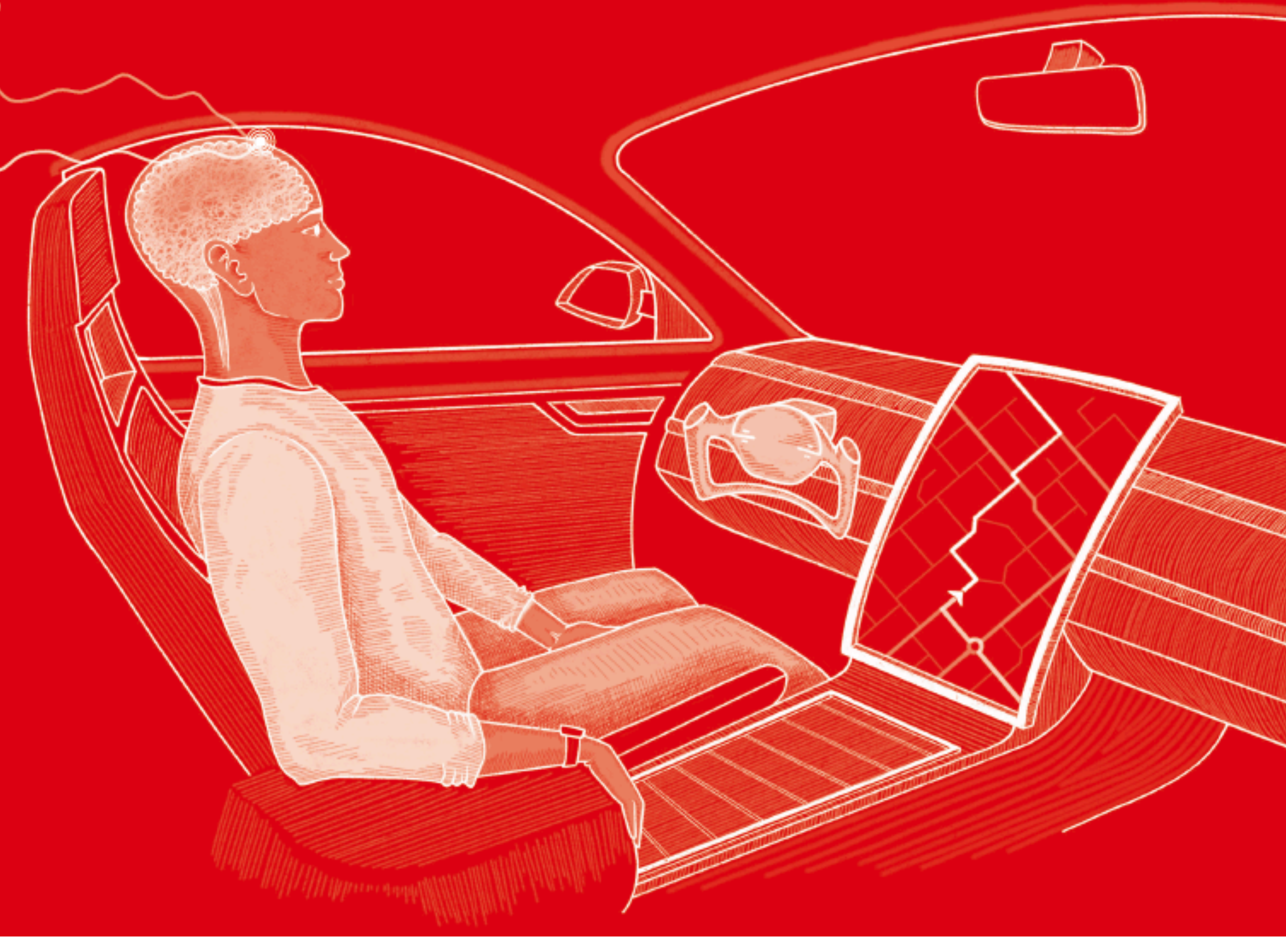
Remo M.A. van der Heiden

AUDITORY ALERTS AND DRIVING

SUSCEPTIBILITY TO AUDIO DURING (AUTOMATED) DRIVING



Cover design: Cropyright Communications – Kristel Mooij
Layout: Remo van der Heiden
Printed by: Ridderprint – www.ridderprint.nl

ISBN: 978-90-393-7287-6

Auditory alerts and driving

Susceptibility to audio during (automated) driving

Ontvankelijkheid voor auditieve informatie tijdens (geautomatiseerd) rijden

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 5 juni 2020 des middags te 2.30 uur

door

Remo Mattheus Arnoldus van der Heiden

geboren op 20 september 1987
te Zürich, Zwitserland

Promotor:
Prof. dr. J.L. Kenemans

Copromotoren:
Dr. C.P. Janssen
Dr. S.F. Donker

Dit proefschrift werd mede mogelijk gemaakt met financiële steun van Rijkswaterstaat - Ministerie van Infrastructuur en Waterstaat

“ Anyone who can drive safely
while kissing
is simply not giving the kiss
the attention it deserves “

Contents

Chapter 1

General introduction

We live in an era where automation is taking over many tasks (e.g., Dekker & Woods 2002; Hancock et al., 2013; Janssen, Donker, Brumby, & Kun, 2019). An example of this trend of automation can be seen in the automotive domain (e.g., Ayoub, Zhou, Bao & Yang, 2019; Janssen, Donker, Chuang, & Ju, 2019; Kun, 2018). More and more driving tasks can be, and are, performed by the car. Examples of systems that are already commercially available are: (adaptive) cruise control, auto steer, lane departure warning, and automatic parking. Such driving assistance systems can support, and partially take over, driving tasks. When several of these systems are combined, the car is able to drive without manual control of the human driver under specific (ideal) road conditions. Examples of such systems are Tesla's Autopilot, Cadillac's Super Cruise, and Nissan's Pro Pilot.

Despite these technological innovations, currently there is no system that is capable of operating fully without human supervision. In effect, current systems require human monitoring of the driving task and interaction between the human and the automated system. One of the main aspects of the human-automation interaction is communication between the system and the human driver, to share the system state and warn whenever intervention is needed (e.g., Walch, Lange, Baumann, & Weber, 2015; Janssen, Boyle, Kun, Ju, & Chuang, 2019). As sound is omnidirectional and can be heard irrespective of where a person's current visual focus is, it is broadly used in cars to communicate a system's state (Nees & Walker, 2011). On the road towards fully automated driving, where human interaction is minimal, there are currently already many driver assisting systems, such as lane departure warning systems, that use auditory signals to warn the human driver for a required intervention. This implies that the impact of such auditory signals on the person in the driver's seat is crucial.

In relation to such impact, throughout this thesis I use the term "susceptibility". In my studies "susceptibility" is operationalized using a specific Event-Related Potential in the electroencephalogram (explained in more detail later). I use the term susceptibility to refer to the extent to which an observer or process controller is in a mode which allows for detection of external signals to such a degree that an adequate behavioral response can be based on the detection. In effect, given the wide use of auditory signals in cars and other automated devices, being susceptible to, and acting on, warnings from the system is an essential prerequisite for a safe ride. This requirement comes on top of that of the susceptibility to unexpected but potentially relevant auditory ànd visual signals that is crucial for safe driving in the traditional sense (i.e., for manual driving). However, its special meaning in the context of automation can be illustrated by the following hypothetical situation.

1.1 A glimpse into the future

As an example of what could happen in the future, I would like to introduce you to Bob, who lives in an era where automated driving is becoming mainstream. Bob is the proud owner of the latest electric car that has advanced automated features, which are advertised to be able to do most of the driving for him and is rated as SAE level 3: Conditional Automation (SAE j3016). In this level of automation, the car is able to drive itself on specific roads, and thereby reduces the amount of monitoring of the driving task that Bob needs to do. However, the car might occasionally reach a limit in its capability to handle a scenario (e.g., a road without markings) and require Bob to intervene or assist. In such situations there is a "transition of control" of (part of) the driving task from the car to Bob. Given the success of auditory warnings for other car purposes (e.g., lane departure), Bob's car uses an auditory warning to signal a transition of control. When Bob fails to make an appropriate response to this signal, it might lead, at a minimum, to a stressful situation and at worst to a crash.

One day, when driving back from a family visit, the car suddenly runs into a fog bank. The visibility decreases and the system is unable to read the lane markings. At this moment the system alerts Bob with a sudden auditory signal that the control of the vehicle will be transferred to him. However, how susceptible is Bob's brain to such unexpected auditory signals? Does the signal trigger detection to such an extent, that an adequate response can be based on it? This question needs an answer.

The aforementioned scenario does not describe what Bob was doing while the car was performing the driving task. Bob believes to be able to occasionally engage in other tasks while the car has control. For example, occasionally he has a serious telephone conversation, he browses his email, uses social media, or plays a game on his phone. Now what if the auditory alert that triggered the transition of control occurred while Bob was engaged in such a task? How susceptible would he be then to this alert? Given that the types of tasks that Bob engages in vary from one situation to the next, it would be useful to have a more general understanding of how tasks that are not related to the driving task influence susceptibility to auditory alerts in particular.

1.2 Auditory alerts and cognitive load

Although Bob's scenario is presented as one of the future, the core elements of this example are already relevant today. Auditory alerts are already commonplace in current cars, and are frequently considered as mechanisms to signal transitions of control (see meta-analysis by Zhang, De Winter, Varotto, Happee, & Martens, 2019). Moreover, humans do also occasionally distract themselves in non-automated vehicles with a variety of non-driving tasks (Dingus et al., 2016; Klauer et al., 2014). This tendency increases when cars drive at higher levels of automation (De Winter, Happee, Martens, Stanton, 2014). At the highest levels of (full) automation, people will want to engage in an even wider variety of activities, including sleeping, similar to activities they do on trains or airplanes (Pfleging, Rang, & Broy, 2016).

One key concept in this context is "cognitive load". Cognitive load refers to the amount of cognitive resources that are absorbed by engagement in a task. The underlying assumption is that there is such a thing as cognitive resources that are limited in the extent to which they can be shared between two or more tasks (Wickens, 1981), such as driving and telephone activity. This prompts the question whether inducing cognitive load by some secondary task affects susceptibility to signals that are relevant for driving. Different non-driving related tasks (e.g., a serious telephone conversation or social media browsing on a smartphone) are hypothesized to make use of cognitive resources. Therefore, if one manipulated cognitive load experimentally, the results should be generalizable to different tasks and activities.

1.3 Measuring susceptibility

How can we validly and feasibly assess susceptibility in a variety of conditions, ranging from driving in a car to a controlled lab setting? One consideration is unobtrusiveness: The assessment should not go at the cost of an additional task requirement on part of the observer/ process controller. In this thesis I have opted for exposing participants to dedicated unexpected, and also unique (i.e., non-repeated) signals, or so-called "novels", without the requirement of a behavioral response. To assess susceptibility within this setting, I looked at the response to such novels of the brain, more specifically, the human cortex, in the form of event-related potentials (ERPs) derived from the EEG (electroencephalogram).

How do ERPs reflect susceptibility to these novels? There are at least two clues in this respect. First, the novels are usually presented as infrequent "oddballs", interspersed within a sequence of frequent identical tones, so-called "standards". To isolate the novelty aspect of the ERP, the ERP elicited by the standard can be subtracted from the ERP elicited by the novel, so that at least it can be excluded that the response to any sound would be qualified as susceptibility. Second, within the resulting difference ERP, the most prominent feature is a positive peak over frontal regions, around 300 ms after auditory-stimulus onset (Allison & Polich, 2008; Squires, Squires, & Hillyard, 1975; Ullsperger, Freude, & Erdmann, 2001). This novelty P3 or frontal P3 (fP3) has been argued to have a rather precise relation to behavioral performance, and can perhaps best be described as reflecting a global cortical reset: "Drop anything you were doing, here's a new situation that needs other actions" (Friedman, Cycowicz, & Gaeta, 2001; Polich, 2007; Kenemans, 2015; Wessel & Aron, 2013).

The frontal P3 has been used to index susceptibility in a variety of conditions and tasks, including driving (Wester, Böcker, Volkerts, Verster, & Kenemans, 2008; Wester, 2009), mental fatigue during driving (Massar, Wester, Volkerts, & Kenemans, 2010) manual tracking (e.g., Scheer, Bülthoff, & Chuang, 2016, 2018), games (e.g., Allison & Polich, 2008; Miller, Rietschel, McDonald, & Hatfield, 2011), and arithmetic (e.g., Ullsperger et al., 2001). The consistent finding of these studies is that auditory susceptibility, as indexed by the frontal P3, is reduced when people engage in visual-manual tasks, compared to, for example, a passive condition. This also includes manual driving. However, how might this change under automated driving conditions, when the driver is not using their hands to drive? And how susceptible are people to auditory signals during cognitive tasks that are not based on visual stimuli or manual interaction?[1]

1.4 Aims and generalizability

One aim of this thesis is to assess auditory susceptibility unobtrusively under conditions of (semi-) automated driving conditions and varying amounts of cognitive load (akin to a concurrent telephone conversation). This is what **Part I (Chapters 2-4; measuring susceptibility)** of this thesis is about. The second aim applies to explicit behavioral responses to auditory alerts or visual in-car warnings: How can we support auditory susceptibility to enhance control transition, and how adequate are current state-of-the-art visual in-car warnings for upcoming road situations? In **Part II (Chapters 5-6; testing interventions)** I attempt to answer these questions. The next section provides a more detailed outline of these chapters. Before that I present some notes on generalizability.

[1] In all my experiments frontal P3 (fP3) refers to the difference in positivity in the peak area around 350 ms post-stimulus at electrode location FCz, as derived from the subtraction of the standard ERP from the novel ERP.

Although the focus of this research is on the domain of semi-automated vehicles, the general principles apply to a broader domain of human-automation interaction, in particular safety-critical systems and other systems with requirement of a time-sensitive response. This type of systems is found in many settings (Janssen, Donker, Brumby, & Kun, 2019), including:

- Various transportation modes (e.g., airplanes, trains, boats, trams), where an 'auto-pilot' is in control with human back-up (e.g., Stanton & Salmon, 2011).
- Factory assembly lines, where machines and humans build a product, and humans handle emergencies (e.g., Boothroyd, 2005).
- Medical settings, where humans monitor automated procedures such as automated cardiopulmonary resuscitation (e.g., Wik et al., 2014).
- Process monitoring, where humans monitor factories, oil platforms, and power plants to prevent errors (e.g., Parasuraman, 1987).

1.5 Chapter outline

Table 1.1. summarizes the main research questions and their motivation for each chapter in this thesis. These are elaborated on more below. In **Part I (Chapters 2-4; measuring susceptibility)** I investigate how susceptible humans are to auditory stimuli. I therefore make use of electroencephalography (EEG) recordings to measure auditory susceptibility in different driving modes and levels of distraction.

More specifically, in **Chapter 2** (Van der Heiden et al., 2018) we investigate whether auditory susceptibility of the brain is reduced during automated driving compared to manual driving and a stationary car. Previous research, including the Utrecht PhD thesis of Wester (2009), has shown that susceptibility is reduced under manual driving compared to stationary conditions (Wester, 2009; Wester et al., 2008). Similarly, other research has shown that susceptibility is reduced under performance of visual-manual tasks (e.g., Scheer, Bülthoff, & Chuang, 2016, 2018; Ullsperger, Freude & Erdmann, 2001). But what happens under automated driving?

In a driving simulator we measured auditory susceptibility using an auditory oddball paradigm (Squires, Squires, & Hillyard, 1975) by measuring the oddball probe induced response in the brain. Beforehand, it was difficult to predict the outcome of this measurement: On the one hand, automated driving is like standing still, or being stationary, since no active manual vehicle control by the human is needed. On the other hand, automated driving is like driving, since there is dynamic visual information. A better insight into auditory susceptibility under automated driving can also contribute to important practical questions as current automated driving systems highly rely on last-minute auditory alerts (Zhang et al., 2019).

In the study from **Chapter 2** (Van der Heiden et al., 2018), the driver is performing no other concurrent tasks. This contrasts with observations in the literature, which predict that people will perform a wide variety of tasks while automation has main control over the driving task (De Winter et al., 2014; Janssen, Iqbal, Kun, & Donker, 2019; Pfleging et al., 2016). Given the wide variety of tasks, there is a need to understand general mechanisms that might be involved. We focus on cognitive load.

Before investigating the effects of the cognitive load inducing tasks combined with automated driving on auditory susceptibility, we first look into the effect of cognitive load itself on auditory susceptibility in **Chapter 3**. The first question that we address is whether cognitive load without visual or manual components reduces auditory susceptibility. Previous work suggests that cognitive load reduces auditory susceptibility (e.g., Allison & Polich, 2008; Miller, Rietschel, McDonald, & Hatfield, 2011; Ullsperger et al., 2001). However, these studies all either have a task where manual interaction is needed, or they present visual stimuli. In our study, we use an auditory task which is known to induce cognitive load (the verb generation task; Abdullaev & Posner, 1998; Snyder, Abdullaev, Posner, & Raichle, 1995). This allows us to measure cognitive load where concurrent manual interaction and visual stimuli are absent. As a third step, we perform a study in which both automated driving and a cognitive load inducing task (somewhat like a telephone conversation) are combined. We investigate how these two factors combined influence auditory susceptibility in **Chapter 4**.

In sum, **Part I (Chapters 2-4; measuring susceptibility)** investigates to what extent auditory susceptibility is affected not only by driving (Chapter 2), but also by automated driving (Chapter 2, 4) and when combined with separate distracting, cognitive load inducing tasks (Chapters 3, 4).

In **Part II (Chapters 5 and 6; testing interventions)** we turn to more applied scenarios and human behavior. What are ways to further support more accurate human behavior?

Chapter 5 (Van der Heiden, Iqbal, Janssen, 2017) proposes the use of auditory early warnings, which we call pre-alerts, to warn the driver of upcoming transitions of control. These pre-alerts start 20 seconds before the actual alert where the human driver has to act on. Such pre-alerts provide a way for mediated, negotiated, or scheduled interruptions (McFarlane, 2002; McFarlane & Latorella, 2002). The pre-alerts buy extra time and should accommodate a more relaxed transition of control in contrast to a more sudden switch of tasks (or “immediate interruption”, McFarlane, 2002; McFarlane & Latorella, 2002). To test whether pre-alerts indeed result in better transitions of control, we performed a driving simulator study where drivers transitioned between automated and manual driving, and in which an auditory warning of a transition was preceded by a pre-alert. We also compared how performance changed under distracted conditions (performing a visual-manual task on a smartphone).

In **Chapter 6** (Van der Heiden, Janssen, Donker, & Merkx, 2019) we tested visual (instead of auditory) in-car warnings in a driving simulator study. Specifically, this work was motivated by a practical scenario. Crash rates are higher at roadwork sites, compared to the same roads without roadworks (Khattak, Khattak, & Council, 2002). Moreover, in the Netherlands, failure to miss roadworks has resulted in various accidents (Van Gent, 2007). In line with current technological developments (IEEE, 2010), a Dutch-German-Austrian initiative (Cooperative ITS Corridor) is investigating the option for beacons at roadwork sites that can trigger in-car visual warnings, for example, to warn of an upcoming lane-closure. It is, however, unknown whether these signals allow for a timely response by the driver.

In our driving simulator, drivers had to respond to a visual warning indicating a closed lane, by changing lanes. We tested how fast drivers responded, including how well they responded under additional cognitive load in a task similar to the one described in Chapters 3 and 4 (verb generation). Our performance metric investigates both how distraction affects the initial response time of the driver (i.e., making the first steering movement) as well as the time needed to make the lane change. Is the reaction time sufficiently short for such a practical scenario?

In sum, **Part II (Chapters 5 and 6; testing interventions)** provides insight in more practical applications. Both chapters demonstrate the need for (Chapter 6) and ability to (Chapter 5) use timely alerts to warn drivers of (automated) vehicles of upcoming events.

	Chapter	Main research question	Relevance
Part I Measuring auditory susceptibility	2	Does automated driving reduce auditory susceptibility?	For safe operation of semi-automated vehicles, the human driver has to respond to a request for intervention ("transition of control"). But we don't know how susceptible human drivers are to auditory information in general.
	3	Does cognitive load reduce auditory susceptibility?	Since drivers tend to engage in a variety of non-driving tasks, it is valuable to study the effect of cognitive load on auditory susceptibility in general.
	4	Does cognitive load during automated driving reduce auditory susceptibility?	Since it is likely that drivers in semi-automated vehicles engage in non-driving related tasks, we study the interaction of automated driving and concurrent cognitive load on auditory susceptibility.
Part II Testing interventions	5	Can pre-alerts improve the transfer of control in semi-automated vehicles?	As drivers might miss auditory signals, we test a potential solution ("pre-alerts") that allows drivers more time to switch between tasks.
	6	Do last-minute visual in-car warnings accommodate a timely lane change?	In current technological developments, beacons along the road can trigger in-car visual warnings. It is, however, unknown whether these signals allow for a timely response by the driver.

Table 1.1. Research questions and relevance to the thesis for each chapter.

References

Abdullaev, Y. G., & Posner, M. I. (1998). Event-related brain potential imaging of semantic encoding during processing single words. *Neuroimage*, 7(1), 1-13.

Allison, B. Z., & Polich, J. (2008). Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biological psychology*, 77(3), 277-283. https://doi.org/10.1016/j.biopsycho.2007.10.014

Ayoub, J., Zhou, F., Bao, S., & Yang, X. J. (2019). From Manual Driving to Automated Driving: A Review of 10 Years of AutoUI. *In: Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications,* 70-90, New York, NY: ACM

De Winter, J. C., 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.

Dekker, S. W., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human–automation co-ordination. *Cognition, Technology & Work*, 4(4), 240-244.

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, *113*(10), 2636-2641.

Hancock, P. A., Jagacinski, R. J., Parasuraman, R., Wickens, C. D., Wilson, G. F., & Kaber, D. B. (2013). *Human-automation interaction research: Past, present, and future. Ergonomics in Design*, 21(2), 9-14.

IEEE, 2010. Ieee standard for information technology– local and metropolitan area networks– specific requirements– part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 6: Wireless access in vehicular environments. IEEE Std 802.11p-2010 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, IEEE Std 802.11n-2009, and IEEE Std 802.11w-2009), pp.1–51.

Janssen, C. P., Boyle, L. N., Kun, A. L., Ju, W., & Chuang, L. L. (2019). A hidden markov framework to capture human–machine interaction in automated vehicles. International Journal of Human–Computer Interaction, 35(11), 947-955.

Janssen, C. P., Donker, S. F., Brumby, D. P., & Kun, A. L. (2019). History and future of human-automation interaction. *International Journal of Human-Computer Studies*, 131, 99-107.

Janssen, C.P., Donker, S.F., Chuang, L. L. & Ju, W. (2019). Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. (392 p.). New York, NY: ACM Press.

Janssen, C. P., Iqbal, S. T., Kun, A. L., & Donker, S. F. (2019). Interrupted by my car? Implications of interruption and interleaving research for automated vehicles. *International Journal of Human-Computer Studies*, *130*, 221-233.

Khattak, A.J., Khattak, A.J., Council, F.M., 2002. Effects of work zone presence on injury and non-injury crashes. *Accident Analysis & Prevention* 34, 19–29. doi:10.1016/S0001-4575(00)00099-3

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England journal of medicine*, 370(1), 54-59.

Kun, A. L. (2018). Human-machine interaction for vehicles: Review and outlook. *Foundations and Trends® in Human–Computer Interaction*, 11(4), 201-293.

McFarlane, D. C. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction*, 17(1), 63-139.

McFarlane, D. C., & Latorella, K. A. (2002). The scope and importance of human interruption in human-computer interaction design. Human-Computer Interaction, 17(1), 1-61.

Miller, M. W., Rietschel, J. C., McDonald, C. G., & Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. *International Journal of Psychophysiology*, 80(1), 75-78.

Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. *Reviews of human factors and ergonomics*, 7(1), 58-99.

Noy, I. Y., Shinar, D., & Horrey, W. J. (2018). Automated driving: Safety blind spots. *Safety Science*, 102, 68–78. http://doi.org/10.1016/j.ssci.2017.07.018

Parasuraman, R. (1987). Human-Computer Monitoring. *Human Factors*, 29 (6), 695–706.

SAE International. (2018). j3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. Retrieved from https://www.sae.org/standards/content/j3016_201806/

Snyder, A. Z., Abdullaev, Y. G., Posner, M. I., & Raichle, M. E. (1995). Scalp electrical potentials reflect regional cerebral blood flow responses during processing of written words. *Proceedings of the National Academy of Sciences*, 92(5), 1689–1693. http://doi.org/10.1073/pnas.92.5.1689

Stanton, N. A., & Salmon, P. M. (2011). Planes, trains and automobiles: Contemporary ergonomics research in transportation safety. *Applied ergonomics*, 42(4), 529-532.

Ullsperger, P., Freude, G., & Erdmann, U. (2001). Auditory probe sensitivity to mental workload changes–an event-related potential study. *International Journal of Psychophysiology,* 40(3), 201-209. https://doi.org/10.1016/S0167-8760(00)00188-4

Van der Heiden, R. M., Iqbal, S. T., & Janssen, C. P. (2017). Priming drivers before handover in semi-autonomous cars. *In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 392-404). ACM.

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., Hardeman, L. E., Mans, K., & Kenemans, J. L. (2018). Susceptibility to audio signals during autonomous driving. *PloS one*, 13(8), e0201963. https://doi.org/10.1371/journal.pone.0201963

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., & Merkx, C. L. (2019). Visual in-car warnings: How fast do drivers respond? *Transportation research part F: traffic psychology and behaviour*, 65, 748-759.

Van Gent, A.L., 2007. *Verkeersonveiligheid bij werk in uitvoering: een literatuurstudie.* Den Haag, The Netherlands: SWOV. Report nr: R-2007-5

Wessel, J.R., & A.R. Aron, Unexpected Events Induce Motor Slowing via a Brain Mechanism for Action-Stopping with Global Suppressive Effects. *The Journal of Neuroscience*, 2013. 33(47): p. 18481-18491.

Wester A.E. Attention Deficit and Impulsivity: Driving, Drugs and Electrophysiology. 2009. Utrecht: Utrecht University. PhD thesis.

Wester, A. E., Böcker, K. B. E., Volkerts, E. R., Verster, J. C., & Kenemans, J. L. (2008). Event-related potentials and secondary task performance during simulated driving. *Accident Analysis & Prevention,* 40(1), 1-7.

Wik, L., Olsen, J.-A., Persse, D., Sterz, F., Lozano, M., Jr., Brouwer, M.A., Westfall, M., Souders, C.M., Malzer, R., Van Grunsven, P.M., Travis, D.T., Whitehead, A., Herken, U.R. and Lerner, E.B. (2014). *Manual vs. integrated automatic load-distributing band CPR with equal survival after out of hospital cardiac arrest. The randomized CIRC trial. Resuscitation*, 85 (6), 741–748.

Walch, M., Lange, K., Baumann, M., & Weber, M. (2015). Autonomous driving: investigating the feasibility of car-driver handover assistance. In *Proceedings of the 7th international conference on automotive user interfaces and interactive vehicular applications* (pp. 11-18).

Zhang, B., De Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation research part F: traffic psychology and behaviour*, 64, 285-307.

Part I

Measuring auditory susceptibility

Chapter 2

Susceptibility to audio signals during autonomous driving

Remo M. A. van der Heiden
Christian P. Janssen
Stella F. Donker
Lotte E. S. Hardeman
Keri Mans
J. Leon Kenemans

Published as:
Van der Heiden, R. M., Janssen, C. P., Donker, S. F., Hardeman, L. E., Mans, K., & Kenemans, J. L. (2018). Susceptibility to audio signals during autonomous driving. *PloS one*, 13(8). doi: 10.1371/journal.pone.0201963

Author contributions:
RMAV designed and programmed the experiment, analyzed the data, and wrote the manuscript with input from the other authors. LESH and KM collected the data and contributed to experimental design. CPJ, SFD, and JLK supervised the project, contributed to experimental design, interpretation of the data, and editing the manuscript. All authors read and approved the final version of the manuscript.

Abstract

We investigate how susceptible human drivers are to auditory signals in three situations: when stationary, when driving, or when being driven by an autonomous vehicle. Previous research has shown that human susceptibility is reduced when driving compared to when being stationary. However, it is not known how susceptible humans are under autonomous driving conditions. At the same time, good susceptibility is crucial under autonomous driving conditions, as such systems might use auditory signals to communicate a transition of control from the automated vehicle to the human driver. We measured susceptibility using a three-stimulus auditory oddball paradigm while participants experienced three driving conditions: stationary, autonomous, or driving. We studied susceptibility through the frontal P3 (fP3) Electroencephalography Event-Related Potential response (EEG ERP response). Results show that the fP3 component is reduced in autonomous compared to stationary conditions, but not as strongly as when participants drove themselves. In addition, the fP3 component is further reduced when the oddball task does not require a response (i.e., in a passive condition, versus active). The implication is that, even in a relatively simple autonomous driving scenario, people’s susceptibility of auditory signals is not as high as would be beneficial for responding to auditory stimuli.

