evenings with me. Also, big thank you to Ingmar van Hengel, whom I met in the hallway so often, but with whom it was so hard to catch up. Of course, Lars Kooijman, a real tovarish, that joined our office towards the end of my journey, but with whom I had enough time to have an exchange of the best of the cultures of our countries. And, Kenko-san, let's talk about opening an onigiri-making business after my defence.

I would like to say 'thank you' from the bottom of my heart to Pradeep. Thank you for being an inspiration on my journey to find myself and willingness to assist me with learning the art of meditation. Especially, the last six months showed me the importance of self-transcendence and the beauty of simple life. Thank you, Sophia, for sharing so many calm moments of beauty with me and reminding me that people can be beautiful inside. Also, thank you, Evgenii and Eduard, for highlighting how slow of a runner I was and pushing me to go faster and never give up. Thank you, Sara, Fahad, Koos, Aldo for baring with me when I was trying to finish my thesis during our travels. Bart, Karen, Long, Josefien, Bart, thank you for the great atmosphere for meditation. Thank you, Kim, Gourav, and Rajat, for your words of wisdom and expression of calmness, which you directed into my heart. Aart and Hanna, it's a little hard to catch you, but I am very happy you are present in my life. Arun, Anne, Roma, thank you for sharing some of the happiest moments with me. Thank you, Elizabeth, for our chats in Dutch. Dank je, Jerry, for your wonderful music. Bedankt, Kyra and Yasmin, for being my great friends in Ireland and pointing me in the direction of the Netherlands, that has now become my second home. Thank you, Oleksii, Lesya and Timka, for being close to me and showing me your support, when it was most needed. Спасибо, Lena and Oleg. Thank you, Taylor, for coming to Delft at the right moment and sharing your evenings with me. Дякую, Taras and Myroslava, for making me feel a little bit closer to home. Спасибо, о. Петр, за Ваше время и понимание.

Огромное спасибо всем моим родным и близким. Без вас я – ничто. Спасибо, бабушка, за твою самоотверженную поддержку на протяжении всей моей жизни. Мысли о твоей улыбке согревает мое сердце даже в самые тяжелые мгновения. Твоя забота и любовь дедушки – фундамент того, кем я есть сегодня. Мама, я стараюсь быть как ты – достойным человеком, который делает этот мир чуточку светлее. Спасибо тебе за все те бессонные ночи и полную поддержку, где бы я не был. Папа, огромное спасибо за то, что ты есть примером мужской силы для меня, я тебя люблю. Дедушка Саша, с теплотой в моем сердце я вспоминаю вечера, наполненные мудростью и улыбками за шахматной доской. Валентина, спасибо, что доверили мне самое дорогое в Вашей жизни и в этом мире; я старался как мог. Дашуля, светлая доброта и любовь в твоих бездонных глазах – моя мотивация. Ты наполнила мой мир чистейшим светом. Теплота твоей ручки наполнила мою жизнь любовью, которая никогда не перестает, хотя и пророчества прекратятся, и языки умолкнут, и знание упразднится.

CURRICULUM VITAE

Born in Chernihiv, Ukraine, 17 February 1991.

Education
2008–2012. BEng in Information Technology at Mikkeli University of Applied Sciences.
2012–2014. Erasmus Mundus Double MSc in Dependable Software Systems at University of St Andrews and Maynooth University.
2014–2018. PhD study at Delft University of Technology.

Work experience
2010–2011. Software developer at MHG Systems Oy.
2011–2012. Software developer at Cluetail.

PUBLICATIONS

Journal articles

Bazilinskyy, P., & De Winter, J. C. F. (2015). Auditory interfaces in automated driving: an international survey. *PeerJ Computer Science, 1*, e13.

Petermeijer, S. M., **Bazilinskyy, P.***, Bengler, K., & De Winter, J. C. F. (2017). Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics, 62*, 204–215.

Bazilinskyy, P., & De Winter, J. C. F. (2017). Analyzing crowdsourced ratings of speech-based take-over requests for automated driving. *Applied Ergonomics, 64*, 56–64.

Bazilinskyy, P., Petermeijer, S. M., Petrovych, V., Dodou, D., & De Winter, J. C. F. (2018). Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays. *Transportation Research Part F: Traffic Psychology and Behaviour, 56*, 82–98.