2.1 Introduction

Human driving involves various tasks, ranging from control (or operational) tasks such as basic vehicle control, to maneuvering (or tactical) tasks such as a lane change, to strategic tasks such as navigation Michon (1985). With the advent of semi-autonomous cars and other automated vehicles, the role of the human driver changes. Some tasks might be automated partially, or even fully, and left to the system to control, whereas other tasks might still remain with the driver. As automation increases, the role of the human changes more from an active driver to that of a system monitor (e.g., Parasuraman, Sheridan, & Wickens, 2000). What the human driver needs to monitor or act on, depends on the degree of automation, or autonomy, of the vehicle. To classify the role division between human and system, the Society of Automotive Engineering (SAE International, 2014) has distinguished six levels of automation. These range from level 0, indicating no driving automation, to full automation at level 5. In between these extremes are four levels of division of control between human and system. Cars at level 1 and 2 are currently being sold commercially, and require, by SAE definition (SAE International, 2014), the human driver to monitor the driving environment. In the next levels, level 3 and 4, the car monitors the driving environment, and the human can be requested to respond to intervene in the driving (also referred to as a transition of control). In level 3 such human intervention is required, whereas in level 4 this is optional and the car needs to have a fallback strategy in case the human does not intervene. At the final level, level 5, human assistance is not needed, but even then the car might occasionally ask for human input (for example, what route to take if there are multiple options). Stated differently, at the upcoming levels of automation (SAE levels 3, 4, and 5), human drivers can be signaled to provide input to the vehicle.

Although there is no specification or standard for how a vehicle might signal that a human's input is requested, a likely candidate are audio signals. One advantage of audio signals is their near omnidirectional character (i.e., perception of auditory stimuli is "gaze-free") which yields a higher chance of being perceived compared to visual signals. Moreover, auditory signals are already widely used in vehicles for notifications (often supported by visual signals), for example to alert that the vehicle is running out of petrol, as a seatbelt warning, or as proximity indication for parking sensors. Therefore, investigating susceptibility to auditory signals for (semi-) autonomous vehicles is timely, given the pace of development of automated vehicles, in which there is shared control between the human driver and the vehicle.

A potential complication is that the extent of auditory detection and processing may depend on the extent of engagement in an ongoing task, such as driving (for a recent review see (Murphy, Spence, & Dalton, 2017). This makes sense from a perceptual-load perspective (Lavie, 1995, 2005), which holds that increasing perceptual demands within an ongoing task reduces the extent of processing of stimuli that are task-irrelevant or that are only of secondary relevance. In such perceptual load experiments, performance indices of load-dependent processing reductions are typically based on interference from the irrelevant stimuli with processing of task-relevant stimuli, or additional task elements such as a posteriori subjective report (see reviews in (Lavie 2005; Murphy, Spence, & Dalton, 2017)). Cognitive load can also be measured by exposing participants to a stimulus without requiring a behavioral response (Wester 2009; Wester, Böcker, Volkerts, Verster & Kenemans, 2008; Wester, Verster, Volkerts, Böcker, & Kenemans, 2010). By not requiring a behavioral response, the approach is unobtrusive. Instead, measures of human electrocortical activity as available in the electroencephalogram (EEG) can be used. EEG reflects the waxing and waning of postsynaptic potentials that necessarily precede possible action potentials, when this is synchronized across ten thousands of cortical neurons. The common interpretation is that both spontaneous EEG and the change in EEG in response to discrete events can reveal the involvement of specific cognitive processes (for a broader introduction see Luck (2014)).

The unobtrusiveness of EEG (i.e., its applicability without additional task demands on part the participant) makes the approach suitable for use in an automotive setting. Note that numerous studies have utilized measures of spontaneous EEG to monitor vigilance in driving or driving-like conditions (e.g., De Waard & Brookhuis, 1991; Khushaba, Kodagoda, & Lal, 2011; Schmidt, Schrauf, Simon, Fritzsche, Buchner, & Kincses, 2009; Wertheim, 1978). Within the present approach, however, susceptibility to a discrete auditory stimulus, is assessed by deriving from the EEG discrete electrocortical responses (or event-related potentials, ERPs) to these very auditory signals. One fruitful implementation of this principle has been demonstrated by Wester and colleagues (Wester, 2009; Wester et al., 2008; Wester et al., 2010) in a variety of driving(-related) conditions, involving the recording of the frontal P3 or fP3.

The fP3 (also known as P3a or novelty P3) is preferably recorded in response to task-irrelevant unique, relatively unexpected, environmental sounds ("novels"; e.g., Dien, Spencer, & Donchin, 2003; Friedman, Cycowicz, & Gaeta, 2001; Heitland, Kenemans, Oosting, Baas, & Böcker, 2013; Kiehl, Laurens, Duty, Forster, & Liddle, 2001; Massar, Wester, Volkerts, & Kenemans, 2010; Schröger & Wolff, 1998). A dominant current view is that the frontal P3 provides relatively widespread inhibition throughout the brain or a neural signal that instigates such widespread inhibition (cf. Friedman et al., 2001; Kenemans, 2015; Polich, 2007; Wessel & Aron, 2017) As a generic mechanism, it is thought to free sufficient neurocognitive and behavioral resources to meet whatever demands a sudden change in the environment (e.g., an unexpected sudden honk of a truck) might impose (Kenemans, 2015, p. 122).

The fP3 to novels is commonly recorded when driving is combined with an auditory "three-stimulus-oddball" novelty/ setup. In a three-stimulus-oddball set-up, a participant hears a series of sounds. The large majority of these sounds are constant tones (also referred to as standard), a small subset are deviant tones of a slightly higher frequency, and a third set contain novel sounds (irrelevant unique, relatively unexpected, environmental sounds) (see also Methods for more details on our implementation).

Within a driving context, Wester and colleagues (Wester, 2009; Wester et al., 2008; Wester et al., 2010) demonstrated that the fP3 component is significantly reduced during driving compared to a stationary control condition. This has also recently been replicated by Scheer, Bülthoff, and Chuang (2016). The interpretation of these results is that the brain is less inclined to meet the demands that a sudden change in the environment might impose when a person is driving compared to when the car is stationary.

A further finding in the Wester et al. studies is that directing attention to auditory information in general restores the driving-reduced fP3 to a considerable extent. In the typical "three-stimulus-oddball" novelty/ fP3 setup, novel stimuli are interspersed (on 10% of the trials) in a sequence of 'standard' tones (e.g., 1000 Hz, 80% of all stimuli) and 'deviant' tones (e.g., 1100 Hz, also on 10% of the trials). It has turned out that, relative to a driving-only condition in which the oddball stimuli are completely irrelevant, instructing participants to manually respond to the deviant tone enhances the fP3 to the novels (Wester, 2009; Wester et al., 2008; Wester et al., 2010). This is important because it suggests that susceptibility for rare but salient and potentially relevant audio signals (e.g., alerts to signal a transition of control) can be augmented by having human controllers engage in a second auditory task.

What might these results predict for autonomous driving? One could argue that autonomous driving is comparable to non-autonomous driving given the aspects of moving through space and being exposed to changing visual input that is safety-critical. Alternatively, the autonomous condition may be more comparable to being stationary, in the sense that, unlike manual driving, there is no need to control a vehicle. How does an automated setting then influence a human's susceptibility to auditory signals? We foresee three possible outcomes. The susceptibility in automated driving can be (1) reduced, similar to the observed reduction in non-autonomous driving, (2) unabated with respect to a non-driving situation, similar to stationary, or (3) lie somewhere in between these extremes.

To investigate how autonomous driving affects human susceptibility to auditory alerts, we conducted an experiment in a driving simulator where participants experience three conditions in a within-subjects manipulation: (1) watching a stationary screenshot of the driving simulator, i.e., a non-driving situation, (2) driving manually in a simulator, and (3) being driven by a simulated autonomous car. Meanwhile, oddball stimuli were presented at a regular pace to allow for measurement of the fP3. In addition, we compared whether there is a difference between a condition where an active response to auditory stimuli is required and a passive condition in which no such response is needed.

2.2 Methods

2.2.1 Participants

To determine the number of participants, we conducted a power test. Starting point was the effect size that was measured in previous research (Wester et al., 2008) on the same metric of interest as we use here (fP3 amplitude, see section on measurement and data analysis). Wester et al. found an effect size (d) of 1.8 for the difference between stationary and regular driving, which we took as input for our power test. We also set our intended level of significance (alpha) at .05, and intended power (one-tailed) at 0.8. If this information is entered into G*Power 3.1.9.2, it yields a required sample of n = 4 participants. Because in the present study the autonomous condition was included (i.e., we have a third driving condition; not just stationary and regular driving as in Wester et al.), and a comparison between the passive and the active conditions was envisaged, we anticipated that more participants were needed to detect effects between these 3 x 2 conditions (see section on Design for details on the experimental design). We decided to include at least four times as many participants (i.e., at least 16).

Eventually eighteen participants (7 M; 11 F) were recruited from a student population using opportunity sampling (ages 20 to 25 years old, M = 22.06 years, SD = 1.39 years). All participants were native Dutch speakers and reported to have normal or corrected-to-normal vision and normal hearing. This study also required the possession of a driver's license (possession varied from 2 weeks to 6 years, M = 3.56 years, SD = 1.55 years). One participant (participant 8) did not adhere to task instructions, and showed reckless driving during the experiment. We therefore recruited another participant (19) to replace this participant. The above statistics of participant set are of the final set of eighteen participants (participants 1–7 and 9–19). The experiment was approved by the ethics committee of the Faculty of Social and Behavioral Sciences of Utrecht University (approval number FETC16-042). All participants gave written informed consent prior to the experiment. Figure 2.1 contains a picture of an individual in the set-up (not a participant). This individual has given written informed consent (as outlined in PLOS consent form) to publish this figure.

2.2.2 Material

Driving simulator

We used a low-fidelity desktop-based driving simulator. Participants were seated in an adjustable chair at about 120 cm from the simulator screen. A 29-inch screen showed a road from the driver's perspective. Three mirrors and analogue speed and rpm indicators were presented on the screen. A Logitech G27 steering wheel and pedals were used to perform the driving task. The simulated car was set to automatic transmission. The set-up is shown in Figure 2.1A and a screen-shot of the driver's perspective on the road is shown in Figure 2.1B.

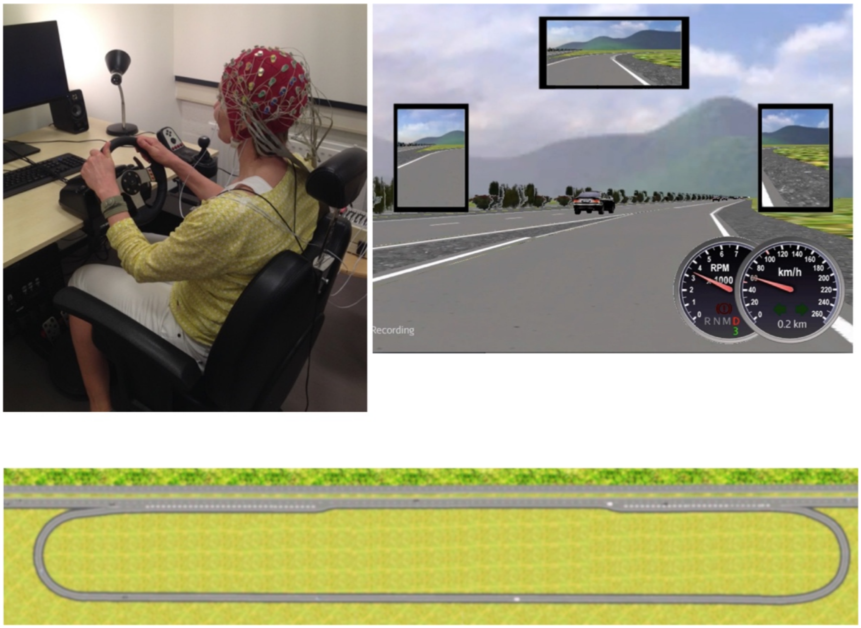


Figure 2.1. Set-up of the driving study. (Top-left: 2.1A) Picture of the driving simulator setup with an individual (not a participant) wearing an EEG cap and 64 electrodes attached. The individual has given written informed consent (as outlined in PLOS consent form) to publish these case details (i.e., to publish this picture). (Top-right: 2.1B) Screenshot of the driver's perspective on the simulated road. (Bottom: 2.1C) A schematic view of the course with starting point indicated and single lane road on the bottom and 2 lane highway on the top.

For the software, we used a modified version of the MotorwayTask in Open Driving Simulator (OpenDS) version 4.0 (Math, Mahr, Moniri, & Müller, 2012). This scenario consisted of a loop, as illustrated in Figure 2.1C, which participants had to drive twice. Participants started on a single-lane road, with a maximum posted speed limit of 70 km/h. Then they entered a two-lane highway via a right-sided ramp. This highway had a maximum posted speed limit of 100 km/h. After driving on the highway for 1000 meters, they had to exit the ramp on the right, to enter the single lane road again. After 1000 meters they were at the start of the scenario again. To increase the impression of driving on a highway, other cars were also present on the road. Each driving trial lasted approximately 5 minutes.

2.2.3 Auditory novelty oddball paradigm

Participants were presented with three types of sounds: a standard tone, a deviant tone, and novel stimuli. The standard tone was a 1000 Hz tone, presented for 400 ms. The deviant tone was 1100 Hz tone, presented for 400 ms. The novel stimuli consisted of 100 different environmental sounds retrieved from a database (Fabiani & Friedman, 1995), such as animal sounds (e.g., dog barking, bird chirping), human sounds (e.g., coughs, laughs, sneezes), or other sounds (e.g., hammering, water running). Novel sounds varied in duration between 159 ms and 399 ms. In each block, 130 oddball stimuli were presented. Of these oddball stimuli, 80% were standard tones, 10% were deviant target tones, and 10% were novel sounds (cf. Wester et al., 2008). The interval between two stimuli was 2.34 s (measured between onsets of the two sounds). All sounds were played binaurally using Presentation software (Neurobehavioural Systems) at 75 dB through Earlink earphones.

2.2.4 Design

Autonomous driving was compared to manual driving and stationary/non-driving using a 2 (oddball response requirement) x 3 (driving mode) mixed factorial design. Between subjects we manipulated the oddball response requirement. That is, half of the participants was instructed to press a button on the top left side of the steering wheel as quickly as possible whenever they heard a deviant tone (active condition), the other half of the participants had no button press requirement at all (passive condition).

Within-subjects, the driving modes consisted of three levels: stationary, driving, and autonomous. In the stationary condition, participants were asked to look at a still frame, which showed a screenshot of the simulated road. In the driving condition, participants were asked to drive on a track in a driving simulator (see Figure 2.1C). In the autonomous condition, participants were given the impression of being driven by a semi-autonomous car by showing them a pre-recorded screen capture video of the track as used in the driving condition (i.e., a pre-recorded screen-recording of one of the experiment leaders driving). To provide the illusion that the autonomous driving condition was indeed a driving condition, the recording also included the startup screens of the simulator, identical to the driving condition. This made it visually indistinguishable from running the actual simulator software. As the driving simulator did not require a driving response from the user, it is ambiguous how to map the autonomous condition directly to a specific level of the SAE framework (SAE International, 2014). Given that the car drives itself (including lane change), the simulation can be considered SAE level 3 or higher.

Each participant performed each driving mode condition four times, creating a total of 12 blocks. The order of blocks (and therefore, conditions) was semi-randomized. Per set of three blocks, all three driving modes (*S*tationary, *A*utonomous, *D*riving) were experienced in the same order (e.g., *S-A-D—S-A-D—S-A-D—S-A-D*). Each ordering was experienced by both a participant in the active condition, and one in the passive condition. Since the total permutations was 6, each unique ordering was experienced by at least one participant in the active condition and one participant in the passive condition, and three orderings were experienced by two participants per active and passive level.

2.2.5 Procedure

After receiving general instructions and walking through the consent procedure, participants started by practicing the driving task. After instructions and adjustment of the simulator seat, participants either were driven autonomously, or drove themselves for one block (order was balanced between participants). The driving practice was followed by an oddball practice. For the oddball practice trials, the oddball task was shortened from 130 to 42 stimuli but the proportion of novel and deviant tones was kept in accordance with the experimental trials (i.e., 80-10-10%). After the practice trials, we checked verbally whether no complaints due to motion sickness occurred. None occurred during the experiment.

After the practice trials, the EEG cap with electrodes was applied. Participants then completed the twelve experimental blocks (see section on design). Between blocks was a small break (roughly 1 minute), and after six blocks a slightly longer break (i.e., about 5 minutes). The aim of the break was to not overstrain the participant. They were asked to stay in their seat and relax. No specific instructions were given on what to do during the break.

For purposes of internship training, after the last block of each driving condition (blocks 10, 11, and 12), we asked participants to fill out a NASA raw-TLX questionnaire (Hart & Staveland, 1988). The results of this questionnaire are not reported here.

After the experimental trials, the EEG cap and electrodes were removed. Participants were then requested to complete a questionnaire that collected demographic data and their subjective experience of the experiment. For purposes of internship training, a questionnaire on impulsivity was also administered. The results of this questionnaire are not reported here. The total procedure took between 120 and 150 minutes, mostly depending on the length of breaks and time needed to fill-out the questionnaire.

2.2.6 Instructions to participants

Participants received various explicit instructions verbally, and later also on the screen of the simulator. For the oddball task, participants in the active response condition were instructed to respond as fast as possible to the deviant tone and that no response was needed to other auditory stimuli. Participants in the passive condition were instructed to respond to none of the auditory stimuli.

With respect to the driving task, participants were told to follow the road and complete the motorway task as described in the materials section. Furthermore, participants were asked to adhere to normal traffic regulations and to drive safely. For the autonomous and stationary conditions, participants were told to stay focused on the screen. In all conditions, participants were instructed to keep their hands on the steering wheel.

2.2.7 Measurement and data analysis

ERP recording and analysis

To be able to measure the susceptibility of the brain to auditory stimuli, event-related potentials (ERP) in response to the oddball stimuli were measured using electroencephalogram (EEG) recordings. EEG was recorded by a BioSemi ActiveTwo system at 2048 Hz. 64 active Ag-AgCl electrodes were positioned according to the international 10/10 system (Sharbrough, Chatrian, Lesser, Luders, Nuwer, Picton, 1991). Electrodes were also placed on left and right mastoids for purposes of offline re-referencing. To be able to compensate for eye-activity related noise, we recorded horizontal and vertical eye movements and eye blinks, for which electro-oculography (EOG) electrodes were placed on both outer canthi of the eyes as well as above and below the right eye in line with the pupil.

After the experiment, ERPs were analyzed using Brain Vision Analyzer 2 software (Brain Products GmbH, München, Germany). Next to a 50 Hz notch filter to compensate for noise from the mains, 0.16 Hz high-pass filter with a slope of 24 dB/oct and a 30Hz low-pass filter with a slope of 24 dB/oct were applied to the EEG data. Data were referenced to the average of left and right mastoid signal. Trials containing false alarms (i.e., button-press responses to the standard or novel stimuli), misses (i.e., failed responses to the deviant target tone), and invalid responses (i.e., responses outside the 100–950 ms interval, after correction for a delay—see below) were removed.

After these initial steps, the data were binned per condition (stationary, autonomous, and driving) and per oddball stimulus (standard, deviant, and novel), resulting in 9 bins. To control for eye movements and blinks, an ocular correction was applied to every segment using the Gratton, Coles and Donchin method (Gratton, Coles, & Donchin, 1983). The data were baseline corrected over the 100 ms interval preceding the stimulus onset. Subsequently, artifacts were automatically rejected (when at least one of the following conditions was met within the epoch: maximum voltage step > 120 μv/ms; maximum difference > 100 μv; minimum activity < 0.5 μv). Average ERP waveforms over all participants were calculated per stimulus type (standard, deviant, and novel) per condition (driving, stationary, and autonomous).

Difference waves were calculated by subtracting the average ERP for the standard tones from the average ERPs for the novels for each driving condition. Based on Wester et al., (2008), we analyzed the mean of the difference wave at electrode FCz in the interval 325–375 ms after stimulus onset. During the analysis of the experiment, we became aware, and cross-tested, that the audio presentation was systematically delayed by 50 ms compared to the logged time. We therefore corrected the reported time log, such that the ERP response wave is relative to actual time presentation, not the logged time (i.e., 50 ms later than logged). In our analysis of the invalid responses (discussed above) we therefore also use an adjusted interval of 100 to 950 ms (compared to 150 to 1000 ms in Wester et al., (2008)).

2.2.8 Reaction time (active response condition)

For the active condition only, we could measure how quickly participants responded to a deviant tone. Specifically, we analyzed the time interval between the onset of the deviant tone, and the onset of the button press on the steering wheel as a measure of reaction time. Button presses were registered by the simulator computer and transferred to the BioSemi as marker codes directly. Reaction times were also corrected by subtracting 50 ms to account for the delayed stimulus presentation.

2.2.9 General questionnaire

The questionnaire data was used to check for inconsistencies (e.g., whether participants had not understood instructions). One participant was removed based on this description (see participants section). No other obvious inconsistencies were found, therefore not providing grounds to a priori remove participants from the analysis.

2.2.10 Statistical analysis

Statistical analyses were conducted in SPSS 24. We used a 2 x 3 mixed multivariate-contrast ANOVA, with Oddball response requirement as a between-subjects factor (two levels: passive and active) and Driving mode as within-subjects factor (three levels: stationary, autonomous, and driving, transformed to linear and quadratic contrasts). Throughout all analyses, a significance level of $\alpha = .05$ was used. For significant effects, the *d* effect size was calculated as the square root of the *F*-value divided by the square root of the sample size.

Given omnibus effects of driving mode, we used pairwise contrast tests for differences between the three driving conditions. In the introduction, we anticipated three possible outcomes. If the fP3 component is reduced in a similar fashion in the autonomous condition as it is during driving, this would be reflected in differences between autonomous and driving on the one hand, and stationary on the other. If the fP3 component is unabated in the autonomous relative to the stationary condition, there would be significant differences with the driving condition only. If the magnitude of the fP3 in the autonomous condition lies somewhere in between the stationary and the regular driving condition, this would manifest in a significant difference between these latter two condition, with smaller differences between either of these conditions and the autonomous condition.

2.3 Results

2.3.1 Susceptibility to novel sounds (fP3)

Driving mode had a significant main effect on fP3 amplitude, $F(2, 15) = 12.4$, $p < .005$. The difference waves of the three driving conditions are shown in Figure 2.2 for the active condition (left panel) and passive condition (right panel). The fP3 was strongest in stationary ($M = 10.3$ μV, $SD = 3.5$ μV), followed by autonomous ($M = 7.8$ μV, $SD = 3.4$ μV), followed by driving ($M = 6.2$ μV, SD = 2.9 μV). Hence, the reduction in fP3 in autonomous conditions was in between that observed in stationary and regular driving (i.e., as in option 3 from the introduction). Pairwise comparisons yielded significant differences between stationary and driving ($p < .001$, d = 1.21), between autonomous and driving ($p = .048$, d = 0.51), and between autonomous and stationary ($p = .008$, d = 0.71). Hence, the effect size for stationary versus driving was somewhat smaller than in the Wester et al. (2008) study (d = 1.8), but still in the large range. The effects sizes for the comparisons involving the autonomous condition, although significant, were clearly smaller (than 1), confirming the pattern specified in option 3.

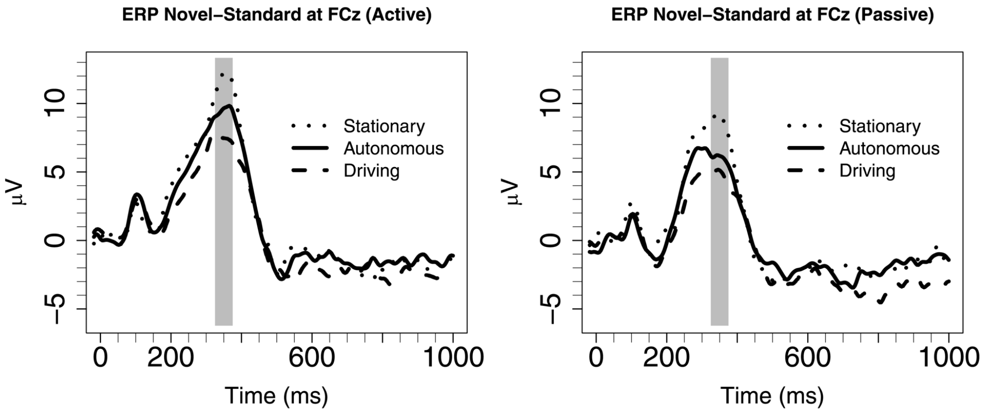


Figure 2.2. Grand-average difference ERPs (novel minus standard) (microvolts) in different driving modes (lines), relative to onset of auditory stimulus (0 s) for the active condition (left panel) and the passive condition (right panel). The shaded area indicates the epoch of frontal P3. Both autonomous (solid line) and driving (dashed line) show a reduced frontal P3 peak compared to stationary (dotted line).

There was also a main effect of oddball response requirement on fP3 amplitude. The active response requirement group had a significantly larger fP3 (M = 9.6 μV, SD = 3.6 μV) compared to the passive group (M = 6.6 μV, SD = 3.2 μV), $F(1, 16)$ = 8.00, p = .01, d = 0.67. There was no significant interaction between the driving mode and Oddball response requirement, p > .1. Figure 2.3 shows a bar plot of the average fP3 response in the active and passive condition for each of the three driving conditions. The figure illustrates the main effect of response requirement (that active conditions generated higher fP3 amplitude compared to the passive conditions) and the main effect of driving condition (i.e., that fP3 amplitudes are highest in the stationary condition, followed by the autonomous condition, followed by the driving condition).

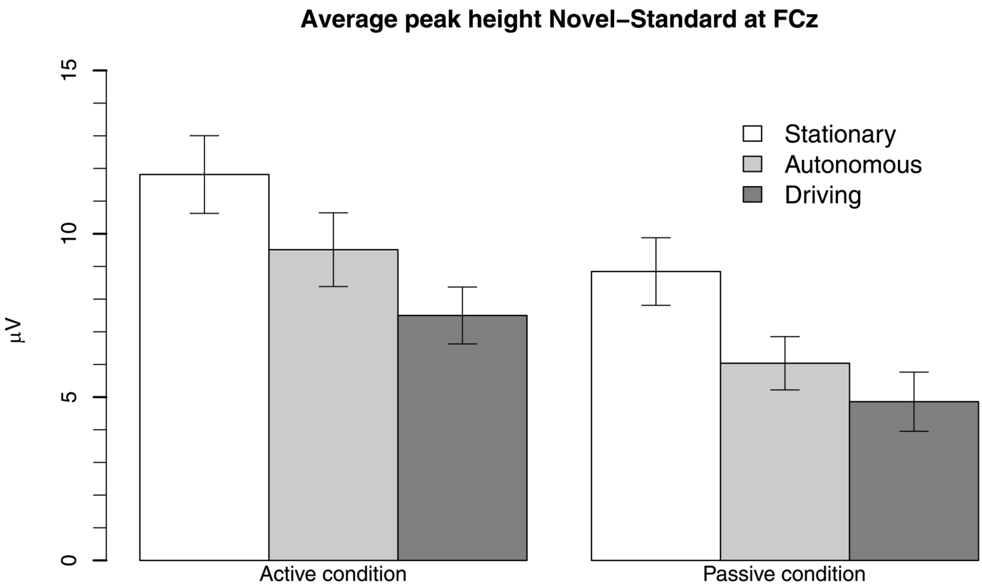


Figure 2.3. Bar plot of average fP3 response per oddball response category and per driving condition. The data show a main effect of oddball response category, and a main effect of driving conditions.

To gain further insight in the wider cortical response to the novel stimuli, Figure 2.4 plots a heatmap of the distribution of amplitude (μV) across the scalp (i.e., all 64 electrodes). Amplitudes concern the difference wave between the novel and standard ERP (i.e., similar to Figure 2.2). The data were based on measurements at all 64 electrodes and binned in bins of 50 ms, starting at 25 ms after stimulus onset. Values were capped at 0 μV and 12 μV (i.e., values outside this range are given the color of the boundary). Dots show the locations of the 64 electrodes in their default reference position. The FCz electrode is the fourth electrode from the nose down. The top three rows are data from the active response condition for stationary, autonomous, and driving. The bottom three rows are the data from the corresponding driving condition in the passive response condition.

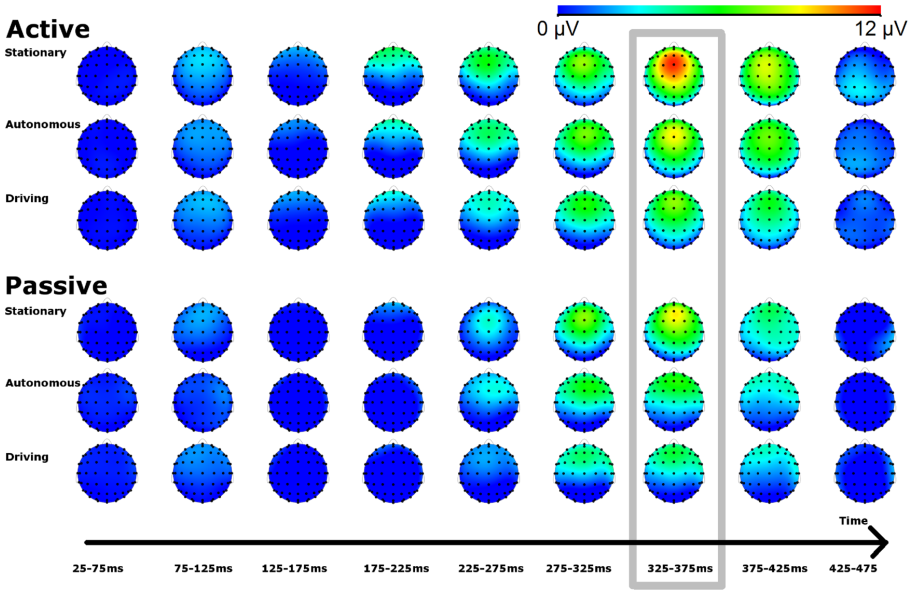


Figure 2.4. Scalp distribution of difference in EEG activity in response to the novel relative to the standard. fP3 activity is strongest around the FCz electrode (4th from the nose down) in the 325-375 ms interval.

Figure 2.4 again highlights that the fP3 activity is strongest in the 325–375 ms interval (7th scalp picture from the left; surrounded by a grey border for emphasis). The figure also illustrates that the amplitude is the highest surrounding the FCz electrode. Moreover, the figure also illustrates that fP3 activity is reduced in the passive response condition compared to the active response condition, and that the fP3 activity is reduced in autonomous compared to stationary, and in driving compared to autonomous.

2.3.2 Oddball response time (RT)

In the active condition, we measured the response time between the moment a deviant sound was presented and the moment a button was pressed. The MANOVA revealed no significant effect of driving condition (p = .127). Looking at the means, reaction time was highest in the driving condition (M = 594 ms, SD = 107 ms), followed by the autonomous condition (M = 551 ms, SD = 108 ms), followed by the stationary condition (M = 545 ms, SD = 98 ms).

We also analyzed response time variability, through analysis of the standard deviation of reaction time in response requirement. The MANOVA again revealed no significant effect of driving condition (p = .122). Looking at the means, standard deviation was highest in the driving condition (M = 140 ms, SD = 25 ms), followed by the stationary condition (M = 127 ms, SD = 32 ms), and the autonomous condition (M = 116 ms, SD = 24 ms).

In previous studies, different patterns were found for both measures: one study found that response time to the deviant tones were more variable (but not longer) during simulated driving (Wester et al., 2008), whereas another found that mean reaction times were longer (but not more variable) during on-the-road driving (Wester, 2009). It should be noted that our reaction time analysis was based on only the 9 subjects in the active condition. Observed power to detect any reaction time differences (for mean reaction time) between driving conditions as reported by SPSS was only 0.384. This may be contrasted to the observed power to detect driving-condition effects on fP3 across the two groups, which was 0.985. To attain more information on the likelihood of not rejecting H0 while H1 was actually true, an additional Bayesian analysis was conducted using JASP (JASP Team, 2018). For the analysis of mean reaction times, the analysis revealed a Bayes factor of 2.6, implicating that the alternative hypothesis (i.e., that driving condition has an effect on mean reaction time) is 2.6 times more likely compared to the H0 hypothesis (i.e., that driving condition has no effect on response time). A Bayes Factor between 1 and 3 is seen as anecdotal evidence (Jeffreys (1961), see also Chapter 7 in Lee & Wagenmakers (2013)). A subsequent Bayesian post-hoc test revealed Bayes factors of 2.3 for stationary versus driving, 1.4 for autonomous versus driving, and 0.4 for autonomous versus stationary. Note that Bayes factors smaller than 2 indicate very low likelihood of rejecting the H0 even with substantial increase in power. Hence, the current analysis suggests that with higher power, a significant omnibus effect of driving condition on reaction time may still be observed, but that this would be entirely due to the difference between stationary and driving; apparently, differences between autonomous and the other conditions are not sufficiently robust across participants (and for the difference between autonomous and stationary it is 1 / 0.4 = 2.2 times more likely that the null hypothesis is true).

When we performed a similar Bayesian analysis for response time variability, we found a Bayes factor of 0.949, suggesting that our data provides no conclusive evidence of whether driving condition has an effect on response time variability (i.e., the H1 and the H0 have a comparable likelihood based on the current data).

2.4 Discussion

We investigated susceptibility to auditory signals during autonomous driving, compared to regular driving and being stationary. We recorded the frontal P3 (fP3) ERP (Event-Related Potential) component as a metric of susceptibility (Kenemans, 2015; Wester 2009; Wester et al., 2008; Wester et al., 2010). Consistent with this earlier work, we found that fP3 was reduced when participants were driving compared to when they were stationary, suggesting that susceptibility to novel stimuli is reduced when driving (Kenemans, 2015). The novel finding is that under autonomous driving conditions the fP3 is also reduced compared to stationary (cf. Wester et al., 2008).

In our introduction, we mentioned three potential outcomes of our study: The susceptibility in automated driving can be (1) reduced, similar to the observed reduction in non-autonomous driving, (2) unabated with respect to a non-driving situation, similar to stationary, or (3) lie somewhere in between these extremes as autonomous driving has characteristics of both driving and non-driving. Taken together, our results rule out alternative 2 and lend most support for alternative 3: there was a significant difference between stationary and autonomous condition, but the effect size of this effect was smaller compared to the difference between regular driving and stationary.

The results of our study also suggest that susceptibility to auditory stimuli is in general better (i.e., frontal P3 components with a larger amplitude) in cases where participants are required to actively respond to some of the task (active condition) compared to when they do not need to respond to such a signal. This improvement was measured as a main effect, and therefore holds for being stationary, when being driving autonomously, and when driving oneself.

The reduced susceptibility during both regular driving and autonomous driving is consistent with the perceptual-load perspective (Lavie, 1995, 2005; Murphy et al., 2017), implying that being sufficiently engaged in an on-going task (i.e., driving yourself, or being driven) reduces the extent of processing of task-irrelevant stimuli. As in the Wester studies (Wester 2009; Wester et al., 2008; Wester et al., 2010), we also found that fP3-indexed susceptibility was enhanced again by making auditory information relevant to a task as well (specifically, fP3 to novels was enhanced by having subjects actively indicating the detection of non-novel target deviants). This is potentially important, as it suggests that susceptibility for very infrequent alerting signals can be enhanced by adding (in a real driving situation) a more or less continuous auditory task. As reported in (Wester 2009; Wester et al., 2008; Wester et al., 2010), introduction of the (active) oddball task has no effect on aspects of basic vehicle control (weaving and speed). Future research must reveal whether this also holds for the tactical and strategical aspects of driving, which become more important in cases where a (semi-) autonomous vehicle takes over mostly the basic vehicle control (e.g., adaptive cruise control and maintaining lateral position in SAE level-2 automation).

The difference between our active and passive conditions may also be compared to that between SAE level 3 of automation in which a response is required (comparable to our active condition), and SAE level 4 in which a response to an alert is optional (comparable to our passive condition). Our results suggest that susceptibility to SAE level 4 alerts may be reduced (cf. our results in the passive condition). Therefore, even though auditory alerts might be provided in a SAE level 4 automated vehicle, their passive nature might make them not that effective. Further empirical research is needed to test this prediction.

It is an empirical question what underlies the better susceptibility in the active condition compared to the passive condition. Possibly, performing a task with respect to a specific auditory event (the deviant target) yields a higher prioritization for auditory information in general (including the novels). In turn, this results in allocation of apparently still free processing capacity (manifest in fP3), without interfering with basic vehicle control. A further possibility is that engaging in an additional task reduces underload (Young & Stanton, 2002). More data and theory are needed to disentangle these hypotheses.

In the active condition, reaction times to the deviant target were shortest in the stationary condition, followed by autonomous, followed by regular driving. However, our MANOVA revealed no main effect, and a Bayesian analysis revealed that although the trend is that there might be a significant effect of driving condition on reaction time, the evidence is not enough to draw solid conclusions. We also did not find a significant effect of driving condition on response time variability.

These results differ slightly from the findings by Wester and colleagues. In their studies, manual response time to the deviant tones were more variable (but not longer) during simulated driving (Wester et al., 2008) and longer (but not more variable) during on-the-road driving (Wester, 2009).

Pulling it all together, it appears that adding a more or less continuous auditory task increases fP3, while fP3 is still smaller during autonomous driving than during stationary. At the same time, within that task, participants show no appreciable decline performance from the stationary to the autonomous conditions. This can be viewed as a dissociation between on the one hand, the ability to detect potentially relevant auditory information (the novels), and on the other, the ability to react to unambiguously relevant auditory information (the deviant targets). Consistently, based on target-elicited ERP results, Wester (Wester, 2009, p. 133) concluded that there may have been “an increased investment of effort to compensate for the increased task-load imposed on the oddball task during driving compared to non-driving."

2.4.1 Implications for practice

Regarding the implications for practice, our results provide four reasons for some caution in the use of auditory alerts in semi-autonomous vehicles as a cue for handover (also referred to as a transition of control). First, the susceptibility in the autonomous condition was significantly lower compared to the stationary condition, thereby suggesting that people's susceptibility to audio signals is reduced. The implication is that there may be a need for more research and development with respect to the design of alerts. For example, in our own work, we have demonstrated how early warnings can be beneficial to allow for a smooth handover between vehicle and driver in level 3 automation (Van der Heiden, Iqbal, & Janssen, 2017).

Second, adding an auditory task seems to enhance susceptibility to some audio signals, but might negatively affect reaction speed (i.e., if the trend that our Bayesian analysis detected persists in future studies). This dissociation needs to be resolved before the current results can be applied in real-life contexts.

Third, our study took place in a controlled setting, where there was only the driving task and the oddball task. However, in their own vehicles, people might do other tasks than driving. More specifically, drivers distract themselves with various non-driving related tasks while driving (Dingus et al., 2016; Klauer et al., 2014), as they do in other non-driving settings (Janssen, Gould, Li, Brumby, & Cox, 2015). Moreover, a recent literature review suggests that as automation of the vehicle increases, people perform non-driving tasks even more frequently, which results in less awareness of the driving environment, and slower response times (De Winter, Happee, Martens, & Stanton, 2014). It is an empirical question whether the susceptibility to audio signals is reduced under conditions of distraction while driving. We plan to study the impact of multitasking on susceptibility.

Fourth, and finally, as argued above, our results suggest that susceptibility to SAE level 4 auditory alerts may be reduced. Therefore, designers might reconsider whether such an alert is effective at all. Not using (auditory) alerts seems a too bold conclusion based on our data. However, at the least, our data suggests that careful testing of these alerts is necessary before they are used in actual cars.

2.4.2 Limitations and future work

Our study was conducted in a driving simulator and therefore suffers from the known limitations of such simulators such as that participants perhaps take more risks compared to when they drive in everyday traffic outside. Nonetheless, previous work by Wester (2009), Chapter 6, has demonstrated that at least in conventional driving, reductions in fP3 that are observed in the simulator setting were also present when measured on the road.

We have studied susceptibility to auditory alerts in a setting without additional distractions. However, given the prevalence of in-car distractions (Dingus et al., 2016; Klauer et al., 2014), including in more automated systems (De Winter et al., 2014), we plan to also study how susceptibility changes when the driver can be distracted. These issues add to other important aspects of driver safety and experience (e.g., Kun, 2018; Kun, Boll, & Schmidt, 2016) including how to avoid confusion about responsibilities in vehicle automation (e.g., Janssen, Boyle, Kun, Ju, & Chuang, 2018), and how to drive efficient and sustainable (e.g., Hu, Wang, & Tang, 2017).With the new gained basic understanding of auditory alert processing, other work can also investigate how the exact type of audio signal impacts susceptibility, which is a large effort given the variety of auditory alerts for in-car settings (for a review of alert types see Nees & Walker (2011)). Moreover, future work can look at alerts in other modalities such as visual (e.g., see Van der Heiden, Janssen, Donker, & Merkx (2018) for an example of last-minute visual alerts) and of multimodal signals (e.g., Politis, Brewster, & Pollick, 2014). Finally, the general limitations of signal processing can be investigated, including the impact of competing tones and of false alarms on susceptibility and reaction time (e.g., Bliss & Acton, 2003; Parasuraman & Riley, 2016; Sorkin, Kantowitz, & Kantowitz, 2016). In the end, providing a signal is not a guarantee that the human acts on it.

2.4.3 Conclusion

Human susceptibility to auditory signals, as measured by the fP3 (frontal P3), is reduced when driving, but also when being driven autonomously, as compared to a stationary situation. As all levels of autonomous driving, except full autonomy, have some level of shared control between the human and the system, where (auditory) signals might play a role, it is important to consider the limitations of the human brain in terms of susceptibility to auditory alerts. One possible venue to enhancing susceptibility to infrequent auditory events (e.g., handover alerts) lies in the implementation of a suitable additional auditory task in the driving situation.

References

Dien, J., Spencer, K. M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal components analysis. Cognitive Brain Research, 17(3), 637-650

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. Proceedings of the National Academy of Sciences, 113(10), 2636-2641.

Fabiani, M., & Friedman, D. (1995). Changes in brain activity patterns in aging: the novelty oddball. Psychophysiology, 32(6), 579-594.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. Neuroscience & Biobehavioral Reviews, 25(4), 355-373.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. Electroencephalography and clinical neurophysiology, 55(4), 468-484.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Advances in psychology (Vol. 52, pp. 139-183). North-Holland.

Heitland, I., Kenemans, J. L., Oosting, R. S., Baas, J. M. P., & Böcker, K. B. E. (2013). Auditory event-related potentials (P3a, P3b) and genetic variants within the dopamine and serotonin system in healthy females. Behavioural brain research, 249, 55-64.

Hu, X., Wang, H., & Tang, X. (2017). Cyber-physical control for energy-saving vehicle following with connectivity. IEEE Transactions on industrial electronics, 64(11), 8578-8587.

Janssen, C. P., Boyle, L. N., Kun, A. L., Ju, W., & Chuang, L. L. (2019). A hidden markov framework to capture human–machine interaction in automated vehicles. *International Journal of Human–Computer Interaction*, *35*(11), 947-955.

Janssen, C. P., Gould, S. J., Li, S. Y., Brumby, D. P., & Cox, A. L. (2015). Integrating knowledge of multitasking and interruptions across different perspectives and research methods.

JASP Team. (2018). JASP (Version 0.8. 6)[Computer software].

Jeffreys, H. (1961). *Theory of probability*, 3rd edn oxford: Oxford university press.

Kenemans, J. L. (2015). Specific proactive and generic reactive inhibition. *Neuroscience & Biobehavioral Reviews*, *56*, 115-126.

Khushaba, R. N., Kodagoda, S., Lal, S., & Dissanayake, G. (2011, September). Intelligent driver drowsiness detection system using uncorrelated fuzzy locality preserving analysis. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 4608-4614). IEEE.

Kiehl, K. A., Laurens, K. R., Duty, T. L., Forster, B. B., & Liddle, P. F. (2001). Neural sources involved in auditory target detection and novelty processing: an event-related fMRI study. *Psychophysiology*, *38*(1), 133-142

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England journal of medicine*, *370*(1), 54-59.

Kun, A. L. (2018). Human-machine interaction for vehicles: Review and outlook. *Foundations and Trends® in Human–Computer Interaction*, *11*(4), 201-293.

Kun, A. L., Boll, S., & Schmidt, A. (2016). Shifting gears: User interfaces in the age of autonomous driving. *IEEE Pervasive Computing*, *15*(1), 32-38.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 451.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in cognitive sciences*, *9*(2), 75-82.

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.

Luck, S. J. (2014). An introduction to the event-related potential technique. MIT press.

Massar, S. A., Wester, A. E., Volkerts, E. R., & Kenemans, J. L. (2010). Manipulation specific effects of mental fatigue: evidence from novelty processing and simulated driving. *Psychophysiology*, *47*(6), 1119-1126.

Math, R., Mahr, A., Moniri, M. M., & Müller, C. (2013). OpenDS: A new open-source driving simulator for research. *GMM-Fachbericht-AmE 2013*, 2.

Michon, J. A. (1985). A critical view of driver behavior models: what do we know, what should we do?. In *Human behavior and traffic safety* (pp. 485-524). Springer, Boston, MA.

Murphy, S., Spence, C., & Dalton, P. (2017). Auditory perceptual load: A review. *Hearing Research*, *352*, 40-48.

Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. *Reviews of human factors and ergonomics*, *7*(1), 58-99.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, *39*(2), 230-253.

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*(2), 140-160.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148.

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

SAE International. (2018). j3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. Retrieved from

Scheer, M., Bülthoff, H. H., & Chuang, L. L. (2016). Steering demands diminish the early-P3, late-P3 and RON components of the event-related potential of task-irrelevant environmental sounds. *Frontiers in human neuroscience*, *10*, 73.

Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention*, *41*(5), 1087-1093.

Schröger, E., & Wolff, C. (1998). Attentional orienting and reorienting is indicated by human event-related brain potentials. *Neuroreport*, *9*(15), 3355-3358.

Sharbrough, F. C. G. E. (1991). American Electroencephalographic Society guidelines for standard electrode position nomenclature. *J clin Neurophysiol*, *8*, 200-202.

Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors*, *30*(4), 445-459.

Van der Heiden, R. M., Iqbal, S. T., & Janssen, C. P. (2017, May). Priming drivers before handover in semi-autonomous cars. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 392-404).

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., & Merkx, C. L. (2019). Visual in-car warnings: How fast do drivers respond?. *Transportation research part F: traffic psychology and behaviour*, *65*, 748-759.

Wertheim, A. H. (1978). Explaining highway hypnosis: Experimental evidence for the role of eye movements. *Accident Analysis & Prevention*, *10*(2), 111-129.

Wessel, J. R., & Aron, A. R. (2017). On the globality of motor suppression: unexpected events and their influence on behavior and cognition. *Neuron*, *93*(2), 259-280.

Wester AE. Attention Deficit and Impulsivity: Driving, Drugs and Electrophysiology. 2009. Utrecht: Utrecht University. PhD thesis.

Wester, A. E., Böcker, K. B., Volkerts, E. R., Verster, J. C., & Kenemans, J. L. (2008). Event-related potentials and secondary task performance during simulated driving. *Accident Analysis & Prevention*, *40*(1), 1-7.

Wester, A. E., Verster, J. C., Volkerts, E. R., Böcker, K. B., & Kenemans, J. L. (2010). Effects of alcohol on attention orienting and dual-task performance during simulated driving: An event-related potential study. *Journal of psychopharmacology*, *24*(9), 1333-1348.

Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Human factors*, *44*(3), 365-375.

Chapter 3
The influence of cognitive load on susceptibility to audio

Remo M. A. van der Heiden
Christian P. Janssen
Stella F. Donker
J.Leon Kenemans

A modified version has been published as:
Van der Heiden, R. M., Janssen, C. P., Donker, S. F., & Kenemans, J. L. (2020). The influence of cognitive load on susceptibility to audio. *Acta psychologica*, *205*, 103058. DOI: https://doi.org/10.1016/j.actpsy.2020.103058

Author contributions:
RMAV designed and programmed the experiment, collected the data, analyzed the data, and wrote the manuscript with input from the other authors. CPJ, SFD, and JLK supervised the project, contributed to experimental design, interpretation of the data, and editing the manuscript. All authors read and approved the final version of the manuscript.

Abstract

In this study we evaluate how cognitive load affects susceptibility to auditory signals. Previous research has used the frontal P3 (fP3) event related potential response to auditory novel stimuli as an index for susceptibility to auditory signals. This work demonstrated that tasks that induce cognitive load such as visual and manual tasks, reduced susceptibility. It is however unknown whether cognitive load without visual or manual components also reduces susceptibility. To investigate this, we induced cognitive load by means of the verb generation task, in which participants need to think about a verb that matches a noun. The susceptibility to auditory signals was measured by recording the event related potential in response to a successively presented oddball probe stimulus at 3 different inter-stimulus intervals, 0 ms, 200 ms or 400 ms after the offset of the noun from the verb generation task. An additional control baseline condition, in which oddball response was probed without a verb generation task, was also included. Results show that the cognitive load associated with the verb task reduces fP3 response (and associated auditory signal susceptibility) compared to baseline, independent of presentation interval. This suggests that not only visual and motor processing, but also cognitive load without visual or manual components, can reduce susceptibility to auditory signals and alerts.

3.1 Introduction

Cognitive load induced by a task can have a negative impact on performance of other tasks. For example, in everyday life, the cognitive load induced by contributing to a phone conversation can degrade driving performance (e.g., Brookhuis, De Vries, & De Waard,1991; Horrey & Wickens 2006). Cognitive load has therefore been studied in detail in the cognitive (neuro-) science literature. Perhaps the best-known task is the psychological refractory period task (e.g., Anderson, Taatgen, & Byrne, 2005; Howes, Lewis, & Vera, 2009; Meyer & Kieras, 1997; Pashler, 1994; Schumacher et al., 1999). The consistent finding in these studies is that processing one task (e.g., a visual-manual task) can impact the processing time of another task (e.g., auditory-vocal task), and that the impact depends on how far apart the presentation of stimuli is.

Another characteristic of the psychological refractory period task, and many other dual-task experiments, is that it requires a participant to process and respond to two tasks. Although fruitful, the exact response pattern and the nature of cognitive load is then dependent on the characteristics of both tasks. For example, whether both tasks are presented in the same or different modalities (e.g., Wickens, 2002, 2008), what difficulty level each of the tasks has (e.g., Lavie, 1995, 2005), and how practiced one is at the task (e.g., Engström, Markkula, Victor, & Merat, 2017). Given the complex interactions of these factors and the need to understand the context of testing, dual-task experiments are limited in their generalizability to measure the cognitive load of a single task.

As an alternative, the load of one task might also be measured through a more passive probe, for which no physical response is required. A promising approach is the use of an auditory novelty oddball paradigm, which has previously been successfully used to measure workload (e.g., Allison & Polich, 2008; Miller, Rietschel, McDonald, & Hatfield, 2011; Ullsperger, Freude, & Erdmann, 2001). In the novelty oddball probe technique, participants are presented with frequent stimuli, called standards, which are interspersed with rare stimuli, or novels (Friedman, Cycowicz, & Gaeta, 2001; Polich, 2007). For example, a common auditory version of the oddball task which we also use in our study plays 1000 Hz tones to the participant as standards, and occasionally intersperses these with unique environmental sounds such as coughs, sneezes, and bark sounds (Fabiani & Friedman, 1995). The consistent finding in the literature is that presentation of a novel stimulus creates a positive peak in electrical activity at the frontal central region of the brain, which peaks around 300 ms after stimulus/probe onset. This resulting peak is also known as the frontal P3 or fP3.

The common interpretation in the literature is that the fP3 peak is related to susceptibility towards new information, and to a process of orienting to novel stimuli (Friedman et al., 2001; Polich, 2007; for a review see: Kenemans, 2015). Moreover, previous research found that the fP3 peak is reduced when people perform visual-manual tasks, such as driving or manual tracking (e.g., Scheer, Bülthoff, & Chuang, 2016, 2018; Van der Heiden et al., 2018; Wester, Böcker, Volkerts, Verster, & Kenemans, 2008), which is interpreted as reflecting a reduction in susceptibility for unexpected but potentially relevant information (cf. Friedman et al., 2001; Kenemans, 2015; Polich, 2007).

The aforementioned studies demonstrated the effect of a concurrent task on susceptibility to novel auditory stimuli. However, these studies involved tasks that had visual stimuli or that required manual action (Scheer et al., 2016, 2018; Van der Heiden et al., 2018; Wester et al., 2008). Moreover, even previous oddball (fP3) studies that manipulated cognitive load all required visual stimuli and/or manual action (e.g., Allison & Polich, 2008; Miller et al., 2011; Ullsperger, Freude & Erdmann, 2001). It is therefore unclear whether the observed lower susceptibility to novel stimuli only occurs when visual or manual processing is involved. Would this same reduction in novelty processing also be present when the task does not involve manual or visual interaction? We address this question in the current study.

Since many tasks that people perform in daily life can create cognitive load, being able to measure the effect of cognitive load in the absence of visual and manual input/tasks on performance allows for a better general understanding of mental processing induced performance degradation. In particular, it is relevant to understand how such load affects susceptibility to sounds. For example, while driving to the supermarket, one might be thinking about what ingredients are needed to cook a meal (i.e., inducing cognitive load), which in turn might influence the ability to notice an unexpected signal, such as a car horn or blind spot warning, and to then react appropriately. Understanding correct perception of and reaction to unexpected signals is even more urgent today, given the rise of automated systems in which humans and (semi-automated) machines perform tasks together (Janssen, Donker, Brumby, & Kun, 2019), and in which humans form a crucial backup to respond to system alerts (Janssen, Iqbal, Kun, & Donker, 2019).

To investigate whether cognitive load in the absence of visual and manual input/tasks also reduces susceptibility to novel processing, we will induce cognitive load using the verb generation task (Abdullaev & Posner, 1998; Bijl, De Bruin, Böcker, Kenemans, & Verbaten, 2007; Snyder, Abdullaev, Posner, & Raichle, 1995). In this task, participants hear or see a noun, and in response either need to *generate* an appropriate verb for that word (e.g., apple-bite), or *repeat* the noun (e.g., apple-apple). Generating verbs in response to a noun (compared to repeating nouns) has been associated with higher cognitive load (Abdullaev & Posner, 1998; Snyder et al., 1995), with increased frontal activity in the cortex (Abdullaev & Posner, 1998; Bijl et al., 2007), and with increased dual-task interference (e.g., Iqbal, Ju, & Horvitz, 2010; Kunar, Carter, Cohen, & Horowitz, 2008; Strayer & Johnston, 2001; Van der Heiden et al., 2018). The common interpretation in the literature is therefore that the generate condition of the verb task induces relatively high cognitive load. We will therefore use the auditory presented generate version of the task to induce cognitive load.

To assess the induced cognitive load while one devises a verb in the verb task, we present an oddball stimulus as a probe and record the subsequent brain activity with an electroencephalogram (EEG). By averaging multiple response measurements to an event related potential (ERP), a high signal to noise ratio can be achieved (e.g., Glatz, Krupenia, Bülthoff & Chuang 2018; Squires, Squires & Hillyard, 1975; Van der Heiden et al., 2018; Wester et al., 2008). Within the verbal response time epoch, a verbal task goes through different stages of mental processing (Abdullaev & Posner, 1998; Bijl et al., 2007; Snyder et al., 1995).To be able to make communication more effective, for example to find the most effective moment to play an auditory warning, it is important to understand the cognitive load over time. To assess this, we manipulated the timing between presentation of the noun and presentation of the probe, akin to manipulations in psychological refractory period task studies (e.g., Anderson, Taatgen, & Byrne, 2005; Howes, Lewis, & Vera, 2009; Meyer & Kieras, 1997; Pashler, 1994; Schumacher et al., 1999).

To summarize, previous research has observed a reduced susceptibility to auditory novel stimuli, which has been interpreted as reduced ability to allocate mental resources to novel stimuli in the environment (Kenemans, 2015). Our research question is whether cognitive load in the absence of visual stimulation and/or required manual action, as created through the verb generation task (Abdullaev & Posner, 1998), also reduces susceptibility towards auditory stimuli. We test this by recording the novel- elicited response following a verb generation task stimulus. As verb generation is a mental process that goes through different stages (Abdullaev & Posner, 1998, Bijl et al., 2007) with already early activity around electrode Fz (Abdullaev & Posner, 1998), we test susceptibility at various inter-stimulus intervals: 0, 200, and 400 ms after noun offset. We compare this to a baseline condition with no noun present. If cognitive load in the absence of visual stimulation and/or required manual action also reduces susceptibility, then we expect to find a fP3 response in the verb generation condition which is reduced compared to the baseline. In line with the psychological refractory period literature, we would also expect to find a higher interference at briefer inter-stimulus intervals.

3.2 Method

3.2.1 Participants

13 participants (8 M; 5 F) were recruited through social media using opportunity sampling (age 19 to 28 years of age, M = 23, SD = 2.6 years of age). All speakers were fluent in Dutch, 12 were native speakers, the one non-native participant spoke Dutch for 16 years. All participants reported normal hearing. The experiment was approved by the ethics committee of the Faculty of Social and Behavioral Sciences of Utrecht University (approval number FETC16-042). All participants gave written informed consent prior to the experiment. Participants were compensated €12 for their time.