Bazilinskyy, P., & De Winter, J. C. F. (2018). Crowdsourced measurement of reaction times to audiovisual stimuli with various degrees of asynchrony. *Human Factors*.

Bazilinskyy, P., Larsson, P., Johansson, E., & De Winter, J. C. F. (2018). Continuous auditory feedback for the status of ACC, lane deviation, and time headway. Under review.

Bazilinskyy, P., Cieler, S., & De Winter, J. C. F. (2018). Sound design process for automotive industry. Under review.

Bazilinskyy, P., Kyriakidis, M., & De Winter, J. C. F. (2018). When will most cars drive fully automatically? An analysis of international surveys. Under review.

* Joint first author.

Conference proceedings

Jeon, M., **Bazilinskyy, P.**, Hammerschmidt, J., Hermann, T., Landry, S., & Wolf, K. E. (2015). Report on the in-vehicle auditory interactions workshop: taxonomy, challenges, and approaches. *Proceedings of AutomotiveUI*, Graz, Austria.

Bazilinskyy, P., Petermeijer, S. M., & De Winter, J. C. F. (2015). Use of auditory interfaces for takeover requests in highly automated driving: A proposed driving simulator study. Workshop on *In-Vehicle Auditory Interactions at the 21st International Conference on Auditory Display*, Berlin, Germany, 25–27.

Bazilinskyy, P., Kyriakidis, M., & De Winter, J. C. F. (2015). An international crowdsourcing study into people's statements on fully automated driving. Proceedings of *the 6th International Conference on Applied Human Factors and Ergonomics (AHFE),* Las Vegas, NV*,* 2534–2542.

Bazilinskyy, P., Geest, L. Van Der, Van Leeuwen, S., Numan, B., Pijnacker, J., & De Winter, J. C. F. (2016). Blind driving by means of auditory feedback.

Proceedings of *13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems*, Kyoto, Japan.

Bazilinskyy, P., Van Haarlem, W., Quraishi, H., Berssenbrugge, C., Binda, J., & De Winter, J. C. F. (2016). Sonifying the location of an object: A comparison of three methods. Proceedings of *13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems*, Kyoto, Japan.

Bazilinskyy, P., Kovácsová, N., Al Jawahiri, A., Kapel, P., Mulckhuyse, J., Wagenaar, S., & De Winter, J. C. F. (2016). Object-alignment performance in a head-mounted display versus a monitor. Proceedings of *IEEE International Conference on Systems, Man, and Cybernetics (SMC) 2016*, Budapest, Hungary.

Bazilinskyy, P., Eriksson, A., Petermeijer, S. M., & De Winter, J. C. F. (2017). Usefulness and satisfaction of take-over requests for highly automated driving. Proceedings of *Road Safety and Simulation conference (RSS) 2017*, The Hague, the Netherlands.

Bazilinskyy, P., Beaumont, C. J. A. M., Van der Geest, X. O. S., De Jonge, R. F., Van der Kroft, K., & De Winter, J. C. F. (2017). Blind driving by means of a steering-based predictor algorithm. Proceedings of *the 8th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Los Angeles, USA

Bazilinskyy, P., Van der Aa, A., Schoustra, M., Spruit, J., Staats, L., Van der Vlist, K. J., & De Winter, J. C. F. (2017). An auditory dataset of passing vehicles recorded with a smartphone. Proceedings of *Tools and Methods of Competitive Engineering (TMCE)*, Las Palmas de Gran Canaria, Spain.

Bazilinskyy, P., Heisterkamp, N., Luik, P., Klevering, S., Haddou, A., Zult, M., Dialynas, G., Dodou, D., & De Winter, J. C. F. (2018). Eye movements while cycling in GTA V. Proceedings of *Tools and Methods of Competitive Engineering (TMCE)*, Las Palmas de Gran Canaria, Spain.

Bazilinskyy, P., Bijker, L., Dielissen, T., French, S., Mooijman, T., Peters, L., Happee, R., Dodou, D., & De Winter, J. C. F. (2018). Blind driving by means of the predicted track angle error. Under review.