3.2.2 Materials

Presentation of auditory stimuli

We used two types of auditory stimuli: an oddball probe and a verb generation task. All auditory stimuli were presented using Presentation (Neurobehavioural Systems) to participants through Earlink earphones at a constant level of 75 dB. Oddball stimuli were presented binaurally, word stimuli for the verb generation task only monaural through the earphone in the right ear.

Oddball probe

We used a 2-stimulus novelty oddball probe (Polich, 2007) in which 80% of the stimuli consisted of a 1000 Hz tone which was presented for 400 ms (labelled: standards). For the remaining 20% of the stimuli (labelled: novels) we presented novel sounds that were obtained from a database by Fabiani et al. (1995). This database consists of 100 unique environmental sounds, such as for example a person sneezing or a dog barking. The duration of the novels varied from 159 ms to 399 ms. All novels and standards were presented at 75 dB.

Verb generation task

In the verb generation task, participants heard a noun in a subset of the trials, and were asked to respond verbally with a verb that would fit the noun they heard (cf. Abdullaev & Posner, 1998). For example, when participants heard "apple" they could reply with "eat" or any other verb that they associated with apple.

As our participants were Dutch speakers, we created a Dutch word list. First, we translated the set of English nouns from Abdullaev and Posner (1998) to Dutch. Next, we removed all words that, as assessed by a research assistant and the researchers, were ambiguous in Dutch. As we wanted the words to be presented in a specific short interval, we also removed words that took long to pronounce, as assessed by having more than two syllables. We then converted the words to wave sound files using text-to-speech website www.texttospeech.io. We threw away words of which presentation took longer than 500 ms using default settings of the text-to-speech algorithm (Dutch female, volume 1, rate 1, pitch 1). The final set consisted of 144 unique words. Next, we determined the imaginability of each word, based on the scores in Van Loon-Vervoorn (1985). As the words differed in their imaginability score, we created 12 sets of 24 words each, such that for each set of words the mean and distribution of imaginability scores were comparable.

The next step was to play words at a constant length, while retaining intelligibility. Such constant length was needed to get the inter-stimulus timings between the verb task and the oddball probe exact. We conducted a small pilot on ourselves in which we changed the playback speed of all verb generation audio files, using Audacity (audacity.sourceforge.net). We eventually settled on a length of exactly 400 ms. This was the minimal length needed to have clear understandable word presentation, and also the same length as a standard tone in the oddball probe. For some words, this meant that they were played slightly faster or slower than the default text-to-speech algorithm. For example, the Dutch word “cel” (English: “cell”), originally took 450 ms to pronounce, which was brought back to 400 ms.

When participants performed the verb generation task in the experiment, our intent was for the word to be easily understandable, yet fast to present. Our intention was not to detect a specific response to a specific word. Therefore, we did not keep track of which words participants replied. We also encouraged participants to respond with the first word that came to mind, as we were only interested in the cognitive load inducing characteristics of the task, not in the actual words that participants came up with.

Intelligibility

To make sure participants were able to understand the words, we tested intelligibility of the words through a separate “repeat” task. In this task, all 144 words were played to the participants with 1600 ms between trials (i.e., offset word A and onset of word B). The task for the participants was to repeat the words. They were instructed to not reply in cases where they did not understand a word. Each participant received the words in the same order. During this stage, the experimenter made notes of mistakes and skipped words.

On average the words were correctly understood in 91% of the cases. The lowest participant score was 85% of words correct, and the highest 94%. There were no words consistently misinterpreted by all participants, rather the words that participants misinterpreted varied between participants.

Verb response reaction time

To be able to register the verbal response reaction time we placed two electrodes on the left cheek of the participants. Based on pilot tests, we set 40 µv for the average of both electrodes as a threshold for the detection of jaw movements, and associated speech. We then tested at what interval (relative to word presentation) jaw movements were detected, and whether such response times differed between experimental conditions.

Design

To measure the effects of the stage of word processing and associated cognitive interference on susceptibility to the oddball novel stimuli, we varied Inter-stimulus interval (ISI) between word offset (of the verb task) and onset of the oddball probes. We used a single factor within-subject design. *Inter-stimulus interval* was used as the factor with 4 levels: 0 ms, 200 ms, 400 ms, and Baseline. Figure 3.1 shows presentation times visually.

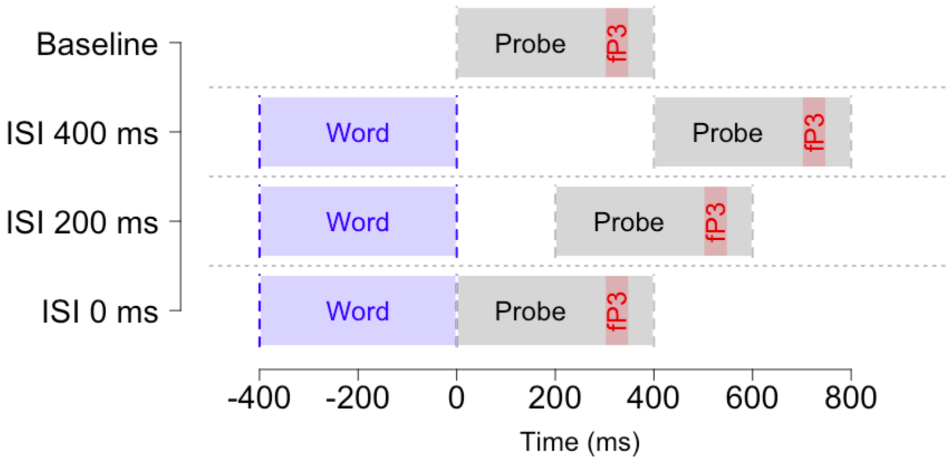


Figure 3.1. Schematic representation of (inter-stimulus interval, or ISI) timings of word (noun from verb generation task) and oddball probe presentation. fP3 response was measured (305-355 ms after probe onset).

Testing blocks

In total there were 12 experiment blocks. In each block, 80 oddball stimuli/trials (64 standards, 16 novels) were presented and 24 nouns were presented. Nouns for the verb generation task were followed by a standard tone of the oddball probe in half the cases (12 trials) and by a novel sound in the other half. One third of these combinations (i.e., 4 standards, 4 oddballs) was used with an ISI of 0 ms, one third with an ISI of 200 ms, and one third with an ISI of 400 ms. The position of these trials within the block was random. The remainder of the trials of a block was filled with presentation of 52 standard tones.

When all measurements of the 12 blocks are combined, each ISI condition had 48 words followed by a standard tone, and 48 words followed by a novel sound. For the baseline condition we also used 48 standard tones and 48 novel responses (which were not preceded by a noun from the verb task).

The interval between two subsequent standard tones was 2000 ms (cf. Van der Heiden et al., 2018; Wester et al., 2008). For trials where a noun was presented, the interval between noun and oddball was set by condition. The subsequent stimulus (either a noun or a standard tone) was presented 4400 ms after the onset of the noun. This longer interval was chosen to prevent any interference from generating a verb response.

3.2.3 Procedure

Participants first received general instructions and an overview of the experiment. They were then asked to read and sign the informed consent form. Subsequently, participants were asked to be seated in a driving simulator and adjust the seat to accommodate their feet and arms in an upright yet comfortable position[2]. Then intelligibility of the words was tested. We then applied 4 ocular electrodes to the face and 2 electrodes at left and right mastoid. A cap with 64 electrodes was placed on the participant's head and the signal from each electrode was inspected. Where needed, slight adjustments were made (e.g., adding extra conductive gel).

After receiving instructions, a practice block was used to allow participants to get accustomed with the verb generation and oddball probe. Twenty trials were played to the participant consisting of 16 standards and 4 novels, 6 of these stimuli were preceded by a word (3 standards, 3 novels). For these 6 stimuli, each ISI from the

[2] For this specific study, a driving simulator setup is not needed. However, some of our other interests are in testing human performance under driving and automated driving conditions (e.g., Van der Heiden et al., 2018). To make our results comparable to future driving studies, we used this similar set-up.

3 possibilities (0, 200, 400 ms) was played in combination with a standard and with a novel in random order.

The main experiment consisted of 12 blocks with a short break after 6 blocks. When all blocks were finished the participants were asked to fill out a short questionnaire on demographics, experience and feedback. The total length of the experiment was 2 hours.

3.2.4 Signal recording and analysis

EEG recordings were made using a BioSemi ActiveTwo system at 2048 Hz with 64 active Ag-AgCl electrodes, positioned following 10/10 system (Chatrian, Lettich, & Nelson, 1985). The average of left and right mastoids was used as reference. Horizontal and vertical EOG electrodes were placed on both outer canthi of the eyes, as well as above and below the right eye in line with the pupil.

EEG signal analysis was performed with Brain Vision Analyzer 2 software (Brain Products GmbH, München, Germany). First, data were re-referenced to the average of left and right mastoid signal. Then, three filters were applied: a 50 Hz notch filter to compensate for noise from the mains, a 0.16 Hz high-pass filter with a slope of 24 dB/oct, and a 30Hz low-pass filter with a slope of 24 dB/oct to remove frequency bands that were not of our interest (cf. Van der Heiden et al., 2018; Wester et al., 2008).

Next, ERP segments of 2500 ms were distilled starting 1000 ms before the oddball stimulus onset, and ending 1500 ms after. To correct for eye movements and blinks, an ocular correction was applied to every segment using the Gratton, Coles and Donchin method (Gratton, Coles, & Donchin, 1983). For all conditions, data were baseline corrected over the 100 ms interval preceding probe presentation. Subsequently, artifacts in individual channels were rejected. Criteria for rejection of an epoch were at least one of the following conditions: maximum voltage step > 120 μv/ms within 200 ms before or after events; maximum difference > 100 μv within 200 ms; minimum activity < 0.5 μv within 100 ms (standard settings in Brainvison Analyzer 2.1).

Finally, grand averages were calculated per probe stimulus type (standard and novel) per condition (Baseline, 0 ms, 200 ms, and 400 ms). Given words at certain ISI, potential ERP artifacts induced by these word stimuli should be equal or novel-elicited and standard-elicited ERPs. Therefore, to annihilate these artifacts *difference waves* were calculated by subtracting the average ERP for the standard tones from the average ERPs for the novels for each condition. Note that according to this logic, only the difference waves can have a meaningful interpretation, not the separate ERPs to novels and standards, respectively. Our analysis focused on the amplitude of the fP3 peak in these difference waves. We used a collapsed localizer that included all conditions to determine the location of the fP3 peak at electrode position FCz for this study (Luck & Gaspelin, 2017). This localizer showed the fP3 peak location was at 305 - 355 ms after probe onset.

Speech reaction time was calculated from the average response across the two electrodes on the cheek. Raw data were filtered by a 50 Hz notch filter to compensate for noise from the mains, a 10 Hz high-pass filter with a slope of 24 dB/oct, and a 60 Hz low-pass filter with a slope of 24 dB/oct. As a threshold to identify speech production we used 40 μv increase from baseline. For each participant in each condition we took the average speech reaction time from all the trials that met the criterium.

3.2.5 Statistical analysis

For statistical analysis, we used a repeated measures ANOVA to analyze the effect of inter-stimulus interval (0, 200, 400 ms and baseline) on fP3 response and speech reaction time. We used an alpha level of .05 for significance. Where needed, this was followed by holm-corrected pairwise comparisons.

3.3 Results

3.3.1 fP3

Figure 3.2 shows the difference waves (novel probe minus standard probe) for each *inter-stimulus interval*. Vertical lines indicate position of probe onset (0 ms) and fP3 peak. For statistical analysis we analyzed the mean fP3 value in the interval 305-355 ms after probe onset. The barplot in Figure 3.3 shows this mean peak activation for each condition.

There was a main effect of *inter-stimulus interval* on fP3 amplitude, $F(3,36) = 4.83$, $p = .006, \eta_p^2 = 0.28$. A holm-corrected post-hoc test showed that the baseline condition (M = 10.4 μv, SD = 4.7) had a significantly higher fP3 amplitude compared to all the other conditions in which the oddball presentation was preceded by a noun (all p's < .05). The amplitudes for the three different conditions where a noun was presented before the oddball did not differ significantly from each other: 0 ms ISI (M = 6 μv, SD = 5.1μv), 200 ms ISI (M = 5.8 μv, SD = 3.9 μv), and 400 ms ISI (M = 5.8 μv, SD = 6.3 μv). In other words: when an oddball probe was presented after presentation of a noun (i.e., when cognitive load was induced), the fP3 response was reduced compared to baseline (no cognitive load), irrespective of *when* the oddball was presented (0, 200, 400 ms ISI).

Figure 3.4 shows activation across the scalp over time. It makes clear that the difference in activity between novel and standard is indeed most pronounced in frontal areas of the brain, during the interval of our analysis, and that the activity is reduced in the conditions with cognitive load (0, 200, 400 ms condition) compared to the baseline condition.

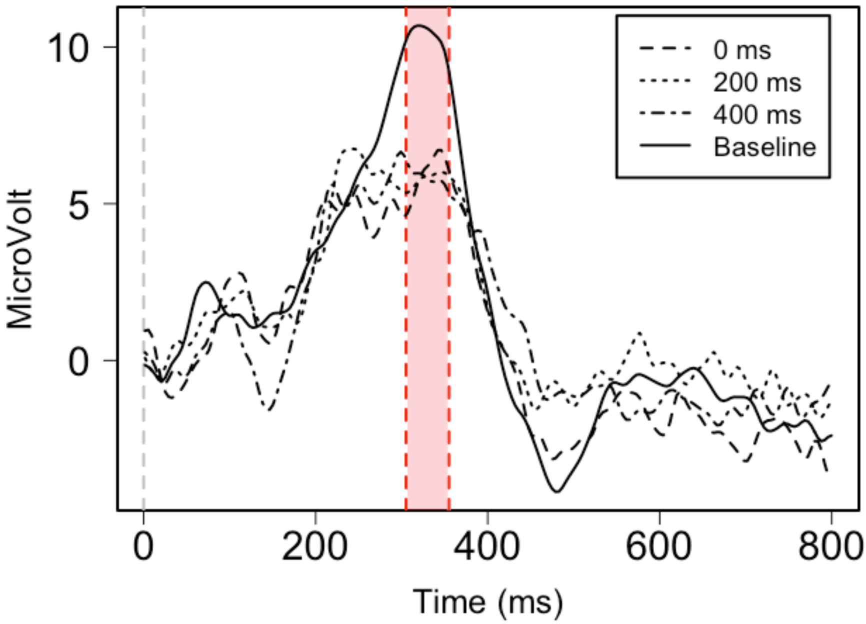


Figure 3.2. Event related potential of 3 ISI conditions and baseline. Vertical lines show onset of oddball stimulus (0 ms) and fP3 peak area used for statistical analysis (305-355 ms).

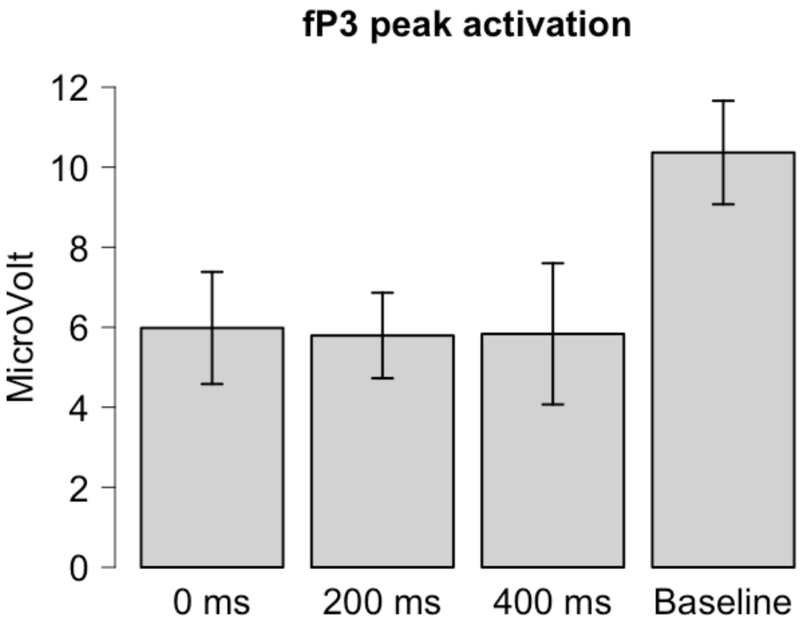


Figure 3.3. fP3 peak activation in 3 ISI conditions and baseline. Error bars show standard error of the mean.

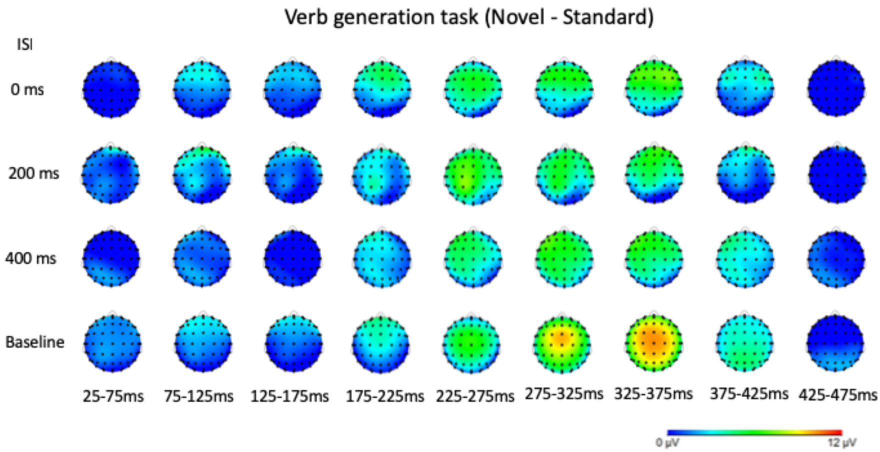


Figure 3.4. Top view of electrode positions and activation over time for ISI conditions and baseline. For each inter-stimulus interval (rows), in 50 ms epochs (columns) the mean difference activation is expressed as a heatmap, with blue being close to zero (i.e. no difference) and red being 12 µv or more difference between the standard and novel oddball probed activation.

3.3.2 Speech reaction time

We analyzed the interval between word presentation onset and onset of speaking, using two electrodes on the cheek (see materials). This measure of speech reaction time was analyzed to ensure that there were no differences in response time between nouns that were followed by a novel or by a standard oddball probe. We also analyzed whether this differed between inter-stimulus interval conditions. Results of a 2 (probe stimulus condition: novel or standard) x 3 (inter-stimulus interval: 0, 200, 400 ms) ANOVA showed that there was no significant difference between probe stimulus conditions (i.e., standard or novel), $F(1,12) = 0.056$, $p > .1$. However, there was a main effect of inter-stimulus interval, $F(2,24) = 30.68$ $p < .001$. A holm-corrected post-hoc test showed that reaction time differed between all three ISI conditions (all p's < .005). The reaction time was fastest for the 0 ms condition (M = 1300 ms, SD = 162 ms), followed by the 200 ms condition (M = 1440 ms, SD = 180 ms), followed by the 400 ms condition (M = 1520 ms, SD = 181 ms). This suggests some influence of ISI on the speech production process. There was no interaction effect between probe stimulus condition and inter-stimulus interval, $F(2,24) = 1.657$, $p > .1$. Further note that the lack of an effect of probe-stimulus condition on reaction time supports the idea that word processing was equal for novels versus standards, and therefore the ERP correlates of word processing can be annihilated by the novel-standard subtraction.

3.4 General discussion

We have tested whether cognitive load in the absence of visual stimulation and/or required manual action influences auditory susceptibility. Our experiment combined a cognitive load inducing task (verb generation task) with an auditory oddball novelty paradigm as probe task for susceptibility. Results show that auditory susceptibility to novel sounds is indeed reduced under conditions of cognitive load, independent of *when* the susceptibility is probed within an interval range of 0-400 ms after presentation of the load inducing stimulus.

Our results are consistent with previous work, which also found that concurrent task performance reduces susceptibility as measured through the novel response (e.g., Scheer et al., 2016, 2018; Van der Heiden et al., 2018; Wester et al., 2008). The result is also consistent with previous work that suggested that cognitive load can be measured with the oddball paradigm and fP3 measurements (Allison & Polich, 2008; Miller et al., 2011; Ullsperger, Freude & Erdmann, 2001). However, all these previous studies used tasks that involved visual stimuli and/or manual responses. As our work does not involve a visual or manual component, to the best of our knowledge it is the first to demonstrate that reduction in fP3 response can also occur without visual stimulation and/or required manual action, thereby strengthening the support for the claim that fP3 response measures the cognitive load associated with a task, and not merely visual- or manual engagement.

3.4.1 Limitations and future work

It is an open question whether the fP3 response can only be used to detect the presence or absence of cognitive workload (i.e., as a binary measure, as used in this paper), or whether it can also be used to detect various levels of workload. Although we have not tested this here, previous work has shown that fP3 response level can be used to differentiate conditions that differ on other aspects of demand such as being stationary, driving manually, or being driven (Van der Heiden et al., 2018) or difficulty level of a visual-manual control task (Scheer et al., 2016).

Although our study has shown that cognitive load reduces the fP3 response, it is less clear what cognitive process underlies this reduction. Within our tested interval (0-400 ms after word onset), there was no distinction between fP3 response levels. However, previous work does suggest that generating a word consists of multiple stages (Abdullaev & Posner, 2008; Bijl et al., 2007; Snyder et al., 1995). Further research is needed to distinguish the cognitive load inducing effects of each stage. For the moment we conclude that when a novel probe is presented anywhere between 400 ms and 800 ms after noun onset, the fP3 that it elicits is reduced relative to a no-verb task conditions irrespective of when exactly between 400 and 800 the probe was presented. This is consistent with a model in which verb-generation processing induces capacity limits (for novel susceptibility) starting at least as early as 700 ms after noun onset (fP3 latency in the 0-ms ISI condition), and lasting at least as long as until 1200 ms after noun onset.

The current fP3 assessment shows the influence of cognitive load on auditory susceptibility. However, no performance measures were applied in relation to the novel-oddball stimuli. Therefore, the ramifications of reduced auditory susceptibility for adequate behavioral reactions to pertinent, possibly relevant auditory stimuli outside the primary-task context, remains a matter of speculation. An interesting option would be to use a three-stimulus oddball set-up including a repeated (e.g., 10% of the trials) deviant pitch, in which participants need to react to the deviant tone (see e.g., Scheer et al., 2016, 2018; Van der Heiden et al., 2018; Wester et al., 2008). These previous studies have found some indications for longer deviant-detection latencies under driving than under stationary conditions. In addition, adding a deviant-related task as such increased fP3 amplitudes in driving conditions to the level of stationary conditions. It would be interesting to find out whether these same dependencies hold when a task like verb generation is added.

3.4.2 Conclusion

Cognitive load without concurrent visual or manual interaction reduces auditory susceptibility, as assessed using a fP3 response. The time between load inducing stimulus and auditory susceptibility probe did not influence this reduction, within the interval we measured (0-400 ms). This suggests that the cognitive process of generating a word provides a comparable level of cognitive load during this interval. This is consistent with a model in which verb-generation processing induces capacity limits (for novel susceptibility) starting at least as early as 700 ms after noun onset (fP3 latency in the 0-ms ISI condition), and lasting at least as long as until 1200 ms after noun onset.

References

Abdullaev, Y. G., & Posner, M. I. (1998). Event-related brain potential imaging of semantic encoding during processing single words. *Neuroimage*, 7(1), 1-13. https://doi.org/10.1006/nimg.1997.0309

Allison, B. Z., & Polich, J. (2008). Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biological psychology*, 77(3), 277-283. https://doi.org/10.1016/j.biopsycho.2007.10.014

Anderson, J. R., Taatgen, N. A., & Byrne, M. D. (2005). Learning to achieve perfect time sharing: Architectural implications of Hazeltine, Teague, and Ivry (2002). *Journal of Experimental Psychology: Human Perception and Performance, 31,* 749–761. https://doi.org/10.1037/0096-1523.31.4.749

Bijl, S., De Bruin, E. A., Böcker, K. E., Kenemans, J. L., & Verbaten, M. N. (2007). Effects of chronic drinking on verb generation: an event related potential study. *Human Psychopharmacology: Clinical and Experimental*, 22(3), 157-166. https://doi.org/10.1002/hup.835

Brookhuis, K. A., De Vries, G., & De Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention*, 23(4), 309-316. https://doi.org/10.1016/0001-4575(91)90008-S

Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *American Journal of EEG technology*, 25(2), 83-92. https://doi.org/10.1080/00029238.1985.11080163

Engström, J., Markkula, G., Victor, T. W., & Merat, N. (2017). Effects of Cognitive Load on Driving Performance: The Cognitive Control Hypothesis. *Human Factors*, *59*(5), 734–764. https://doi.org/10.1177%2F0018720817690639

Fabiani, M., & Friedman, D. (1995). Changes in brain activity patterns in aging: the novelty oddball. *Psychophysiology*, *32*(6), 579-594. https://doi.org/10.1111/j.1469-8986.1995.tb01234.x

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, 25(4), 355-373. https://doi.org/10.1016/S0149-7634(01)00019-7

Glatz, C., Krupenia, S. S., Bülthoff, H. H., & Chuang, L. L. (2018). Use the Right Sound for the Right Job: Verbal Commands and Auditory Icons for a Task-Management System Favor Different Information Processes in the Brain. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (paper 472). New York, NY: ACM. https://doi.org/10.1145/3173574.3174046

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology*, *55*(4), 468-484. https://doi.org/10.1016/0013-4694(83)90135-9

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human factors*, 48(1), 196-205. https://doi.org/10.1518%2F001872006776412135

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review*, 717–751. https://doi.org/10.1037/a0017187

Iqbal, S. T., Ju, Y. C., & Horvitz, E. (2010). Cars, calls, and cognition: investigating driving and divided attention. *In Proceedings of the SIGCHI conference on human factors in computing systems*(pp. 1281-1290). New York, NY: ACM. https://doi.org/10.1145/1753326.1753518

Janssen, C. P., Donker, S. F., Brumby, D. P., & Kun, A. L. (2019). History and future of human-automation interaction. *International Journal of Human-Computer Studies, 131*, 99-107. https://doi.org/10.1016/j.ijhcs.2019.05.006

Janssen, C. P., Iqbal, S. T., Kun, A. L., & Donker, S. F. (2019). Interrupted by my car? Implications of interruption and interleaving research for automated vehicles. *International Journal of Human-Computer Studies, 130*, 221-233. https://doi.org/10.1016/j.ijhcs.2019.07.004

Janssen, C.P., Van der Heiden, R.M., Donker, S.F., Kenemans, J.L. (2019). Measuring Susceptibility to Alerts While Encountering Mental Workload. In: Adjunct *proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. New York, NY: ACM. https://doi.org/10.1145/3349263.3351524

Kenemans, J. L. (2015). Specific proactive and generic reactive inhibition. *Neuroscience & Biobehavioral Reviews*, *56*, 115-126. https://doi.org/10.1016/0013-4694(75)90263-1

Kunar, M. A., Carter, R., Cohen, M., & Horowitz, T. S. (2008). Telephone conversation impairs sustained visual attention via a central bottleneck. *Psychonomic bulletin & review*, 15(6), 1135-1140. https://doi.org/10.3758/PBR.15.6.1135

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Learning and Memory*, *21*(3), 451–468. http://doi.org/10.1037/0096-1523.21.3.451

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82. http://doi.org/10.1016/j.tics.2004.12.004

Levy, J., Pashler, H., & Boer, E. (2006). Central interference in driving: Is there any stopping the psychological refractory period? *Psychological Science*, *17*(3), 228-235. https://doi.org/10.1111%2Fj.1467-9280.2006.01690.x

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, 54(1), 146-157. https://doi.org/10.1111/psyp.12639

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance .2. Accounts of psychological refractory-period phenomena. *Psychological Review*, *104*(4), 749–791.

Miller, M. W., Rietschel, J. C., McDonald, C. G., & Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. *International Journal of Psychophysiology*, 80(1), 75-78. https://doi.org/10.1016/j.ijpsycho.2011.02.003

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin*, *116*(2), 220.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148. https://doi.org/10.1016/j.clinph.2007.04.019

Scheer, M., Bülthoff, H. H., & Chuang, L. L. (2016). Steering demands diminish the early-P3, late-P3 and RON components of the event-related potential of task-irrelevant environmental sounds. *Frontiers in human neuroscience*, 10, 73. https://doi.org/10.3389/fnhum.2016.00073

Scheer, M., Bülthoff, H. H., & Chuang, L. L. (2018). Auditory Task Irrelevance: A Basis for Inattentional Deafness. *Human factors*, 60(3), 428-440. https://doi.org/10.1177%2F0018720818760919

Schumacher, E. H., Lauber, E. J., Glass, J. M., Zurbriggen, E., Gmeindl, L., Kieras, D. E., & Meyer, D. E. (1999). Concurrent response-selection processes in dual-task performance: Evidence for adaptive executive control of task scheduling. *Journal of Experimental Psychology Human Perception and Performance*, *25*(3), 791–814.

Snyder, A. Z., Abdullaev, Y. G., Posner, M. I., & Raichle, M. E. (1995). Scalp electrical potentials reflect regional cerebral blood flow responses during processing of written words. *Proceedings of the National Academy of Sciences*, 92(5), 1689–1693. http://doi.org/10.1073/pnas.92.5.1689

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and clinical neurophysiology*, *38*(4), 387-401.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological science,* 12(6), 462-466. https://doi.org/10.1111%2F1467-9280.00386

Ullsperger, P., Freude, G., & Erdmann, U. (2001). Auditory probe sensitivity to mental workload changes–an event-related potential study. *International Journal of Psychophysiology*, 40(3), 201-209. https://doi.org/10.1016/S0167-8760(00)00188-4

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., Hardeman, L. E., Mans, K., & Kenemans, J. L. (2018). Susceptibility to audio signals during autonomous driving. *PloS one*, *13*(8), e0201963. https://doi.org/10.1371/journal.pone.0201963

Van Loon-Vervoorn, W. A. (1985). Voorstelbaarheidswaarden van Nederlandse woorden: 4600 substantieven, 1000 verba en 500 adjectieven. Swets & Zeitlinger.

Wester, A. E., Böcker, K. B. E., Volkerts, E. R., Verster, J. C., & Kenemans, J. L. (2008). Event-related potentials and secondary task performance during simulated driving. *Accident Analysis & Prevention*, *40*(1), 1-7. https://doi.org/10.1016/j.aap.2007.02.014

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors*, *50*(3), 449–455. http://doi.org/10.1518/001872008X288394

Chapter 4

The effect of cognitive load on auditory susceptibility during automated driving

Remo M. A. van der Heiden
Christian P. Janssen
Stella F. Donker
J. Leon Kenemans

Under review

Author contributions:
RMAV designed and programmed the experiment, collected the data, analyzed the data, and wrote the manuscript with input from the other authors. CPJ, SFD, and JLK supervised the project, contributed to experimental design, interpretation of the data, and editing the manuscript. All authors read and approved the final version of the manuscript.

Abstract

Objective: We experimentally test the effect of cognitive load on auditory susceptibility during automated driving.

Background: Being susceptible to auditory information is important for safe operation of automated vehicles, as auditory alerts are frequently used as a way to get the human involved in cases where the automated vehicle is unable to perform (part of) the driving task. Previous work using EEG Event Related Potential techniques (ERP) found that such susceptibility is slightly reduced during automated driving, but not as much as during manual driving. However, in practice, human drivers also perform non-driving tasks during automated driving, which can distract them and negatively impact their attention and their susceptibility to auditory information. It is therefore important to study the effect of mental distraction during automated driving on auditory susceptibility.

Method: We tested 24 participants who were driven in a simulated automated car on a simulated highway. Concurrently, they were performing a cognitive task that required to either *repeat* a noun or *generate* a verb that expresses the use of the noun stimulus. On each repeat and generate trial, a probe stimulus was presented after the noun to elicit the frontal P3 (fP3), a peak around 300 ms after stimulus onset using an EEG ERP oddball task. The fP3 amplitude can be used as an indication for the level of auditory susceptibility.

Results: We found that fP3 was significantly lower during automated driving with cognitive load compared to automated driving without cognitive load. fP3 was reduced for both conditions that induced load: the repeat and the generate condition.

Conclusion: Engaging in other tasks (i.e., distracting yourself) during automated driving decreases auditory susceptibility of the brain.

Application: For automated driving and other human-machine interaction settings, system designers should take into account potential reduced auditory susceptibility of the distracted human. This contrasts with current semi-automated vehicles, which require humans to respond immediately to alerts, and which therefore assume full attention.

4.1 Introduction

The amount of automation in everyday life is rapidly increasing. Although automation can take away or reduce some tasks from the human, there are many forms of automation that involve both the human and the system (e.g., Dekker & Woods, 2002; Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2000; Sheridan & Verplank, 1978). Such shared control systems require the human operator to be informed of the system state. Although in the past these tasks were typically left to skilled, well-trained, professional users such as airplane pilots and control room monitors, today more and more automation finds its way to consumer products which are operated by non-professional users. These users lack such extensive training (Janssen, Donker, Brumby, & Kun, 2019). Therefore, intuitive design of these systems becomes perhaps even more important.

Take for example the domain of automated driving, which is one of the fields that has seen an increasing amount of automation. For automated driving, six levels have been identified that differ in tasks that are performed by the driver (human) and tasks that are performed by the car (machine). Higher levels of automation require less frequent human input but might still rely on human monitoring (SAE j3016). Because of the omnidirectional nature of audio, auditory signals are widely used to request human input.

Given the reliance of current semi-automated vehicles on auditory alerts, it is important to understand how well the human brain processes such alerts. Previous research has shown that human drivers are less susceptible to auditory information while involved in visual-manual tasks such as driving and manual tracking, compared to being stationary and not performing any task (Scheer, Bülthoff, & Chuang, 2016, 2018, Wester, Böcker, Volkerts, Verster, & Kenemans, 2008). In addition, susceptibility is also found to decrease during automated driving compared to being stationary, with even further decrease in susceptibility during manual driving (Van der Heiden et al., 2018).

This previous research has typically measured auditory susceptibility when performing one single task only. In practice, however, humans will change their behavior when interacting with automation (Bainbridge, 1983). For (semi-) automated driving specifically, drivers tend to engage more frequently in additional non-driving related tasks, such as smart-phone use or having a conversation with a passenger (De Winter, Happee, Martens, & Stanton, 2014). This can create distraction from driving. However, it is not known how such additional distraction impacts auditory susceptibility under automated driving conditions. We therefore test the effect of distraction on susceptibility to auditory stimuli.

To assess auditory susceptibility, we use the novelty oddball paradigm, consisting of a stream of at least identical standard tones, mixed with (semi-) unique novels. A concurrent brain activity recording (EEG/ ERP) can then be used to quantify the novel-probe-elicited cortical activation (corrected for the standard-elicited activation). The most prominent feature of this novelty-oddball response is the so-called frontal P3 (fP3) response in the ERP: a positive peak over frontal regions (e.g., electrode FCz) around 300 ms after stimulus onset (Allison & Polich, 2008; Squires, Squires, & Hillyard, 1975; Ullsperger, Freude, & Erdmann, 2001). This peak in activation is interpreted to reflect the process of orienting to novel stimuli, especially in the sense of behavioral interrupt (Friedman, Cycowicz, & Gaeta, 2001; Polich, 2007; Kenemans, 2015; Wessel & Aron, 2013). This fP3 has been widely used to index susceptibility in a variety of conditions and tasks, including driving (Van der Heiden et al., 2018; Wester, Böcker, Volkerts, Verster, & Kenemans, 2008), mental fatigue during driving (Massar et al., 2010) manual tracking (Scheer, Bülthoff, & Chuang, 2016, 2018), games (e.g., Allison & Polich, 2008; Miller, Rietschel, McDonald, & Hatfield, 2011), arithmetic (e.g., Ullsperger, Freude, & Erdmann, 2001), and during cognitive tasks without visual or manual components (Van der Heiden, Janssen, Donker, & Kenemans, Submitted).

While previous work found a reduction in fP3 response under driving and automated driving conditions (Van der Heiden et al., 2018; Wester et al., 2008), it has not been explored how additional distraction during automated driving (e.g., a telephone call) affects auditory susceptibility. In-car distraction can take many forms and is expected to increase with higher levels of automation (De Winter et al., 2014). To be able to measure the effects of distraction during (automated) driving on auditory susceptibility we therefore need to induce mental distraction in a systematic way. To this end, we use the verb task (Abdullaev & Posner, 1998; Petersen, Fox, Posner, Mintun, & Raichle, 1989; Snyder, Abdullaev, Posner, & Raichle, 1995). In this task, participants hear nouns, and either need to *repeat* the noun (e.g., apple - apple), or *generate* a verb that is related to the noun (e.g., apple - eat). The generate task is known to induce cognitive load (Abdullaev & Posner, 1998; Snyder et al., 1995), which can interfere with dual-task performance (cf. Iqbal, Ju, & Horvitz, 2010; Kunar, Carter, Cohen, & Horowitz, 2008; Strayer & Johnston, 2001; Van der Heiden et al., 2018), and increase activity in the frontal cortex when compared with the easier repeat task (Abdullaev & Posner, 1998; Bijl et al., 2007). Furthermore, in a very recent study, the generate task impacted auditory susceptibility in non-driving conditions (Van der Heiden et al., submitted). This makes it a perfect candidate to assess susceptibility during distraction, which was the aim of the current work. We included both *generate* and a *repeat* condition to obtain better insight in the mechanism by which the distracting task reduces susceptibility: Is it the mere production of a vocal response (common to generate and repeat), or more specifically active search within the semantic network (only in generate)?

To summarize, in the current study, we test how induced cognitive load influences the susceptibility to auditory stimuli while people are driven by an automated vehicle. As previous work has shown that susceptibility reduces during manual tasks (e.g., Allison & Polich, 2008; Miller, Rietschel, McDonald, & Hatfield, 2011), during automated driving (Van der Heiden et al., 2018), and during cognitive tasks without visual or manual components (Van der Heiden et al., submitted), we hypothesize the following effects:

1. Automated driving will reduce the fP3 response compared to stationary driving (cf. Van der Heiden et al., 2018)
2. The fP3 response is further reduced when automated driving is combined with either the *repeat* or the *generate* condition of the verb task, as this task is known to induce cognitive load (Abdullaev & Posner, 1998; Snyder et al., 1995). Alternatively, the generate condition may reduce the fP3 even more than the repeat condition (active semantic search scenario).

4.2 Method

4.2.1 Participants

In order to determine the number of participants to be tested in this study, we conducted a power test using G*Power 3.1.9.4, for a one-sided t-test with difference between means (as our ANOVA is followed-up by pairwise comparisons). We used the effect size from previous work that also measured the fP3 in a driving simulator (Van der Heiden et al., 2018), where an effect size (d) of 0.71 was found for the difference between stationary and automated. As alpha-level we used .0125 (the level used in pairwise comparisons). For power we used 0.8. This returned a required sample of 22 participants. To counterbalance our design, we decided to invite 24 participants.

24 participants (22 F; 3 M) were recruited through various channels (e.g. social media, posters, flyers) using opportunity sampling. Participants were 23 years old on average (ages 18 to 25, SD = 7.2 years of age). All but one participant were native speakers of Dutch (the non-native speaker indicated speaking Dutch since primary school) and all but one had possession of a driver's license. All participants indicated to have normal or corrected to normal vision. One participant indicated to have hearing difficulties, but performance on the intelligibility test (see procedure) was on par with the other participants.

The ethics committee of the Faculty of Social and Behavioral Sciences of Utrecht University gave approval for this study (FETC16-042). All participants gave written informed consent prior to the experiment. Participants were compensated with either €12 or were provided with course credits for their time.

4.2.2 Materials

Driving simulator

A customized medium fidelity fixed base driving simulator, based on an original Green Dino three screen setup, was used. The setup is shown in Figure 4.1, and consisted of a steering wheel, pedals, shift stick and an adjustable seat combined with 40 inch screens and surround sound. OpenDS 4.5 (www.opends.eu) was used as driving simulator software. The driving environment was custom programmed and consisted of a three-lane highway that followed the trajectory of two semi circles, with a radius of 1135.9 m (one clockwise, one counterclockwise). The car drove in the middle lane of the highway at a speed of 80 km/h. There were no other cars in the driver's lane, but cars occasionally drove in the other lanes (left 87 km/h and right 73 km/h) to enhance the impression of driving on a highway. The scenario for the driving environment is shared in the supplementary material of this article.

We used a simulated driving environment, as the systems that we study currently do not exist on the road. But, even if they would, then the driving simulator setup allows higher precision with replicability and a higher level of control over (driving related) workload compared to an on-road study. Moreover, meta-reviews suggest that results that are found in a driving simulator map well to comparable on-road study results (e.g., Caird, Willness, Steel, & Scialfa, 2008; Horrey & Wickens, 2006). In addition, results with ERP studies in simulated manual driving have been replicated in on the road driving (Wester, 2009).

Figure 4.1. Driving simulator setup with participant wearing 64 electrode EEG cap.

Presentation of auditory stimuli

Two types of auditory stimuli were used in this experiment: oddball probe stimuli and verb task stimuli. All stimuli were presented using Presentation (Neurobehavioral Systems) at 75 dB trough Earlink earphones.

Oddball probe

We used a two-stimulus novelty oddball probe (Polich, 2007). In 75% of cases, the stimuli consisted of a standard sound: a 1000 Hz pure tone of 400 ms. In 25% of cases, the stimuli consisted of novel sounds: environmental sounds such as dog barking or human sneezing, that were taken from a database by Fabiani and Friedman (1995). The database consisted of 100 unique sounds that were between 159 ms and 399 ms in duration.

Verb generation and noun repetition task stimuli

A set of 96 spoken nouns was used in the verb generation and noun repetition task. In the verb generation task (Abdullaev and Posner, 1998), participants were instructed to generate a verb that fitted with the noun they heard. For example: hammer → pound, apple → eat. In the noun repetition task, participants were instructed to repeat the exact noun they heard (i.e., hammer → hammer, apple → apple).

Since our participants were Dutch, we used a Dutch translation of spoken nouns based on an English set used by Abdullaev and Posner (1998). This lead to the development of a list of 144 nouns that was used in our previous work (Van der Heiden et al., submitted). For the current study, we only used 96 nouns, as each block had 32 words (see design), so the total number of words had to be a multiple. Out of the 144 words, we selected the 96 words which in previous work (Van der Heiden et al., submitted) had the fewest errors on trials where participants had to repeat the words.

The procedure for coming up with the original Dutch set was as follows. We first translated all nouns from Abdullaev and Posner (1998) from English to Dutch. We then removed all nouns that the experimenters assessed as being ambiguous in Dutch. As we intended to present the spoken nouns during an exact interval of 400 ms we removed nouns that took long to pronounce, as identified by words that had more than two syllables. We then converted the nouns to wave sound files using text-to-speech website www.texttospeech.io with default settings of the text-to-speech algorithm (Dutch female, volume 1, rate 1, pitch 1). We threw away nouns of which presentation took longer than 500 ms as they were too long to be adaptable to the 400 ms interval limit. We then changed the tempo of the remaining files, such that each noun had a playback time of 400 ms.

4.2.3 Design

In order to assess the effect of cognitive load (due to the generate or the repeat task) that is added on top of an automated driving condition we used a single factor (*task*) within-subjects design with 4 levels: Stationary, Automated, Automated + repeat, and Automated + generate. This allowed us to compare the effect of automated driving with additional cognitive load to automated driving without additional distraction, and a baseline stationary control.

Testing blocks

In total, there were 12 experimental blocks. For each participant, the first set of four blocks contained all four unique conditions (Stationary, Automated, Automated + repeat, and Automated + generate). The order of these first four blocks was counterbalanced across participants, ensuring that each unique order was performed by one participant. For the remaining sets of blocks (2), orders were shuffled, and each participant experienced an order that they did not experience before.

Within each experimental block, 64 oddball probes were presented. In the verb generation and noun repetition blocks, 16 nouns were played in conjunction with a standard oddball probe, 16 with a novel oddball probe (i.e., 32 nouns total per block); for these blocks, 48 standards were played without a prior noun presentation. If a probe followed a noun presentation, the next probe was presented 4400 ms after the onset of the preceding oddball stimulus in order to prevent interference from speech production on the EEG measurement. On all other trials, the interval between the onset of two probe stimuli was 2000 ms (cf. Van der Heiden et al., 2018; Wester et al., 2008). In each noun trial, the 400 ms duration noun stimulus was immediately (0 s interval) by the probe stimulus (novel or standard).

For the word task, 96 different nouns were used. In total there were six sets of 32 nouns, three for the generate that combined contained all 96 words, and three for repeat that contained the same 96 words but shuffled between sets. The order of words was randomized for each participant within a set. In the end, in effect each word was used twice per participant: once in the generate condition, and once in the repeat condition.

4.2.4 Procedure

After receiving an overview of the experiment, participants were asked to read and sign the consent form. Next, for the intelligibility test we asked participants to get seated in the driving simulator. After connecting the Earlink Earphones, all nouns were played to the participant who was tasked to repeat each noun after playback. In order to validate that all nouns were intelligible, the experimenter in the meantime made notes of nouns that were incorrectly replied to.

The experimenter then applied the EEG electrodes. Participants were asked to take a seat in a standard office seat to accommodate easy access to all electrode locations. Two electrodes were placed on mastoids, for later re-referencing to average mastoids. Four ocular electrodes were applied to enable offline ocular-artifact control with horizontal and vertical electrooculography (HEOG and VEOG). After measuring the head circumference, a matching EEG cap was applied and straightened. Conductive gel was applied to all electrode locations before the corresponding electrodes were plugged in.

Participants were then given task specific instructions and asked to sit in the driving simulator. After adjusting seat position, a practice block was started where participants performed the verb generation task for 1 minute, while they were also driven by the automated vehicle and while the oddball probes were used. The participant then performed the 12 experimental blocks, with a few minutes rest after every four blocks. After the experiment, participants were asked to fill out a questionnaire on demographics and general feedback. The total experiment lasted just under two hours.

4.2.5 Signal recording

EEG setup

EEG was recorded using BioSemi ActiveTwo system with 64 active Ag-AgCl electrodes positioned following the international 10/10 system (Sharabrough, 1991), and the standard BioSemi CMS/DRL on-line reference, at a sample rate of 2048 Hz. Signal analysis was done in BrainVision Analyzer 2.1 (Brain Products GmbH, München, Germany), following similar procedures as in earlier work (Van der Heiden et al., 2018; Van der Heiden et al., submitted; Wester et al., 2008). We first downsampled the data to 256 Hz (after anti-alias filter). Data were then re referenced to average mastoids signal. A high-pass filter of 0.16 Hz, a low-pass filter of 30 Hz, and a notch filter of 50 Hz were applied to remove signal bands that were not of our interest. We then created segments for each of the four conditions for both standard and novel probes starting 1000 ms before the onset of the oddball probe, and ending 1500 ms after onset. Before creating the ERPs, we applied the Gratton & Coles ocular correction to compensate for eye movement during the recorded segments (Gratton, Coles, & Donchin, 1983). Artifacts in individual channels were rejected by the following criteria in an epoch: maximum voltage step > 120 µv/ms within 200 ms before or after events; maximum difference > 100 µv within 200 ms; minimum activity < 0.5 µv within 100 ms. Finally, grand averages were created for each of the conditions and the difference wave was obtained by subtracting the standard ERP from the novel ERP.

To obtain the correct fP3 peak at electrode location FCz, we used a collapsed localizer. The interval 285-335 ms after stimulus onset was found to best represent the fP3 peak area when the ERPs for all four conditions were collapsed. We took the average value in the fP3 interval for statistical peak analysis.

Speech response time

In order to check our cognitive-load inducing task manipulation, we measured response time. We used a microphone, connected to the auxiliary input of the BioSemi. This microphone was placed at the bottom of the left screen of the simulator, at close proximity to the participants' head location. We used an average level (i.e., calculated by a moving average) of 1000 μv over 15 samples as threshold for speech production. As speech response time we took the interval starting from the marker which was set at noun offset (which is equal to oddball probe onset) and ending at the start of speech production. We excluded the first four participants from this analysis as no microphone was present during that time.

Statistical analysis

For statistical analysis, we use R statistics (R Core Team, 2014), with an alpha level of .05. For pairwise comparisons, we use four paired t-tests for the fP3. To control for the family-wise error, our criterion for calling a difference significant was alpha / 4 (i.e., .05 / 4 = .0125).

4.3 Results

4.3.1 frontal P3

For each of the four conditions, we calculated the difference wave of fP3 ERP at electrode FCz (i.e., difference between response to the novel probe and standard probe). Figure 4.2 shows the fP3 peak, the area of the statistical analysis is indicated with dashed lines. The mean value of this dashed area is provided in Figure 4.3 per condition. There was a main effect of condition on the mean fP3 peak activation, $F(3,69) = 16.1$, $p < .001, \eta_p^2 = 0.58$. Subsequently, we performed four pairwise comparisons to test which conditions differed from each other. Tests were done in order for the conditions where we predicted the highest fP3 value (stationary) to where we expected the smallest fP3 value (automated with generate). fP3 was highest during single task stationary (M = 11.5 μv, SD = 6.1 μv). Pairwise comparisons between condition revealed that stationary did not differ significantly from single task automated (M = 9.9 μv, SD = 4.4 μv, p = .049). fP3 in the Automated + repeat condition (M = 5.1, SD = 4.9) was significantly lower than Automated (p < .001). Automated + repeat did not differ significantly from Automated + generate (M = 5.6, SD = 2.9, p = .57). Automated + generate did also differ significantly from Automated (p < .001). That is, our results suggest that performing a concurrent task under automated driving conditions reduces fP3 response and associated auditory susceptibility. Figure 4.4 shows for various time intervals how electrical activity is distributed across the scalp as a difference between the response to the novel compared to the standard. The figure illustrates that the fP3 response is indeed the highest in the frontal area of the brain, near electrode FCz that we analyzed. Moreover, it shows again how this activity is reduced under distracted conditions.

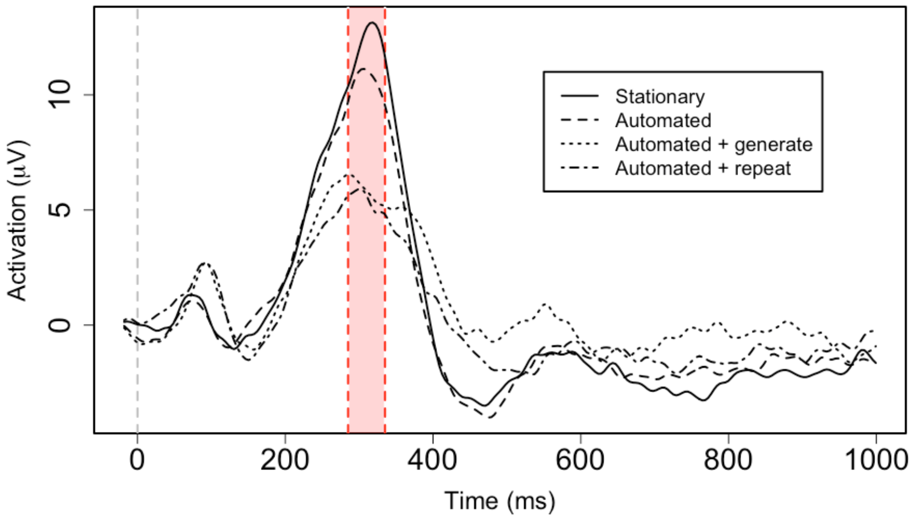


Figure 4.2. Event related potential of the four conditions (Stationary, Automated, Automated + generate, Automated + repeat). Vertical lines show onset of oddball stimulus (time point 0 ms), noun stimulus (onset at 400 ms in gray), and fP3 peak area used for statistical analysis (285-335 ms).

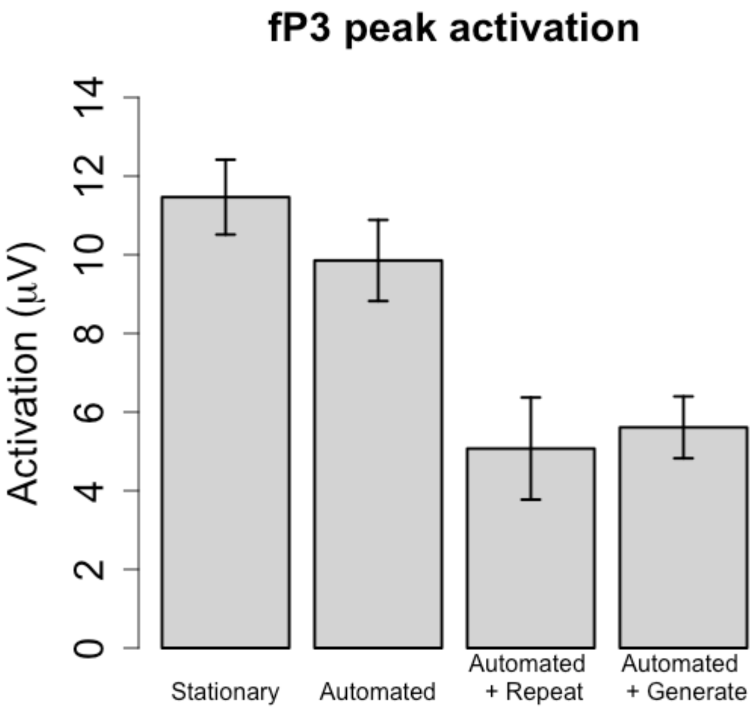


Figure 4.3. Barplot showing the average activation in the fP3 peak area (285 - 335 ms after stimulus onset). Error bars show standard error of the mean.

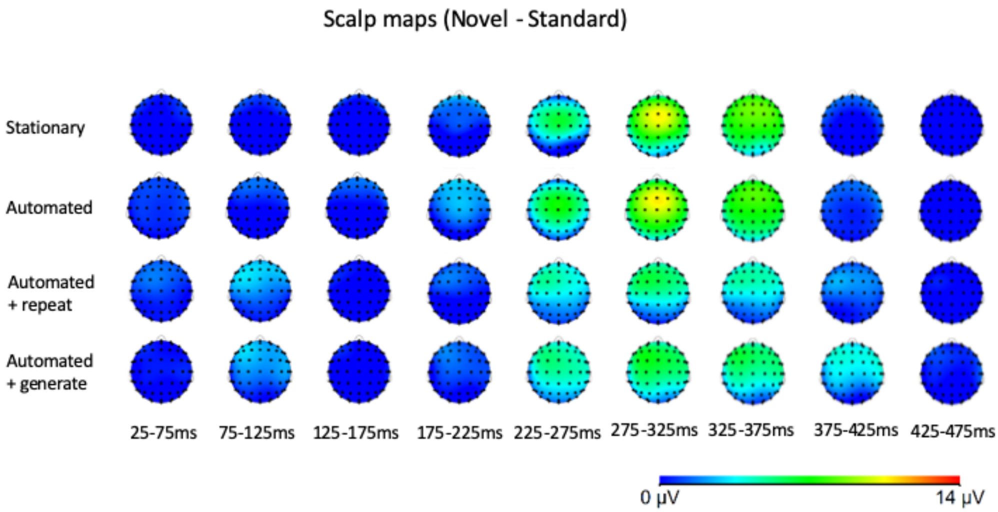


Figure 4.4. Scalp maps for various 50 ms time intervals from 25 ms after oddball probe onset to 475 ms after oddball probe onset. Average mastoid is used as reference value.

4.3.2 Speech response time

Figure 4.5 shows the average speech activation level for the different conditions over time, as measured from the point of noun offset and oddball probe onset (i.e., at time point 0 ms). As a control, the green line shows speech activation when no noun was presented. This control line shows that there is no consistent noise in the background. By comparison, during speech there is a clear response. We therefore dropped all no-word conditions for statistical analysis.

A 2 (Oddball probe: standard or novel) x 2 (Mental distraction task: repeat or generate) ANOVA showed that there was no main effect of oddball probe $F(1,19) = 3.24$, $p = .09$, $\eta_p^2 = 0.15$. There was a main effect of mental distraction task $F(1,19) = 174.1$, $p < .001$, $\eta_p^2 = 0.90$. Speech response time was higher under the generate condition (*Mdn* = 680 ms) compared to the verb generation time (*Mdn* = 287 ms). There was no significant interaction effect, $F(1,19) = 0.13$, $p = .72$, $\eta_p^2 = 0.007$

As speech response time does not significantly differ between Novel and Standard (i.e., also if we look at Figure 4.5 they show a similar pattern) potential concurrent muscle related activity from the verbal response is cancelled out in the difference wave that we use for all our fP3 analyses.

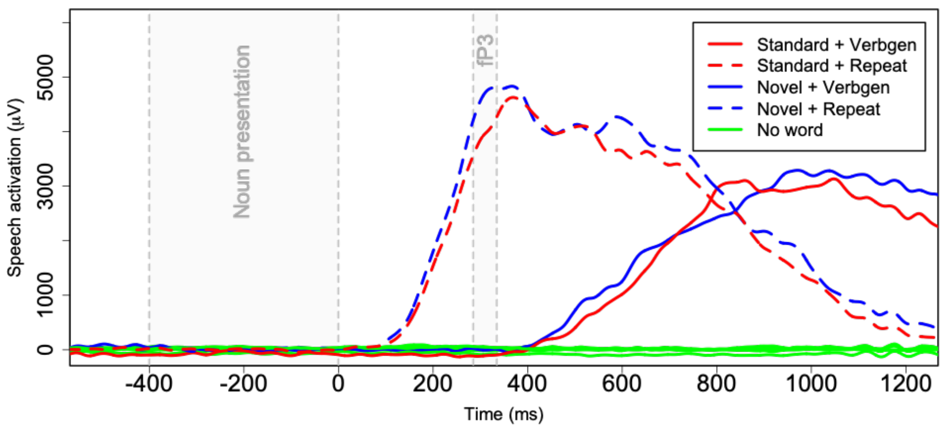


Figure 4.5. Average speech activation level for different conditions, no speech activation is expected in the no word condition. Dashed lines show activation for the repeat condition, solid lines show activation for verb generation condition. Red lines show task combined with a standard tone, blue lines show task combined with a novel sound. Note that time point 0 corresponds to noun offset and probe onset.

4.4 General Discussion

This study found that the fP3 peak is reduced when drivers are performing a distracting task under automated driving conditions. We did not statistically replicate the finding that automated driving by itself (i.e. without the addition of a secondary task) causes lower auditory susceptibility, as indicated by a decrease in the fP3 peak, compared to being stationary (Van der Heiden et al., 2018), although the pattern was in the expected direction. It is conceivable that this difference was less clear in the current study because the context of the verb-generation task induces a general relevance of all auditory stimulation. In a similar vein, the reduction of fP3 when driving compared to when stationary has been reported to disappear when the sequence of probes contains additional stimuli that have to be responded to behaviorally (Wester et al., 2008; Van der Heiden et al., 2018). In the present study we did discover that performing an additional mental task during automated driving reduces susceptibility. This is a relevant finding, given people's tendency to perform other non-driving tasks in semi-automated driving settings (De Winter et al., 2014), and the reliance of current automated vehicles on auditory alerts to request human assistance.

Previous research on the verb task had suggested that the generate condition should lead to more cognitive load compared to the repeat condition (cf. Iqbal, Ju, & Horvitz, 2010; Kunar et al., 2008; Strayer & Johnston, 2001; Van der Heiden et al., 2018). We therefore expected that possibly fP3 response would be lower in the generate condition compared to the repeat condition. In contrast to our expectations and previous research, our study did not find a difference between the generate and repeat conditions on fP3 peak. This is unlikely to reflect cognitive load induced by response production; whereas this could hold for repeat, overt responses and therefore preparatory response production processes were much later in generate, and very probably too late to affect the production of the fP3. Rather, the lack of differential fP3 could reflect equal cognitive load in repeat and generate, but induced by response production in the former and by semantic search (preceding response production) in the latter.

An alternative view is inspired by our analysis of speech data, which revealed a median voice-onset latency of 287 ms during repeat, relative to probe onset (see Figure 4.5). This indicates that a considerable amount of voice response was delivered while information might still being sampled from the probe stimulus, or immediately after that. This may have induced a form of (backward) masking that reduced the difference between novel- and standard fP3, perhaps to an extent comparable to that in the generate condition (in which median voice-onset latencies were much later, i.e., 680 ms). Further work is needed to see if, and how strongly, the repeat and generate conditions can be differentiated. Or, more generally, how different levels of cognitive load affect fP3 response and associated susceptibility under automated driving conditions.

If we compare the magnitude of the fP3 response to that observed in previous work that used the same paradigm (Van der Heiden et al., 2018), the fP3 peak that we found for distracted automated driving (i.e., automated + repeat and automated + generate in Figure 4.3) is comparable to, or even lower than, what was found for manual driving in previous work (i.e., compare with Figure 4.3 in Van der Heiden et al., 2018), which previously was observed as a condition with low susceptibility (cf. Wester et al., 2008). Having a low level of susceptibility might be problematic during manual driving as the associated brain process is interpreted to reflect the process of orienting to novel stimuli and the susceptibility to new information (Friedman, Cycowicz, & Gaeta, 2001; Polich, 2007; Kenemans, 2015). So, for example, the ability to orient (and subsequently respond) to an unexpected alert or sound in the driving environment such as a dog running after a ball. A reduced susceptibility is probably even more problematic under distracted automated driving conditions, as here the driver is already engaged to a lesser extent in the driving task. Therefore, there might be reduced situational awareness, and even stronger limitations on the ability to act appropriately to an unexpected alert.

The reduced susceptibility during automated driving (found in our previous study), combined with the reduction during an added cognitive task as observed presently is alarming, as these systems rely on the human ability to detect alerts. Although reduced susceptibility may not always lead to failed detection, in an ideal scenario (where alerts are critical), susceptibility should be high. System designers should take this reduced susceptibility into account, and develop strategies to overcome this, for example, by using multi-modal alerts or pre-alerts (Borojeni, Weber, Heuten, & Boll, 2018; Van der Heiden, Janssen, & Iqbal, 2017).

4.4.1 Limitations & future work

Although in the current study both conditions in which distraction is present (i.e., automated + generate and automated + repeat) showed a reduction in fP3 response compared to automated driving and to stationary, we did not find a difference between the two distracted conditions. This might be due to the timing of our probe; as outlined above this may have induced masking effects in the repeat condition. One way to avoid this, is to apply a delayed-response setting in which voice onsets during repeat are forced to occur much later, although admittedly this could induce undesired working memory load. Another option is to use longer intervals between noun and probe. Our previous study (Van der Heiden, et al., submitted) showed that this does not affect fP3 during generate, but this may be expected to not hold for repeat (after the voice response fP3 may well recover to a single-task level).

The point in time that we measure is limitation of our work in general. We probed susceptibility at a fixed interval: 0 ms after presentation of the noun stimulus. This interval was chosen, as previous work that involved only the generate task found that extending the interval between stimulus and probe to 200 or 400 ms (i.e. in contrast to directly after) does not influence the level of measured susceptibility (Van der Heiden, Janssen, Donker, & Kenemans, submitted).

Future work could also look into the effect over longer time spans, such as 1 second after stimulus offset. It is an open question whether susceptibility is fully restored after the oral response to the verb task (i.e., whether it is a phasic response process), or whether some level of distraction remains (i.e., a tonic process).

A limitation of our set-up, in which the generate and repeat task trials are always succeeded by an oddball probe, is that the noun might function as a cue for an oddball probe, and thereby affect fP3 response. This way, the oddball stimulus is more predictable. Moreover, at that time, listening to an auditory sound is behaviorally relevant (because a response to the noun is needed). Previous work suggests that actively engaging in an auditory task at random times (i.e., occasionally pressing a button in response to a specific tone) can increase auditory susceptibility in general (Van der Heiden et al., 2018). Therefore, if anything, having a predictable probe might have resulted in relatively higher fP3 activation. If the effect of the cue would be controlled, then even lower levels of fP3 activation might be found in the repeat and generate conditions.

4.4.2 Implications for practice

Our results show that mental distraction can reduce susceptibility to alerts. Therefore, it is important for safety-critical systems to take into account the possibility of delayed or absent response from the human operator due to such reduced susceptibility. In the case of automated driving, safety critical alerts such as handover of control requests might therefore build in resilient mechanisms, such as multi-modal alerts, or using earlier “pre-alerts” to forewarn a driver about an upcoming transition of control (Borojeni et al., 2018; Van der Heiden, Janssen, & Iqbal, 2017).

References

Abdullaev, Y. G., & Posner, M. I. (1998). Event-related brain potential imaging of semantic encoding during processing single words. *Neuroimage*, 7(1), 1-13.

Allison, B. Z., & Polich, J. (2008). Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biological psychology*, 77(3), 277-283. https://doi.org/10.1016/j.biopsycho.2007.10.014

Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man–machine systems* (pp. 129-135). Pergamon.

Bijl, S., De Bruin, E. A., Böcker, K. E., Kenemans, J. L., & Verbaten, M. N. (2007). Effects of chronic drinking on verb generation: an event related potential study. *Human Psychopharmacology: Clinical and Experimental*, 22(3), 157-166. https://doi.org/10.1002/hup.835

Borojeni, S. S., Weber, L., Heuten, W., & Boll, S. (2018, September). From reading to driving: priming mobile users for take-over situations in highly automated driving. *In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (p. 14). ACM.

Burns, P. C., Parkes, A., Burton, S., Smith, R. K., & Burch, D. (2002). *How Dangerous is Driving with a Mobile Phone?: Benchmarking the Impairment to Alcohol* (Vol. 547). TRL.

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, 40(4), 1282-1293.

De Winter, J. C., 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.

Dekker, S. W., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human–automation co-ordination. *Cognition, Technology & Work*, 4(4), 240-244.

Fabiani, M., & Friedman, D. (1995). Changes in brain activity patterns in aging: the novelty oddball. *Psychophysiology*, 32(6), 579-594.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, 25(4), 355-373. https://doi.org/10.1016/S0149-7634(01)00019-7

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology*, 55(4), 468-484.

Hancock, P. A., Simmons, L., Hashemi, L., Howarth, H., & Ranney, T. (1999). The effects of in-vehicle distraction on driver response during a crucial driving maneuver. *Transportation Human Factors*, 1(4), 295-309.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human factors*, 48(1), 196-205.

Iqbal, S. T., Ju, Y. C., & Horvitz, E. (2010, April). Cars, calls, and cognition: Investigating driving and divided attention. *In Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1281-1290). ACM.

Janssen, C. P., Donker, S. F., Brumby, D. P., & Kun, A. L. (2019). History and future of human-automation interaction. *International Journal of Human-Computer Studies*.

Janssen, C. P., Van der Heiden, R.M., Donker, S. F., & Kenemans, J. L. (2019). Measuring susceptibility to alerts while encountering mental workload. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (pp. 415-420). New York, NY: ACM.

Kenemans, J. L. (2015). Specific proactive and generic reactive inhibition. *Neuroscience & Biobehavioral Reviews*, 56, 115-126. https://doi.org/10.1016/0013-4694(75)90263-1

Kunar, M. A., Carter, R., Cohen, M., & Horowitz, T. S. (2008). Telephone conversation impairs sustained visual attention via a central bottleneck. *Psychonomic bulletin & review*, 15(6), 1135-1140.

Massar, S. A., Wester, A. E., Volkerts, E. R., & Kenemans, J. L. (2010). Manipulation specific effects of mental fatigue: evidence from novelty processing and simulated driving. *Psychophysiology*, 47(6), 1119-1126.

Miller, M. W., Rietschel, J. C., McDonald, C. G., & Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. *International Journal of Psychophysiology*, 80(1), 75-78.

Parasuraman, R., & Riley, V. (1997). Humans and automation: *Use, misuse, disuse, abuse*. *Human factors,* 39(2), 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans,* 30(3), 286-297.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of cognitive neuroscience,* 1(2), 153-170.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148. https://doi.org/10.1016/j.clinph.2007.04.019

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

SAE International. (2018). j3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. Retrieved from https://www.sae.org/standards/content/j3016_201806/

Scheer, M., Bülthoff, H. H., & Chuang, L. L. (2016). Steering demands diminish the early-P3, late-P3 and RON components of the event-related potential of task-irrelevant environmental sounds. *Frontiers in human neuroscience*, 10, 73. https://doi.org/10.3389/fnhum.2016.00073

Scheer, M., Bülthoff, H. H., & Chuang, L. L. (2018). Auditory Task Irrelevance: A Basis for Inattentional Deafness. *Human factors*, 60(3), 428-440. https://doi.org/10.1177%2F0018720818760919

Sharbrough, F. (1991). American Electroencephalographic Society guidelines for standard electrode position nomenclature. *J clin Neurophysiol*, *8*, 200-202.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

Snyder, A. Z., Abdullaev, Y. G., Posner, M. I., & Raichle, M. E. (1995). Scalp electrical potentials reflect regional cerebral blood flow responses during processing of written words. *Proceedings of the National Academy of Sciences*, 92(5), 1689-1693.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and clinical neurophysiology*, 38(4), 387-401.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological science*, 12(6), 462-466.

Ullsperger, P., Freude, G., & Erdmann, U. (2001). Auditory probe sensitivity to mental workload changes–an event-related potential study. *International Journal of Psychophysiology,* 40(3), 201-209. https://doi.org/10.1016/S0167-8760(00)00188-4

Van der Heiden, R. M., Iqbal, S. T., & Janssen, C. P. (2017). Priming drivers before handover in semi-autonomous cars. *In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 392-404). ACM.

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., Hardeman, L. E., Mans, K., & Kenemans, J. L. (2018). Susceptibility to audio signals during autonomous driving. *PloS one*, 13(8), e0201963. https://doi.org/10.1371/journal.pone.0201963

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., & Kenemans J. L. (submitted) The influence of cognitive load on susceptibility to audio. Manuscript submitted for publication.

Wessel, J.R. and A.R. Aron, Unexpected Events Induce Motor Slowing via a Brain Mechanism for Action-Stopping with Global Suppressive Effects. *The Journal of Neuroscience*, 2013. 33(47): p. 18481-18491.

Wester AE. Attention Deficit and Impulsivity: Driving, Drugs and Electrophysiology. 2009. Utrecht: Utrecht University. PhD thesis.

Wester, A. E., Böcker, K. B. E., Volkerts, E. R., Verster, J. C., & Kenemans, J. L. (2008). Event-related potentials and secondary task performance during simulated driving. *Accident Analysis & Prevention*, 40(1), 1-7. https://doi.org/10.1016/j.aap.2007.02.014

Part II

Testing interventions

Chapter 5

Priming Drivers before Handover in Semi-Autonomous Cars

Remo M. A. van der Heiden
Christian P. Janssen
Shamsi T. Iqbal

Published as:
Van der Heiden, R. M., Iqbal, S. T., & Janssen, C. P. (2017). Priming drivers before handover in semi-autonomous cars. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 392-404). DOI: https://doi.org/10.1145/3025453.3025507

Author contributions:
RMAV designed and programmed the experiment, collected the data, analyzed the data, and wrote the manuscript with input from the other authors. STI supervised the project, contributed to experimental design, interpretation of the data, and editing the manuscript. CPJ contributed to interpretation of the data and editing the manuscript. All authors read and approved the final version of the manuscript.

Abstract

Semi-autonomous vehicles occasionally require control to be handed over to the driver in situations where the vehicle is unable to operate safely. Currently, such handover requests require the driver to take control almost instantaneously. We investigate how auditory pre-alerts that occur well before the handover request impact the success of the handover in a dual task scenario. In a study with a driving simulator, drivers perform tasks on their phone while the car is in an autonomous mode. They receive a repeated burst audio pre-alert or an increasing pulse audio pre-alert preceding the standard warning for immediate handover. Results show that pre-alerts caused people to look more at the road before the handover occurred, and to disengage from the secondary task earlier, compared to when there was no pre-alert. This resulted in safer handover situations. Increasing pulse pre-alerts show particular promise due to their communication of urgency. Our detailed analysis informs the design and evaluation of alerts in safety-critical systems with automation.

5.1 Introduction

With the advent of novel (in-)vehicle technologies, new challenges emerge in managing driver distraction. Research has shown that drivers perform a variety of tasks that may distract from driving including interacting visually and manually with their mobile phones (Dingus et al., 2016; Klauer et al., 2014). A recent meta-review suggests that distraction is likely to increase even further in 'self-driving' or 'autonomous' vehicles, where cars assume more of the driving responsibilities (De Winter, Happee, Martens, & Stanton, 2014). In such vehicles, gradations of automation can be identified, as defined in various standards (Gasser & Westhoff, 2012; NHTSA, 2013; SAE International, 2014). At the lowest level of automation (e.g., no automation, or level 0 in (SAE International, 2014)), the human driver is in full control of the car. In full automation (e.g., level 5 in SAE International, 2014), the human driver is not involved in any driving task anymore. This is, for example, the vision of the Google autonomous car (e.g., Urmson, 2015) For the levels in between these extremes (e.g., levels 1-4 in SAE International, 2014) there is some form of shared responsibilities. For example, even if the car is driving by itself, the driver might need to intervene when the system is uncertain about what action to take. For ease of reference we will use the umbrella term "semi-autonomous vehicle" to describe this wide category. The important characteristics of these systems for our work is that the vehicle can drive relatively independently for some time, but can request the human driver to assist or take over control through an alert.

A natural question that is of relevance to the CHI community is then: what should the specifics of such an alert be? Indeed, various aspects of handover (Gold, Dambock, Lutz, & Bengler, 2013) and in-car alerts (Politis, Brewster, & Pollick, 2015; Walch, Lange, Baumann, & Weber, 2015) have received attention in the literature. Currently alerts, such as in the Tesla model S, include a brief alert and immediately handover control to the driver. However, given that the driver may not have the proper situational awareness to be able to immediately take the proper action, immediate alerts as such can potentially result in fatal outcomes. Moreover, with higher levels of automation (e.g., level 3 in SAE International, 2014) such alerts are even more crucial as the car is monitoring the environment for large time intervals and drivers might have disengaged (cf. De Winter et al., 2014).

In this research we investigate whether providing a *pre-alert*, an additional alert that commences well before the actual handover request (in our study: 20 s) can help drivers be better prepared to safely navigate the incident for which the handover of control occurs. Designing such a pre-alert for semi-autonomous vehicles presents unique challenges, as drivers distract themselves more with other tasks as automation of the car increases (De Winter et al., 2014). If they are then asked to take-over, they should not have remnants of their preceding task (e.g., checking e-mail, reading the news) that inhibit their ability to successfully take over (cf. Strayer, Cooper, Turrill, Coleman, & Hopman, 2015). At the same time, tasks that drivers engage in while in an autonomous vehicle may be more important than the typical secondary task in today's manual cars, making it more challenging to suddenly disengage (Hancock, 2013).

In a simulator study, we study the effects of early warnings, or pre-alerts, on handover performance in dual task scenarios. We investigate three research questions: (1) how do pre-alerts affect behavior before take-over including eye-gaze and suspension of the non-driving task, (2) how do drivers perceive the pre-alerts, and (3) how do the pre-alerts affect driving performance. We present two types of pre-alerts: (A) A *repeated burst audio pre-alert*, and (B) an *increasing pulse audio pre-alert*. While the car drives itself, drivers occasionally perform a video transcription or calendar entry task on a mobile phone. Results showed that pre-alerts helped drivers prepare better for taking over control by increasing gaze on the road and earlier suspension of the phone task, followed by quicker reaction to traffic incidents compared to having no pre-alerts.

In the remainder of this paper we first provide more background on multitasking, managing multitasking, and in-car alerts. We then describe our study. Finally, we discuss the results, including their implications for theory and design, limitations, and potential for future work.

5.2 Related work

5.2.1 Driving and Multitasking

Multitasking is a prevalent practice while driving (Dingus et al., 2016; Klauer et al., 2014). Multiple studies have documented the detrimental effects of cell phone conversations, texting, and interacting with in-vehicle systems while driving (Alm & Nilsson, 1995; Briem & Hedman, 1995; Brookhuis, De Vries, & De Waard, 1991; Salvucci, Markley, Zuber, & Brumby, 2007; Reed & Robbins, 2008). Switching from a non-driving task to driving is often challenging. For example, drivers who engage in phone conversations have slower braking reaction time (Alm & Nilsson, 1995; Lee, Caven, Haake, & Brown, 2001), degraded steering performance (Brookhuis et al., 1991), and a higher likelihood of accidents (Redelmeier & Tibshirani, 1997) than drivers with no distractions.

In the majority of this preceding work, the driving task has typically been considered the 'primary task' where the driver's focus is expected at all times, with other tasks as 'secondary'. However, as the automation technology matures, human drivers might be required less and less to take-over control of the car. Therefore, they might increasingly engage in other tasks (cf. meta-review in De Winter et al., 2014), and those other tasks (e.g., checking e-mails, reading the news, having a conversation, watching a video) might even feel as a primary task with driving as a distraction (cf. Hancock, 2013).

5.2.2 Managing Multitasking during driving

A known strategy to manage multitasking is to interleave activities, which has been well documented in conceptual frameworks (e.g., Altmann & Trafton, 2002; Boehm-Davis & Remington, 2009; Borst, Taatgen, & Van Rijn, 2015; Salvucci & Taatgen, 2011). In the domain of driving, multiple studies have looked at how drivers interleave non-driving, secondary tasks with driving. A common strategy is to wait for 'natural breakpoints' in the task to switch attention (Brumby, Howes, & Salvucci, 2007; Brumby, Salvucci, & Howes, 2009; Iqbal, Ju, & Horvitz, 2010; Janssen & Brumby, 2010; Janssen, Brumby, & Garnett, 2012; Salvucci, 2005); for example, Iqbal et al. showed that drivers chunked a task of providing directions while driving into multiple steps and reoriented to driving at the boundaries between chunks (Iqbal et al., 2010). There are many advantages of interleaving at natural breakpoints: it reduces mental workload (Bailey & Iqbal, 2008; Salvucci & Bogunovich, 2010) as it reduces information that needs to be maintained in memory (Borst, Taatgen, & Van Rijn, 2010), it frees mental resources such as visual attention for other tasks (Salvucci & Taatgen, 2011; Wickens, 2008), it reduces stress (Bailey & Konstan, 2006), it reduces the time needed for later task resumption (cf. Altmann & Trafton, 2002), and it can offer speed-accuracy trade-offs in dynamic environments such as driving (Janssen et al., 2012).

In autonomous vehicles, however, while people are not driving, the non-driving task might capture most of the driver's attention (Hancock, 2013). In a hand-over scenario, people might therefore not want to immediately let go of whatever they were working on. Priming a handover in a timely manner through pre-alerts, as we propose in this paper, has the advantage that it allows drivers to disengage from their 'primary' (non-driving) task at a pace that suits them. Moreover, gradually disengaging from the task and waiting for a natural breakpoint can also benefit the driving task. If interleaving at a more opportune task reduces workload (Bailey & Iqbal, 2008; Salvucci & Bogunovich, 2010) and stress (Bailey & Konstan, 2006), then people are in a better mental state to resume driving.

5.2.3 (In-car) Alerts

The idea of using alerts to gain user attention has been explored in many domains. Mediation or alerting has been proposed as one of the four interruption management methods in McFarlane's work (McFarlane, 2002). In the driving domain, researchers have explored the effectiveness of systems for aiding driving by providing local danger alerts (Cao, Castronovo, Mahr, & Müller, 2016), mediating communications among car passengers (Mahr, Pentcheva, & Müller, 2009), or persuading people to drive in a more economical manner (Meschtscherjakov, Wilfinger, Scherndl, & Tscheligi, 2009). For example, Iqbal et al. explored how alerts can gain user attention while driving and conversing on a cellphone simultaneously, and found that alerts reduced driving errors, while also reducing conversation quality (Iqbal, Horvitz, Ju, & Mathews, 2011).

In the domain of autonomous driving, the idea of using alerts before handover of control has recently gained traction, based on the concern that the current designs of immediate take over may not yield the desired outcome. Most work has focused on the design of the alert, in terms of the timing and modalities so that it conveys the required urgency (Gold et al., 2013; Koo et al., 2014; Naujoks, Mai, & Neukum, 2014; Politis et al., 2015; Walch, Lange, Baumann, & Weber, 2015). For example, Gold et al. (2013) showed that alerts happening 7 seconds before the incident resulted in more successful take overs compared to alerts that were presented 5 seconds before. Walch et al. (2015) also investigated different timings and modalities. While they found no difference in driving performance, drivers had a preference for the alerts reinforced through both auditory and visual means. There was no significant advantage in driving performance.

Perhaps closest to our work is Politis et al.'s study which tested language based alerts for upcoming incidents (Politis, et al., 2015). Alerts were delivered via audio, visual, or tactile means. Results showed that drivers quickly transitioned to the driving task for warnings that conveyed urgency, and performance was worst for unimodal visual alerts. Other work has looked at using auditory cues to provide drivers in autonomous vehicles awareness about the environment (Beattie, Baillie, & Halvey, 2015). This work does not separate the specific scenario of handover where the driver's awareness is put to test by having to react to an incident in very short notice.

Compared to existing work, our research focuses less on the exact timing or ideal conveyance of urgency via the alerts. Rather, we draw upon designs of alerts that have been effective in conveying urgency in a timely manner (Fagerlönn, Lindberg, & Sirkka, 2012; Haas & Casali, 1993). Our goal is to understand if a pre-alert (i.e., an early alert well before the final warning) is useful in general, and if so, why and how it supports ease of disengagement from a secondary task. Such a scenario can be crucial in autonomous vehicles.

5.3 User study

In our study of the effects of pre-alerts on preparing for handover, we aim to address three research questions:

RQ1: What do drivers do before handover? We therefore analyze eye-gaze and time on secondary task.
RQ2: How do drivers experience the handover? We therefore look at subjective ratings and physiology.
RQ3: How successful is the handover? We therefore look at the first reaction time, speed reduction, and an analysis of unsafe incidents.

To answers these questions, we conducted a user study using a driving simulator with autonomous capabilities where drivers engage in a non-driving task on a mobile device and are required to take over control in certain driving situations where the car is unable to continue.

5.3.1 Users

Twenty-four drivers (12 M; 12 F) with an average age of 32.5 years (*SD* = 9.6) were selected by quota sampling. Each driver had a valid driver's license and drove on average 5.4 days a week (*SD* = 2.4). All drivers provided informed consent and were compensated with a $50 gift card.

5.3.2 Tasks

Drivers performed two task types: driving (part manual, part autonomous), and a non-driving task on a mobile phone.

5.3.3 Driving task

A simulated driving task was developed in a medium fidelity simulator. Three 47'' TVs projected the driving environment. Drivers sat in an adjustable car seat behind a full Ford dashboard. During manual driving, drivers used a steering wheel, gas and brake pedal (transmission was automatic).

Simulation software consisted of the STIsim simulator software. Data was recorded at a rate of 10 data points per second. An eye tracker was mounted on the dashboard to capture eye gazes on the driving scene. A custom scenario was developed, consisting of a drive on a two lane country road with bends and straight road segments without intersections. Oncoming traffic was presented on the opposite lane, not in the driver's lane.

The scenario consisted of manual driving and automated driving trajectories. The car started stationary and drivers had to press the gas pedal to start driving manually. Drivers had to maintain the posted speed limit and remain in lane. Occasional curves were included in each scenario. The curves were subtle enough so braking was not needed but input of drivers was required to stay in their lane safely.

After 1500 feet (25-30 s after the start), an automated voice would state "Automation enabled", and the car assumed driving control. At that point, if desired, drivers could release the steering wheel and the pedals. The car continued to drive itself until it warned drivers when they were needed to take over the driving task. While the car was in auto drive there was no driver initiated way to get back control. In the event of a handover, the car would start by warning the driver and the controls were handed over back to the driver after a voice said: "Automation disabled". Optionally, such a warning could be preceded by a pre-alert, 20 s before handover. The pre-alerts are described in more detail later.

There were four handover scenario varieties, each with two instances – resulting in eight scenarios in total. Examples are shown in Figure 5.1. The fog scenarios had light or heavy fog, and required drivers to slow-down, maintain their lane, and avoid other cars. The construction works scenario had cones along the road and required drivers to slow down and steer accurately. One version also included a lane change. The parked car scenario had a car that blocked either part of the road, or the full road, requiring the driver to slow down or stop. The dog scenario had a dog abruptly crossing the street, requiring the driver to stop in order to avoid hitting the dog.

The scenarios required different types of responses, such as braking and accurate steering. Some scenarios allowed for multiple responses such as braking and steering away from an accident. However, although the simulator allowed for this variety of maneuvers, some were not safe. This is similar to how a driver in real traffic can sometimes respond in different ways, of which only some are safe.

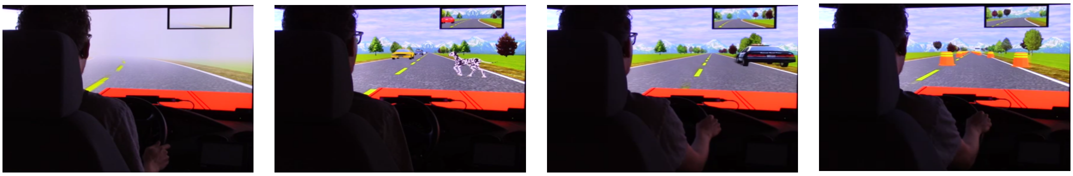


Figure 5.1. Four handover scenarios: (a) Fog, (b) Dog crossing, (c) Parked car on the side of the road, (d) Construction works.

5.3.4 Non-driving tasks

Recent reviews suggest that drivers distract themselves with non-driving tasks more when automation in the car increases, which impacts situational awareness and response times (De Winter et al., 2014). We therefore also included conditions in which drivers performed non-driving tasks. As the structure of the task can also impact when drivers look at the road (Brumby et al., 2009; Janssen & Brumby, 2010; Janssen et al., 2012), we used two tasks. Half the drivers performed a video transcription task while the other half performed a calendar task. These tasks were chosen because they represent tasks that people might prefer to do in an autonomous car – such as watching videos (Politis et al., 2015) and performing short typing tasks, as reported on a pre-survey. All secondary tasks were conducted on a Nokia Lumia 1520 phone with Windows 8.1.

Video task. A custom developed app (see Figure 5.2, left) showed a video screen and an input box. Drivers had to play the video (which showed elementary statistics lectures (Illowsky & Dean, 2008)) and had to transcribe it in the textbox. Controls to play, pause, and forward in the video were embedded in the player. We used the standard keyboard from Windows phone 8.1, with auto-correction and -completion disabled. This allowed for more reliable measurement of writing performance. We logged the timestamp of each keypress.

Calendar task. The alternative task was a calendar task (see Figure 5.2, right). Drivers were asked to enter event information in a simple calendar interface. The interface had two separate screens, one showed all the upcoming events as a digital flyer, the other showed the input boxes and saved items. We again logged the timestamp of each keypress.

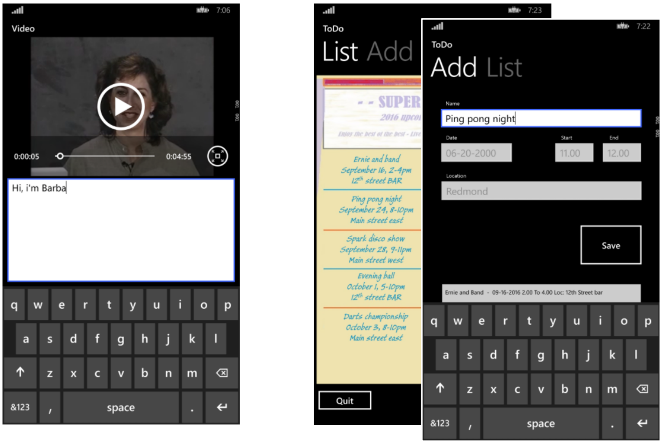


Figure 5.2. Layout of Video task (left) & Calendar task (right)

5.3.5 Warning & Pre-alerts

At each hand-over instance, in all conditions, drivers received a voice warning: "automation disabled" 1 s before the handover. Drivers were instructed to take over the driving from the system at this moment. In the *pre-alert* conditions, drivers also received either of two pre-alerts 20 s prior to this final warning. For the pre-alerts we used individual 500 Hz beeps that lasted 150 ms per beep. Beep frequency and length were recommended in the warning literature (Mondor & Finley, 2013).

Repeated burst audio pre-alert: In this condition, bursts of 3 beeps are played 3 times, with silence in between burst sets. The burst sets started playing at 20, 10, and 1 seconds before the final warning (i.e., in Figure 5.4: 0, 10, and 19 seconds relative to start of pre-alert phase). Repeated burst alerts are already used in the car context, for example, to notify when a door is not closed firmly or when a car is almost out of fuel.

Increasing pulse audio pre-alert: In this condition, beeps are given throughout the 20 second pre-alert time. However, the interval between consecutive beeps is reduced gradually over time as the driver gets closer to the critical moment. The initial inter-stimulus interval is 1000 ms. The final inter-stimulus interval is 50 ms. An increasing pulse audio alert is already used at other places in the car domain to suggest increased urgency. For example, park assistant alerts decrease the inter-stimulus interval between beeps when a car gets close to another object to suggest urgency to stop. In other studies, increasing pulses (e.g., heartbeats) have also been used to successfully convey urgency (Janssen, Iqbal, & Ju, 2014).

No pre-alert: In this baseline condition, no pre-alert is given and drivers are only warned by the final warning voice 1 s before handover of control.

Figure 5.3 shows a schematic of the different stages in a single run. There are two experimental segments in each run. Each run started with a period of driving by the driver, followed by a period of auto drive during which the driver could engage in a non-driving task (depending on condition). During the auto drive there would be optionally pre-alerts (depending on experimental condition), followed by a final voice warning declaring the handover to the driver. This would be followed by the handover event during which the driver had to start driving. After a while, the second segment started, following the same procedure. The entire run was about 6 minutes long (i.e., roughly 3 minutes per segment).

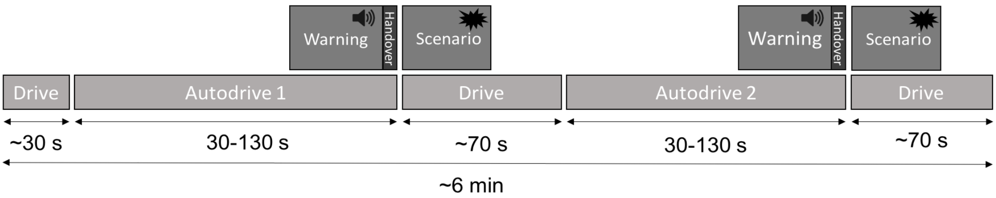


Figure 5.3. A schematic representation of a run for one condition with two consecutive segments.

5.3.6 Experimental design

We used a 3 x 2 within-subjects design. We manipulated Pre-alert type (*Repeated burst*, *Increasing pulse*, or *No pre-alert*), and number of tasks (*single-task* driving, or *dual-task* driving with secondary task). For each of the six combinations of pre-alerts and number of tasks, we developed one drive/run. Each drive consisted of two hand-over moments that had a critical incident. In sum, each participant had six drives and twelve handover situations. Ordering of the six conditions was counterbalanced following a Latin Square design, to compensate for learning effects. As we had eight critical events, we randomly assigned these to drivers with the requirement that each run had one incident that might require braking (i.e., dog or parked car) and one incident that required accurate steering (i.e., cones or fog). Finally, we also measured single-task performance as baseline (explained in procedure). For the secondary task, drivers either performed the video or calendar task (12 drivers per task, randomly assigned).

5.3.7 Procedure

On arrival, drivers were given an overview of the study and asked to sign an informed consent form as well as to fill out a questionnaire about current driving behavior. They were then asked to make themselves comfortable in the car seat of the simulator, which was adjusted so their foot could reach the pedals and their eyes were visible for the eye tracker.

We then calibrated the eye tracker and set participants up with the Microsoft band2, which was used for measuring heart rate. This was followed by a 2-minute training session on the secondary task (video or calendar, depending on group), followed by training with the driving task. In this single-task practice drive, drivers practiced all 4 types of handover events, and were introduced to the two pre-alerts. The first event was preceded by an increasing pulse pre-alert, the second with a repeated burst pre-alert, the last two had no pre-alert.

The remainder of the experiment consisted of the six experimental drives/runs. In addition, single-task performance with either the video or calendar task (depending on assignment) was performed at a random position in between the experimental drives. After all trials ended the drivers filled out a general questionnaire on their overall experience and preferences about the pre-alerts.

5.3.8 Measurements

Below we define our exact measurements, sorted by research question (RQ). Unless otherwise noted, we use a 3 (pre-alert) x 2 (number of tasks) within-subjects analysis of variance (ANOVA) with an alpha-level of .05 for significance. Where needed we use Holm-Bonferroni-corrected post-hoc tests. Error bars in all plots show standardized error of the mean.

Gaze during driving (RQ1) We used an SMI REDn eye tracker which reported tracking status (eyes tracked or not) and X,Y gaze-coordinates at 30 samples/second. The eye tracker was positioned on the dashboard just above the steering wheel. For short drivers an extra cushion was used to make sure their eyes stayed visible during the entire experiment, as tested before the experiment.

For all gaze metrics, the eye-tracker could only track the eyes when the user was looking at the simulator screen (not at the secondary task). We therefore define "looking at the road" as gaze samples where the eye tracker was at least tracking one eye, and "not looking at the road" as moments where no eyes were tracked. Given the large size of the screen and the peripheral location of the phone, this crude metric is a good approximation of actual looking at the road. Based on this information, we calculate what *percentage of drivers* look at the road (e.g., Figure 5.4) and what *percentage of the time* drivers on average look at the screen. In pilot studies we crosschecked with 2 eye trackers that the eyes were consistently detected when watching the simulator screen.

Disengaging from the secondary task (RQ1). Based on logs of touchscreen keypresses, we determined the interval between the start of the pre-alert phase and the last keypress. Shorter intervals indicate a faster disengagement.

User preferences (RQ2). In a questionnaire, drivers indicated their preferences after the experiment. The questions used a five-point scale ranging from low/ poor (1) to high/ good (5).

Heart rate (RQ2). A Microsoft Band2 measured the number of beats per minute. One value was logged per second.

Initial reaction time (RQ3). For each handover event we measured the reaction time as the time interval between the moment automation was disabled and the first action of the driver (either a brake press or steering wheel input).

Driver speed reduction (RQ3). For each handover event, we measured at what speed the driver drove, as logged by the simulator at a rate of 10 samples/second.

Unsafe incident analysis (RQ3). For the first 10 seconds after handover, we manually labeled whether observed behavior was unsafe, following conservative but realistic pre-defined rules. In all scenarios, driving more than 10 mph over the posted speed limit, or leaving the highway was labeled unsafe (some drivers drove on the grass). In addition, in one *parked car scenarios* only a full stop avoided a collision; not doing so was labeled unsafe. Also, for one *dog scenario*, some drivers crossed into the lane of incoming traffic, despite that cars might come in. This was also labeled unsafe.

5.4 Results

We measured how handover performance changes with the use of pre-alerts and the use of a secondary tasks. We discuss our results in the context of our three research questions.

5.4.1 RQ1: Gaze & phone task engagement before handover

Percentage of drivers looking at the road

Figure 5.4 shows the percentage of drivers looking at the road in the different stages of handover, relative to the start of the pre-alert (time point 0). In single-task trials (grey lines) we found that at each time point, on average 70 to 80% of the drivers looks at the road. There is a slight increase in the phase before the handover, but this is not different between the various alert conditions.

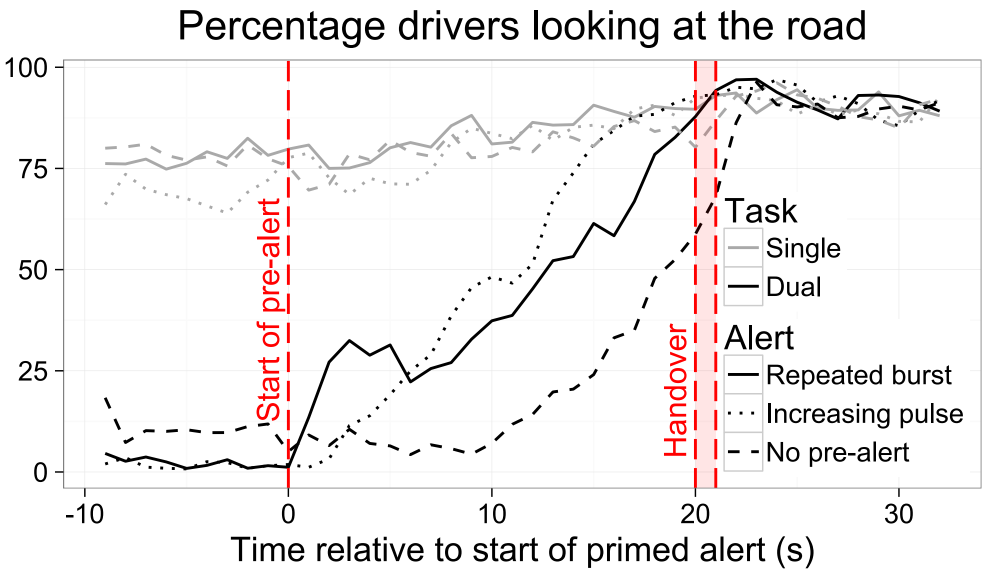


Figure 5.4. The percentage of drivers looking at the road relative to the start of the primed warning.

The pattern is different for the dual-task condition (dark lines). In the phase before the pre-alert (-9 to 0 s), drivers look more at the road when they are in the no pre-alert condition (dashed line) compared to the two pre-alert conditions. This is likely because drivers know that they have no pre-alert to rely on, and want to be prepared for a later handover request.

During the period where the pre-alert is active (0 to 20 s), in both pre-alert conditions (Repeated burst, solid line; increasing pulse, dotted line), drivers look more frequently at the road compared to the no pre-alert condition (dashed line). Moreover, this accumulates over time, as drivers get closer to the handover itself. The two pre-alert conditions are hard to distinguish from one another. Qualitatively, in the repeated burst pre-alert condition drivers start looking at the road after each burst of tones (0, 10, and 19 s). The strongest bump is after the first burst. By contrast, for the increasing pulse pre-alert there is a more gradual increase over time.

In the condition where there is no pre-alert, the percentage drivers looking at the road initially is similar to before the start of the pre-alert. However, 10 seconds before handover the percentage of drivers looking at the road increases, even though they have not yet received an alert. This is because in the simulator parts of the critical event gradually become visible even before the handover request occurs. This is similar to how in real driving visible cues are sometimes available ahead of time (e.g., a traffic jam ahead). Such visual cues make non-distracted drivers look at the scene and prepare. The take-away message from this graph though is that in both pre-alert conditions, drivers look up a lot *earlier* and do *not* rely on input from the road *at the last moment*.

Gaze during auto-drive before the pre-alert period

We quantified the preceding results. First, we looked at the entire auto-drive period before the pre-alert. For each driver, we calculated the percentage of time that they looked at the road. An ANOVA showed that there is a significant effect of number of tasks on the percentage of time drivers spend watching the road $F(1,23) = 344.3$, $p < .001$, $\eta_p^2 = 0.94$. Drivers looked at the road more than ten times as much in single task condition (M = 71%, SD = 19%) compared to dual task condition (M = 6%, SD = 13%). There was also a significant effect of pre-alert, $F(2,46) = 4.441$ $p = .002$, $\eta_p^2 = 0.16$. Post-hoc tests found that gazes at the road during No pre-alert (M = 9%, SD = 4%) was significantly higher than Repeated burst (M = 4%, SD = 3%, $p = .001$) and Increasing pulse (M = 5%, SD = 4% , $p = .016$), as in the no pre-alert condition drivers cannot rely on a signal to warn them, they interleave the tasks more often to check for hazardous situations. The two pre-alerts did not differ from each other ($p > .1$). There was no significant interaction effect ($p > .1$).

Gaze during pre-alert phase

During the pre-alert phase (i.e., 0-20 s in Figure 5.4), there was again a significant main effect of number of tasks on percentage of time drivers spend watching the road $F(1,23) = 168.7$, $p < .001$, $\eta_p^2 = 0.88$. In general, drivers were looking at the road about twice as much in single task condition (M = 83%, SD = 18%) compared to dual task condition (M = 42%, SD = 29%). There was also a significant main effect of pre-alert, $F(2,46) = 16.28$, $p < .001$, $\eta_p^2 = 0.41$. However, both main effects were affected by a significant interaction effect, $F(2,46) = 15.95$, $p < .001$, $\eta_p^2 = 0.41$. Post-hoc tests revealed that in single-task there was no significant difference between the three pre-alert conditions (all $ps > .1$). However, in dual-task, the percentage gaze at the road was significantly lower in the no pre-alert condition (M = 23%, SD = 15%) compared to repeated burst (M = 49%, SD = 19%) and increasing pulse

(M = 54%, SD = 16%) all ps < .001. This is expected, as the pre-alerts warn drivers to look at the road and drivers therefore indeed gaze more at the road.

Disengaging from the secondary task

To test whether the alert helped drivers disengage from the secondary phone task, we tested how long they continued after the alert had started using a 3 (Pre-alert type) x 2 (Secondary task type) ANOVA. There was a significant effect of pre-alert on the time drivers continue their phone task after alert onset, $F(2,44) = 30.09$, $p < .001$, $\eta_p^2 = 0.56$. Figure 5.5 shows the data. Post-hoc tests showed that all three conditions differed significantly from each other (with the difference between increasing pulse and repeated burst with $p = .038$, all other $ps < .001$). As the figure shows, in the increasing pulse condition drivers quit the secondary task twice as fast compared to the no pre-alert condition. The ANOVA also revealed a marginal effect of secondary task, $F(1,22) = 3.81$, $p = .06$, $\eta_p^2 = 0.15$. Disengagement was slightly faster in the calendar task (M = 10.5s, SD = 4.7s) than the video task (M = 12.7s, SD = 6.0s). There was no interaction effect, $F(2,44) = 1.5$, $p > .1$.

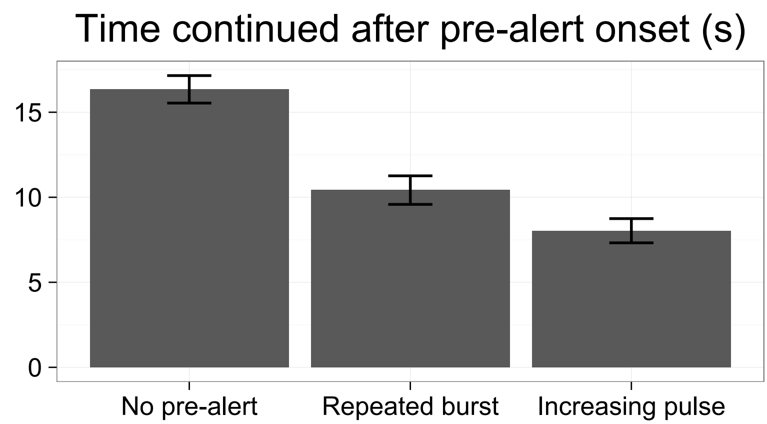


Figure 5.5. The average time secondary task continued after different alert onsets.

5.4.2 RQ2: User experience

User preferences

Subjective feedback revealed overall preference for pre-alerts was divided. Twelve drivers preferred the increasing pulse pre-alert. Feedback included that these drivers liked that they could finish their task and prepare for handover. Twelve other drivers preferred the repeated burst pre-alert, as it felt less disruptive than the increasing pulse.

In the post questionnaire drivers provided various scores for the different pre-alerts on a five-point scale with anchors for low (1) and high (5). In Figure 5.6 we present the histograms of the score for the metrics (1) annoyance, and (2) disruptiveness of the pre-alert. The responses are again divided, as reflected in the broad distributions. Some trends are that the increasing pulse is reported more frequently as conveying high to too much urgency, and being more frequently considered as highly annoying and disruptive. However, there are also drivers who reported an inverse pattern (e.g., rated annoyance low).

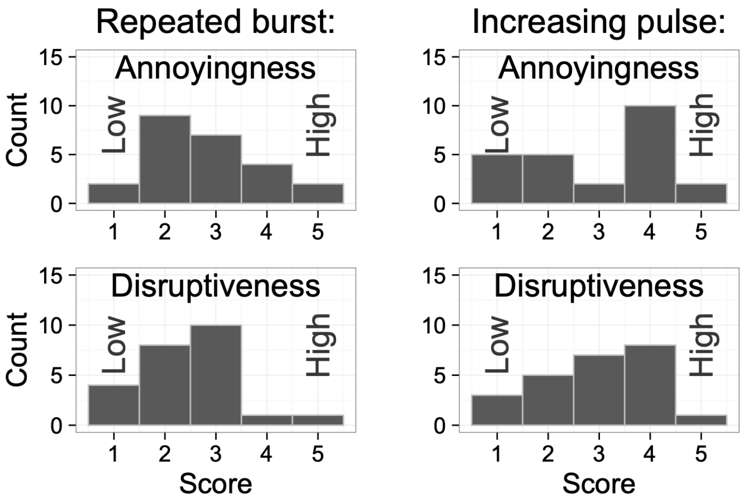


Figure 5.6. Subjective impression of pre-alerts

Heart rate

We also tested if the pre-alerts had any effect on drivers' physiology, specifically the average heart rate. Two drivers had to be excluded from this analysis because the measurement stopped during the experiment. The ANOVA found a significant effect of pre-alert on average heart rate $F(2,40) = 4.72$, $p = .015$, $\eta_p^2 = 0.18$. Post-hoc tests found that in No pre-alert (M = 69.8 bpm, SD = 6.7 bpm) the heart rate is significantly higher than in Repeated burst (M = 68.9 bpm, SD = 6.6 bpm, $p = .034$) and Increasing pulse pre-alert (M = 69.2 bpm, SD = 6.8 bpm, $p = .046$). There was no difference between Repeated burst and Increasing pulse ($p > .1$) In addition, heart rate was significantly higher in Dual-task conditions (M = 69.8 bpm, SD = 6.6 bpm) compared to Single-task (M = 68.8 bpm, SD = 6.8 bpm), $F(2,40) = 4.719$, $p = .031$, $\eta_p^2 = 0.21$. There was no significant interaction effect ($p > .1$) In summary, heart rates are slightly increased in dual-task conditions, and when there is no pre-alert and the user needs to do the extra task of frequently checking the road. This suggests that extra workload increased heartrate (cf. Mehler, Reimer, & Coughlin, 2012). Though the effect is small, the trend is consistent with subjective data.

5.4.3 RQ3: Success of handover

For the next two measures, we only analyzed the first segment of each driving scenario, as due to a coding error the driving speed was slightly lower on the second segment (60ft/s vs 65ft/s), which affected the time between the end of the pre-alert and the time given to hand-over. In the second segment, there was some delay between the warning that automation was turned off and the time that drivers could actually control the car. This reduces the reliability of the reaction time data on these metrics. This is not the case for the segments that we analyzed.

Initial reaction time

Previous work has mostly analyzed reaction times as a performance metric for handover. Due to the varied nature of our task, reaction to an event can be either braking or steering. We measure reaction time as time until either of these two actions occurs. We combine the data of single- and dual-task trials, resulting in three histograms for each pre-alert condition in Figure 5.7. Each plot shows data from 48 trials: 24 single, 24 dual. The bars cover 200 ms intervals. In general, most drivers respond within 200 ms (i.e., the first bar is the highest bar in each setting).

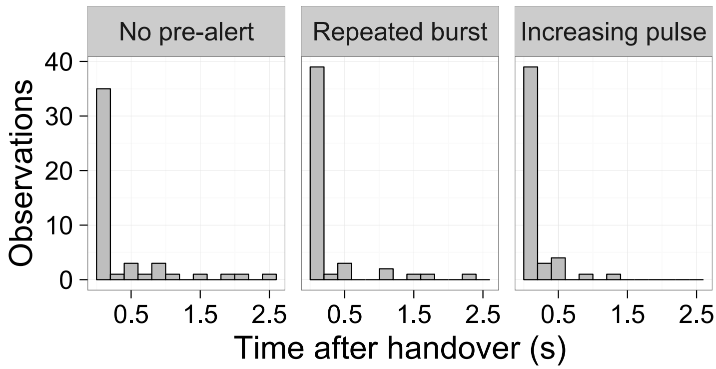


Figure 5.7. Time until first action after handover.

However, in driving analyses, we do not only care about the mean and majority of behavior, but also about extremes. This is where conditions differ. In the no pre-alert condition (left), the distribution is more right-tailed (9 trials with response longer than 600 ms, with extremes up to 2.5 s) compared to the repeated burst (5 trials) and increased pulse pre-alert (2 trials). Stated differently: in most cases, most drivers respond timely, but the trend is that more drivers respond timely when a pre-alert is given, with the number of late responses around twice as high in the no pre-alert condition.

We also analyzed driving speed after handover. In critical situations, reducing speed is a smart strategy as this creates more time for an effective response (as less distance is covered per time interval). Figure 5.8 shows how the average speed reduces over time in single-task (top figure) and dual-task (bottom figure) for the three pre-alert conditions.

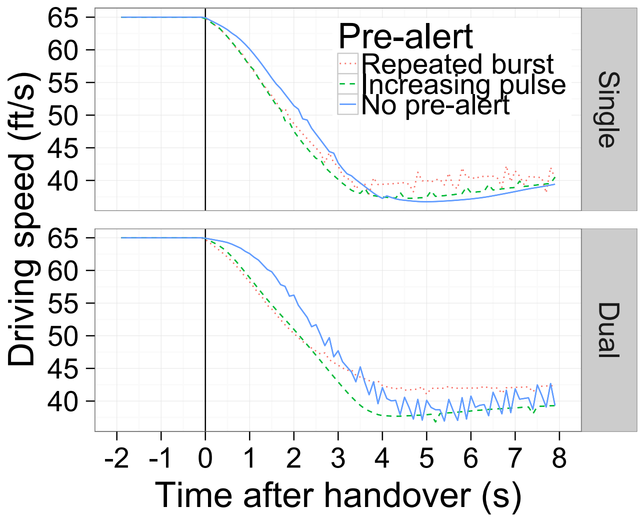


Figure 5.8. Driving speed after handover.

Before the handover (point 0 in the graph), the car drives automatically at a constant speed. Then, at point 0, drivers can take over. In the single-task condition, drivers reduce speed immediately in each condition, with a slight delay in the no pre-alert condition. However, if 95% CI intervals would be drawn, these would overlap between all three conditions, suggesting there is no difference. It shows that without distraction, drivers are prepared to respond.

This is different in the dual-task condition. Here, the brake response in the no pre-alert condition is delayed compared to the two pre-alert conditions. That is, drivers respond later. In fact, if 95% CIs would be drawn around the lines, the confidence intervals between the no pre-alert condition do not overlap with those of the two pre-alert conditions. A standard interpretation of a lack of overlap is that the conditions differ at a 95% confidence level (or with an alpha of .05) (Cumming, 2013).

The intervals of the two pre-alert conditions do overlap, indicating there is no significant difference.

Unsafe incident analysis

The final question is: Did the pre-alerts lead to less unsafe behavior? If drivers had more time to look at the road before handover to build situational awareness (RQ1), do they perform better in handling the handover incident?

For this analysis, we looked at all scenario segments (288) and labeled whether unsafe behavior had occurred such as not reducing speed, going into a lane where incoming traffic might occur, or crashing into an object. Of the 288 segments, 34 segments (11.8 %) were marked as unsafe. Given the low numbers, we report the frequency of unsafe behaviors in Table 5.1, split up by pre-alert type (rows), and single- or dual-task (columns). There is no clear emerging pattern. Specifically, incidents still occur in the two active pre-alert conditions. Moreover, it is not the case that there are more unsafe behaviors in the dual-task condition compared to single-task condition. If any, the trend in the data is that there were fewer unsafe behaviors in the increasing pulse condition (third row).

Due to this variety of results, we also analyzed whether unsafe behaviors were more frequent for some scenarios compared to others. This could not be done in combination with the type of pre-alert or single-/dual-task, due to the low numbers. In general, unsafe behavior was shown most in scenarios involving braking for a dog (23 segments). This was much higher compared to segments with a parked car (5 segments), construction works (6 segments), or fog (0 segments). Our interpretation of this result is that the dog scenario allowed more freedom to the user of what to do: they could either try to brake, or try to avoid it by driving past it. Moreover, as the object on the road (a dog) was relatively small compared to, for example, a parked car, drivers might not have had extra benefit from the early view of the object due to a pre-alert.

	Single-task	Dual-task	Total
None	7	6	**13**
Repeated burst	5	8	**13**
Increasing pulse	6	2	**8**
Total	**18**	**16**	**34**

Table 5.1. Number of segments involving unsafe behavior

5.5 General discussion

5.5.1 Summary of results

We investigated how pre-alerts affect the hand-over of control from a semi-autonomous car to a human driver. A recent review showed that drivers distract themselves more with other tasks as automation in the car increases, and that this impacts their situational awareness and ability to respond correctly (De Winter et al., 2014). Based on theory, we expect that a pre-alert has four benefits (1) it allows a driver the necessary time to disengage from a secondary task (Boehm-Davis & Remington, 2009), (2) this can reduce mental workload (Bailey & Iqbal, 2008; Salvucci & Bogunovich, 2010) and stress (Bailey & Konstan, 2006) and leave drivers in a better state to manage the handover, (3) this allows drivers more time to reorient to the driving task and gain relevant situational awareness, and (4) with sufficient time distracting effects from the non-driving tasks may be reduced (cf. Strayer et al., 2015). Our results demonstrate that pre-alerts are indeed beneficial. During the alerting phase (RQ1), drivers disengage from their non-driving tasks earlier when a pre-alert is given (Figure 5.5), and they look earlier at the road (Figure 5.4). In their experience (RQ2), drivers were divided in which alert they preferred, though both conveyed some urgency (Figure 5.6). We also found an effect on heart rate, which was slightly higher in the condition without a pre-alert. This finding needs to be replicated before conclusions can be drawn, but might indicate that a situation with no pre-alert is more stressful, as drivers cannot rely on the pre-alert to notify them. Finally, when looking at driving performance (RQ3), drivers responded faster (by braking or steering) to incidents when they were warned by a pre-alert (Figure 5.7), and reduced their speed more quickly (Figure 5.8). In effect this allows them more time to respond to an incident (i.e., at a lower speed it takes longer before an incident location is reached). Finally, there were still unsafe behaviors in all conditions, but these were the lowest with an increasing pulse pre-alert (Table 5.1).

Between the two pre-alert types, the preferences of drivers were divided. However, the trend in most metrics is that the increasing pulse pre-alert leads to slightly safer performance. For example, in this condition people disengage the earliest from a secondary task (Figure 5.5), more people experienced it as conveying high urgency (Figure 5.6), it had the lowest number of slow handovers (Figure 5.7), and the lowest number of incidents (Table 5.1).

Given this pattern, a general conclusion is that pre-alerts are useful compared to not getting a pre-alert. However, there is still room for improvement, as unsafe actions still occur.

5.5.2 Implications for theory

Our results confirm the classical result in driver distraction research (e.g., Alm & Nilsson, 1995; Briem & Hedman, 1995; Brookhuis et al., 1991; Janssen et al., 2012; Janssen , Iqbal, & Ju, 2014; Lee et al., 2001; Redelmeier & Tibshirani, 1997; Reed & Robbins, 2008; Salvucci, Mandalia, Kuge, & Yamamura, 2007) that secondary tasks distract from looking at the road (RQ1, Figure 5.4) and result in longer response times to incidents (RQ3, Figure 5.7). Our results also suggest that pre-alerts can mitigate some of these problems. This is particularly useful, as driver distraction occurs frequently in regular cars (Dingus et al., 2016; Klauer et al., 2014) and increases with an increase in autonomy of the car (De Winter et al., 2014).

The use of alerts for hand-over situations in semi-autonomous cars is of course not new. However, in contrast to earlier work (e.g., Gold et al., 2013), we focus on alerts that happen ahead of time (*pre*-alerts). Sending an alert too early (e.g., minutes) in advance might not make sense, as there is no situation that the driver can notice and start to anticipate. We focused on a pre-alert of 20 seconds, as previous work has suggested that distraction of secondary tasks can continue up to 27 seconds after the task was finished, with exponential decay (Strayer et al., 2015). Twenty seconds is therefore an interval that is needed to recover from distractions and to focus on the road.

In our study, drivers still incurred incidents in some of the alert conditions. There are multiple explanations, which require further testing in future studies. First and foremost, there were incidents even in single-task situations, suggesting that some tasks were difficult to handle in general. Second, in the distraction conditions drivers might have persisted too long with the secondary tasks even after the pre-alert and thereby not have taken enough time to react. Finally, similar to (Strayer et al., 2015), even when drivers did finish secondary tasks, drivers might have had remnants of distraction. The balance that needs to be found here is between giving a just-in-time alert such that it is meaningful, while also giving sufficient time to overcome any negative effects of distraction. Further studies are needed to get this balance just right.

Although both pre-alerts that we offered were effective, the increasing pulse performed better on some metrics compared to the repeated burst (e.g., Figure 5.5,5.7). Our interpretation is that this is because increasing pulse more clearly conveyed urgency (as also confirmed by the drivers, see Figure 5.6). This is also in line with theory (e.g., Politis et al., 2015).

As our study takes place in a simulator, there are concerns about how the findings generalize. For example, there are no serious consequences of a crash. However, it also offers many advantages for our setting. First, we can test behaviors in cars that have a level of autonomy beyond those that are currently widely available.

Second, we can test extremely dangerous situations such as hand-overs preceding crashes that would be unethical to test on the road. Third, we can measure behavior in-depth with multiple measures including eye-gaze, physiology, preferences, and driving performance. Finally, meta-reviews have demonstrated that situations that are shown to be dangerous in simulator conditions are also dangerous on the road (Caird, Willness, Steel, & Scialfa, 2008; Horrey & Wickens, 2006). However, the effect size of the performance decline can differ between the two situations. A specific prediction that our work makes for the regular road is that pre-alerts can be beneficial, but also that even an interval of twenty seconds might not be enough to respond appropriately to an incident in a handover situation. This is particularly the case because everyday traffic is more diverse and perhaps less predictable than our simulator scenarios.

5.5.3 Implications for design

Our findings suggest that pre-alerts can be helpful in managing handover situations. This opens up a large space of future work that explores what the exact nature of pre- alerts can be. Below, we discuss some relevant parameters.

Convey urgency

Our results suggest that pre-alerts should provide a sense of urgency (cf. Janssen et al., 2014; Politis et al., 2015), as in our increasing pulse pre-alert. However, there is still a wide design space to explore regarding exact choices. Relevant parameters include the exact length and timing of the pre-alert, the modality of the pre-alert, and the ability to perhaps also turn a pre-alert off.

Encoding more information in pre-alerts

Our pre-alert only used beeps to indicate that handover had to take place. However, it might be beneficial to also inform the user about why the pre-alert is raised (for example, is a sensor not working, does the car notice traffic), and what concrete actions they are to take (e.g., "Scan your surroundings to see whether you can come to a stop, or can go to another lane"). Preceding work has suggested that such concrete alerts are helpful in the automotive domain (Iqbal et al., 2011). We also used only one modality (audio) for the alert. Exploring multimodal alerts may result in better outcomes (Walch et al., 2015).

Balance effectiveness with less annoyance

Although the increasing pulse pre-alert trended to be slightly more effective out of the two pre-alerts, many drivers did not like it because it appeared to be annoying. Although the primary function of an alert is to increase safety, a distaste of the pre-alert might disrupt drivers too much. This would also counter two of the benefits of the pre-alert: to allow time to finish a task and to get in a low workload, low stress state. Future work should explore how design can counter this.

Timing of pre-alerts

While we selected a fixed time interval for the pre-alert (20 s before the handover) the timing could also be made dynamic depending on the type of event, level of distraction of the driver, and complexity of the required action. Dynamic timing helps address situations where a pre-alert occurring too early may diminish its urgency, or a pre-alert occurring too late may be deemed useless.

Although our focus has been on the automotive domain, our results can also be applied to other domains in which there is (A) shared control between humans and systems and (B) potential distraction. The implication there is that a pre-alert can benefit the shift of control from system to user. Our results suggest that pre-alerts that convey urgency, such as our increasing pulse pre-alert, are valuable. However, for each domain more tests are needed to determine the timing of these pre-alerts. This should take into account (1) the remnant effects of distraction (Strayer et al., 2015), (2) the time needed to finish any preceding task, and (3) the time that is needed to gain situational awareness in the domain at hand.

5.5.4 Limitations & Future work

We conducted our work in a driving simulator to allow for an in-depth study of human behavior, in an environment where the users cannot incur harm, but that is known from meta-reviews to translate to everyday driving (Caird et al., 2008; Horrey & Wickens, 2006). However, this also has limitations. First, there are no real risks for drivers and they might therefore have acted slightly riskier than in normal life. Second, given the experimental set-up, they might have anticipated some incidents and hand-overs which lowered response times compared to driving on the road. In everyday life, alerts might be rarer (Wolfe et al., 2007), which impacts response time. Third, the study was measured over a relatively short interval (90 minutes per participant) with various incidents, whereas normal driving has incidents less frequently. These limitations do not differ from other valuable driver distraction studies that used simulators, but are to be taken into account nonetheless.

The pre-alerts that we tested were limited in scope. We have discussed relevant parameters to explore. Some specific limitations are the following. First, the use of other modalities and multi-modal pre-alerts (cf. Politis, Brewster, & Pollick, 2014a; Politis, Brewster, & Pollick, 2013; Politis et al., 2015) needs to be tested. Second, we did not test voice-based commands despite their potential (e.g. Jeon, Gable, & Davidson, 2015; Politis, Brewster, & Pollick, 2014b; Politis et al., 2015). Finally, our drivers were not able to turn off the pre-alert, whereas this might be a relevant option in real cars, for example, to signal to the car that you noticed the pre-alert.

In our measurements we have tried to give a detailed description of human behavior. However, there is room to go even more detailed. First, our physiological state results were subtle and need to be replicated before solid conclusions can be drawn. Second, our eye-tracking results gave insight in whether and when drivers looked at the road, but more detailed analyses regarding *where* they look would also be beneficial (i.e., to understand what information people gather, what information they might have overlooked).

Finally, in our study we tested technology based on what technology looks like today (e.g., the Tesla model S) and predictions of what future states of shared control are possible (e.g., see Gasser & Westhoff, 2012; SAE International, 2014). However, the history of HCI has shown that interaction between humans and technology can change when disruptive technologies are introduced (e.g., GUIs, touch screens, smartphones). Similarly, currently unanticipated developments in the automotive domain may arise that might fundamentally change the interaction between drivers and cars. This might be particularly the case if fully automated cars without handover are developed (e.g., as in the Google vision). However, until that day we benefit from a basic understanding of human capacity (e.g., when do humans pay attention, how distracting is technology?).

5.6 Conclusion

Our results show that semi-autonomous cars benefit from pre-alerts that warn for a future handover situation. In particular, pre-alerts that reflect urgency, such as an increasing pulse signal, show high promise.

References

Alm, H., & Nilsson, L. (1995). The effects of a mobile telephone task on driver behaviour in a car following situation. *Accident Analysis & Prevention*, 27(5), 707-715.

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive science*, 26(1), 39-83.

Iqbal, S. T., & Bailey, B. P. (2008, April). Effects of intelligent notification management on users and their tasks. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 93-102).

Bailey, B. P., & Konstan, J. A. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in human behavior*, 22(4), 685-708.

Beattie, D., Baillie, L., & Halvey, M. (2015). A comparison of artificial driving sounds for automated vehicles. *In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 451-462).

Boehm-Davis, D. A., & Remington, R. (2009). Reducing the disruptive effects of interruption: a cognitive framework for analysing the costs and benefits of intervention strategies. *Accident Analysis & Prevention*, 41(5), 1124-1129.

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2010). The problem state: a cognitive bottleneck in multitasking. Journal of Experimental Psychology: *Learning, memory, and cognition*, 36(2), 363.

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2015). What makes interruptions disruptive? A process-model account of the effects of the problem state bottleneck on task interruption and resumption. *In Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 2971-2980).

Briem, V., & Hedman, L. R. (1995). Behavioural effects of mobile telephone use during simulated driving. *Ergonomics*, 38(12), 2536-2562.

Brookhuis, K. A., De Vries, G., & De Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention*, 23(4), 309-316.

Brumby, D. P., Howes, A., & Salvucci, D. D. (2007). A cognitive constraint model of dual-task trade-offs in a highly dynamic driving task. *In Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 233-242).

Brumby, D. P., Salvucci, D. D., & Howes, A. (2009). Focus on driving: How cognitive constraints shape the adaptation of strategy when dialing while driving. *In Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1629-1638).

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, 40(4), 1282-1293.

Cao, Y., Castronovo, S., Mahr, A., & Müller, C. (2009). On timing and modality choice with local danger warnings for drivers. *In Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 75-78).

Cumming, G. (2013). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis. Routledge.

De Winter, J. C., 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.

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, *113*(10), 2636-2641.

Fagerlönn, J., Lindberg, S., & Sirkka, A. (2012). Graded auditory warnings during in-vehicle use: using sound to guide drivers without additional noise. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 85-91).

Gasser, T. M., & Westhoff, D. (2012). BASt-study: Definitions of automation and legal issues in Germany. In *Proceedings of the 2012 road vehicle automation workshop*. Automation Workshop.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 1938-1942). Sage CA: Los Angeles, CA: Sage Publications.

Haas, E. C., & Casali, J. G. (1993). The perceived urgency and detection time of multi-tone and frequency-modulated warning signals. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 37, No. 9, pp. 544-548). Sage CA: Los Angeles, CA: SAGE Publications.

Hancock, P. A. (2017). Driven to distraction and back again. In *Driver Distraction and Inattention* (pp. 33-50). CRC Press.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human factors*, *48*(1), 196-205.

Illowsky, B., & Dean, S. L. (2008). *Collaborative statistics*. Rice University.

Iqbal, S. T., Horvitz, E., Ju, Y. C., & Mathews, E. (2011). Hang on a sec! Effects of proactive mediation of phone conversations while driving. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 463-472).

Iqbal, S. T., Ju, Y. C., & Horvitz, E. (2010). Cars, calls, and cognition: Investigating driving and divided attention. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1281-1290).

Janssen, C. P., & Brumby, D. P. (2010). Strategic adaptation to performance objectives in a dual-task setting. *Cognitive science*, *34*(8), 1548-1560.

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

Janssen, C. P., Iqbal, S. T., & Ju, Y. C. (2014). Sharing a driver's context with a caller via continuous audio cues to increase awareness about driver state. *Journal of experimental psychology: applied*, *20*(3), 270.

Jeon, M., Gable, T. M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2015). Menu navigation with in-vehicle technologies: Auditory menu cues improve dual task performance, preference, and workload. *International Journal of Human-Computer Interaction*, *31*(1), 1-16.

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England journal of medicine*, *370*(1), 54-59.

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *9*(4), 269-275.

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

Mahr, A., Pentcheva, M., & Müller, C. (2009). Towards system-mediated car passenger communication. In Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 79-80).

McFarlane, D. C. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction*, *17*(1), 63-139.

Meschtscherjakov, A., Wilfinger, D., Scherndl, T., & Tscheligi, M. (2009). Acceptance of future persuasive in-car interfaces towards a more economic driving behaviour. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 81-88).

Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: an on-road study across three age groups. *Human factors*, *54*(3), 396-412.

Mondor, T. A., & Finley, G. A. (2003). The perceived urgency of auditory warning alarms used in the hospital operating room is inappropriate. *Canadian journal of anaesthesia*, *50*(3), 221-228.

National Highway Traffic Safety Administration. (2013). *Preliminary statement of policy concerning automated vehicles*. Washington*, DC*, 1-14.

Naujoks, F., Mai, C., & Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. *Advances in Human Aspects of Transportation*, 7(Part I), 431.

Politis, I., Brewster, S., & Pollick, F. (2014a). Speech tactons improve speech warnings for drivers. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1-8).

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

Politis, I., Brewster, S., & Pollick, F. (2013). Evaluating multimodal driver displays of varying urgency. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 92-99).

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

Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine*, *336*(7), 453-458.

Reed, N., & Robbins, R. (2008). The effect of text messaging on driver behaviour: A simulator study. TRL.

SAE International. 2014. J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. Retrieved from http://standards.sae.org/j3016_201401/.

Salvucci, D. D. (2005). A multitasking general executive for compound continuous tasks. *Cognitive Science*, *29*(3), 457-492.

Salvucci, D. D., & Bogunovich, P. (2010). Multitasking and monotasking: the effects of mental workload on deferred task interruptions. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 85-88).

Salvucci, D. D., & Taatgen, N. A. (2010). *The multitasking mind*. Oxford University Press.

Salvucci, D. D., Markley, D., Zuber, M., & Brumby, D. P. (2007, April). iPod distraction: Effects of portable music-player use on driver performance. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 243-250).

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013). Measuring cognitive distraction in the automobile.

Chris Urmson, (2015, May 11). The View from the Front Seat of the Google Self-Driving Car. Medium.com. Retrieved from https://medium.com/backchannel/the-view-from-the-front-seat-of-the-google-self-driving-car-46fc9f3e6088

Walch, M., Lange, K., Baumann, M., & Weber, M. (2015). Autonomous driving: investigating the feasibility of car-driver handover assistance. In Proceedings of the 7th international conference on automotive user interfaces and interactive vehicular applications (pp. 11-18).

Wickens, C. D. (2008). Multiple resources and mental workload. Human factors, 50(3), 449-455.

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. Journal of experimental psychology: General, 136(4), 623.

Chapter 6

Visual in-car warnings: How fast do drivers respond?

Remo M. A. van der Heiden
Christian P. Janssen
Stella F. Donker
Chantal L. Merkx

Published as:
Van der Heiden, R. M., Janssen, C. P., Donker, S. F., & Merkx, C. L. (2019). Visual in-car warnings: How fast do drivers respond?. *Transportation research part F: traffic psychology and behaviour*, *65*, 748-759. DOI: https://doi.org/10.1016/j.trf.2018.02.024

Author contributions:
RMAV designed and programmed the experiment, collected the data, analyzed the data, and wrote the manuscript with input from the other authors. CPJ, SFD, and CLM supervised the project, contributed to experimental design, interpretation of the data, and editing the manuscript. All authors read and approved the final version of the manuscript.

Abstract

We investigate how quickly drivers can change lanes in response to a visual in-car warning. Our work is motivated by technological developments, in which beacons along the road can trigger in-car warnings, for example when a driver is approaching a lane closure. What is not known, however, is at what distance such an in-car warning still allows for a timely lane change. We measured how quickly drivers respond to a visual in-car warning in a driving simulator. The driving task was combined with an audio task that provided different levels of cognitive distraction. We found that the initial reaction time to in-car warnings was significantly larger for drivers that were distracted by the audio task. Although the majority of drivers responded in time for a safe lane change, some drivers occasionally missed these signals, pointing at a serious potential hazard. Indeed, the results of a simulation model, used to investigate how this might extrapolate to regular traffic conditions, suggest that around 50% of drivers might not make a timely lane change in response to a last-minute warning. This indicates that these signals might be insufficient on their own when applied in the real world. This work can inform the design and evaluation of safer roads and in-car interfaces.

6.1 Introduction

Roadwork sites have a higher crash rate than the same roads without roadworks (Khattak, Khattak, & Council, 2002). For example, a meta-review of accidents in the Netherlands (Van Gent, 2007) suggests that a failure to miss roadwork signs can result in various accidents such as rear-end collisions and crashing into roadwork safety trailers. Given the safety risks of such accidents, it is important to consider ways to prevent them.

One option to potentially reduce these accidents is the use of in-car technology to timely signal a critical event, such as an upcoming lane closure, to the driver. A specific technology that has been identified for this effort is to place beacons that can transmit wireless messages to cars equipped with an appropriate receiver (cf. IEEE, 2010). This technology can be used to display visual warnings in a car and as such allows for warning a driver for an upcoming traffic event. These beacons could be used on roadwork safety trailers (Figure 6.1) to inform drivers of an upcoming lane closure.

Figure 6.1. Roadwork safety trailer. Used with permission from Rijkswaterstaat. Source: https://beeldbank.rws.nl, Rijkswaterstaat

What is unknown, however, is whether an additional warning within the reach of the trailer beacon would allow the driver sufficient time to make a timely, and safe lane change. Current estimates of the reach of such beacons is approximately 500 meters (Gozálvez, Sepulcre, & Bauza 2012; Paier, Faetani, & Mecklenbrauker, 2010). Moreover, in some countries like the Netherlands these roadwork trailers are also preceded by portable rumble strips (also known as sleeper lines), which serve as a tactile warning and are the last resort to warn people to change lanes. Usage of rumble strips is prescribed by guidelines (e.g., for the Netherlands these are: CROW, 2013). In the Netherlands rumble strips are placed 150 meters before the trailer, which already takes up 150 meters of the beacon's reach. In effect, usage of rumble strips reduces the distance to make a lane change after receiving the in-car signal to 350 meters (500

- 150 meters), as also shown in Figure 6.2. However, is this enough distance to make a fast and safe lane change? We investigate this in the current study, by studying how fast drivers are able to respond to a sudden in-car warning that informs them about upcoming roadwork.

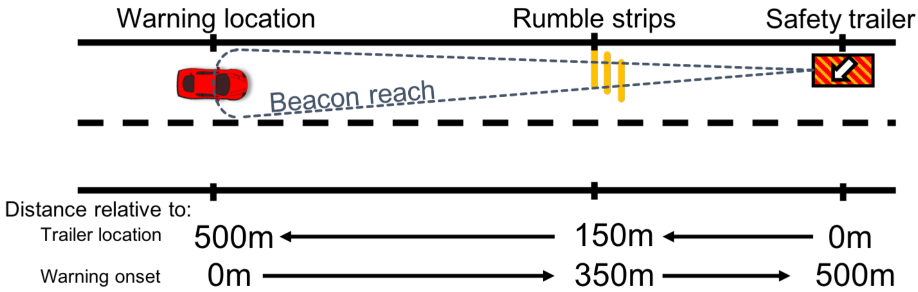


Figure 6.2. A typical roadwork scenario as you can encounter it in the Netherlands: a beacon, placed on a roadworks trailer, warns for an upcoming lane change. Distances can be expressed relative to the trailer and relative to the location where the car first received the warning, as determined by the beacon's reach (i.e., 500 m). Please note that in Dutch roadwork scenarios, a trailer is typically preceded by rumble strips. Our experiment provides a controlled version of this scenario, in which a warning is given, but no trailer or rumble strips are visible to the participant.

More specifically, we investigated lane change performance while placing drivers under different levels of cognitive distraction using an audio task (cf. Kunar, Carter, Cohen, & Horowitz 2008). Previous naturalistic driving studies have shown that drivers are distracted by various tasks such as eating, smoking, conversations, and making phone calls (Dingus et al., 2016; Klauer et al., 2014), similar to how distraction plays a role in many other professional and private situations (e.g., Janssen, Gould, Li, Brumby, & Cox, 2015). Even tasks that do not require you to take your hands off the wheel, such as holding a conversation (e.g., Iqbal, Ju, & Horvitz, 2010; Janssen, Iqbal, & Ju, 2014) or responding to cued words (e.g., Kunar et al., 2008; Strayer & Johnston, 2001), can distract from driving and result in longer response times. In effect, such distractions lead to a higher accident risk (Dingus et al., 2016; Klauer et al., 2014). It is therefore important that the performance of changing lanes within a distance of 350 meters is investigated under different levels of cognitive distraction.

6.1.1 Preceding studies on lane changing

Lane changing is, next to lane keeping, a common practice on motorways. A lane change is defined as a driver maneuver that moves a vehicle from one lane to another where both lanes have the same direction of travel (Fitch, Lee, Klauer, Hankey, & Sudweeks, 2009). Lane change has been the topic of several studies (e.g., Finnegan & Green, 1990; Hetrick, 1997; Wakasugi, 2005) in which, for example, timings, crash rate, and the use of indicators has been studied.

The Human Factors Guideline for Safer Road Infrastructure states that a lane change can be split up in two stages: an orientation and an approach stage. Each stage takes around 2-3 s before actual movement is initiated (Birth, Pflaumbaum, Potzel, & Sieber, 2009). However, in practice multiple action stages can be identified such as checking the mirrors, shoulder check, indicating direction of the lane change, and making a steering movement. Under some circumstances, a lane change could also involve braking or accelerating (e.g., to avoid crashing into a car that unexpectedly gets in a driver's lane).

In our simulator study, we look at the most basic version of a lane change, which consists of just the steering action. Our motivation for this choice is that measuring this action provides an estimate of the *minimum* time that is needed to change lanes. If even this minimum time is not sufficient to change within 350 to 500 meters (our context of interest), then this suggests that actual implementation of systems that require a lane change within 350 to 500 meters might need to be reconsidered.

6.1.2 Intended contribution and overview

The intended contribution of this work is twofold. First, we gain more theoretical insight on lane change behavior. Second, a practical contribution is to identify whether this lane change time is sufficient in our context of interest: beacons that provide a lane change alert 500 meters before a roadworks site and 350 m before rumble strips appear. This investigation is particularly needed, as 350 m is a lot shorter than current guidelines that are used in Dutch road design (Dubbeldam, 2007; Kroon, Brookhuis, Hagenzieker, & Martens 2014) allow (namely: around 675 m). Stated differently, our work can inform the evaluation of these guidelines in the face of new technology.

In this study we use a visual signal to deliver an alert. Although other types of alerts are possible in everyday systems (e.g., auditory alerts), we chose for a visual alert, for practical reasons. Our distracting task (discussed next) was presented auditory, and we did not want direct competition for cognitive resources (cf. Wickens, 2002, 2008). By separating visual signals (lane change warning) from auditory signals (secondary task), we minimize such interference.

In the remainder of this document, we will outline our driving simulator study in which we test how much time drivers need to react (i.e., a plain steering action without mirror checking or other gaze behavior) to an in-car signal and change lanes. After we present our empirical results, we use empirical data from Robinson and colleagues (Robinson, Erickson, Thurston, & Clark, 1972) to predict how these results would extend to real traffic situations where there is (need for) gaze behavior. Robinson and colleagues made detailed measurements of eye-gaze under different traffic conditions, which we apply to see how much longer lane changes are expected to last when made under traffic conditions. Hence, this simple model allows us to interpret how our empirical results might scale to an on the road situation where a lane change involves more steps, for example due to the presence of other traffic and the need for mirror- and shoulder checks.

6.2 Method

We tested how quickly drivers react to a visual in-car warning in a driving simulator set-up. Participants had to change lanes after receiving a visual in-car warning of an upcoming lane closure. We combined this lane change task with an audio task that created different levels of central cognitive interference: no audio task, simply repeating a word, or generating a word based on the last letter of the audio stimulus word (Iqbal et al., 2010; Kunar et al., 2008; Strayer & Johnston, 2009). Since in-car and external distraction is diverse (Dingus et al., 2016; Klauer et al., 2014), we chose this audio task as it is known to create central interference (Kunar et al., 2008; Strayer & Johnston, 2009). Central interference is independent of the modality of the task interaction (e.g., is not specific to visual, manual, or auditory interaction). By studying central interference, our setting therefore generalizes to various task interaction and modality settings. Such generalization is needed, given the diverse types of tasks that drivers perform in their cars (Dingus et al., 2016; Klauer et al., 2014).

6.2.1 Participants

Twenty-four participants (9 women; 15 men) ranging in age from 28 to 70 years (M = 46.5 years, SD = 12.4 years) were recruited by a recruiting company using quota sampling. The sample matches the distributions of highway drivers in the Netherlands on age, gender and yearly driving distance as found in a population-based study (CBS, 2013). Each participant had a driver’s license for at least 3 years, drove at least 5,000 km per year (M = 14,000 km), and had normal or corrected-to-normal vision. Only Dutch native speakers were recruited as the test involved verbal reaction to the audio task in Dutch. Participants self-reported to have not visual or auditory difficulties. All participants gave written informed consent, and were compensated with 35 euro.

A. B. C.

Figure 6.3. **A:** Driving simulator hardware setup, **B:** Simulated highway, with the simulated navigation display situated in the bottom right, **C:** Close-up of the simulated navigation device with closed lanes warning.

6.2.2 Material & Stimuli

Participants sat on an adjustable fixed chair in front of a Logitech G27 racing wheel and a 29" monitor (see Figure 6.3A). In both single and dual-task condition the lane change performance was studied in the context of a visual warning. In the dual-task trials, participants were asked to steer a simulated car while performing an audio task at the same time. We will explain these individual tasks next.

Driving task. The driving task was developed in a modified version of OpenDS 2.5[3]. OpenDS (Math, Mahr, Moniri, & Müller, 2013) has been applied in over 40 psychological studies worldwide (See http://opends.de/community/studies-using-opends). The driving task was to stay in the middle lane (3.5 m wide) of a straight three-lane highway. A simulated navigation system was shown at the bottom right of the screen (see Figure 6.3B). This representation approximates the size of a Dutch navigation system. The symbols (visual cues) were about 1 degree in visual angle. When a lane closure was imminent, the interface showed which lanes were closed (red crosses), following symbols that are used on Dutch highways to indicate closed lanes. Participants were instructed to change to the open lane once they noticed the alert.

[3] We changed the physics of the default OpenDS implementation of a car, such that our car would accelerate instantly instead of gradually.

The open lane was either to the left or the right of the central lane. The visual cue for lane closure was only shown on the interface, not on the road.

The car drove at a constant speed of either 80 km/h or 130 km/h (on Dutch highways the maximum speed is always between these values). Within a block of trials, participants changed lanes 6 times (half left, half right). For each lane change a trajectory of 1000 m was used. The upcoming lane closure was signaled at one of six locations (225, 275, 325, 375, 425 or 475 meters after trial start), which were balanced over trials and subjects. At the end of each trial the car was automatically reset to the middle lane to begin a new trial, and the navigation screen was cleared (showing only a green background).

Audio task. We compared lane change performance in reaction to a visual warning in single-task driving to two situations where an audio task created distraction. The audio task is similar to the verb-generation task that is commonly used in psychology (Abdullaev & Posner, 1998; e.g., Snyder, Abdullaev, Posner, & Raichle, 1995) and is a task that is known to result in central interference in cognitive processing (Kunar et al., 2008; Strayer & Johnston, 2009). In both our audio conditions (i.e., *repeat* and *generate*), participants heard a stream of words, presented at a steady pace of 1 word every 4 s. In the repeat condition, participants simply had to repeat the word they heard. In the generate condition, participants had to respond with a new word that started with the last letter of the word they heard. For example, if they heard “plot”, they could respond with “top”. Previous work has suggested that the generate condition creates more central cognitive interference compared to the repeat condition in dual-task situations (cf. Kunar et al., 2008; Strayer & Johnston, 2009).

To generate the word set, we selected 4 letter Dutch words from an online database (www.woordenraden.nl). For each letter of the alphabet, we identified a set of words that started with that letter. There were no words that started with Q or V. From the initial set, we then manually removed words that were uncommon in Dutch language, or emotionally loaded, as judged by the experimenter. If more than 20 words remained for a particular starting letter, at most 20 words were selected. For some letters (e.g., words starting with 'x') fewer words remained. The final set consisted of 315 words.

The words were converted to audio via a text to speech processor and were played over headphones. The participants’ vocal response was recorded with the headphone's microphone. Recording was done in two separate streams: one stimulus stream and one response stream. In addition, the experiment leader used a paper form to keep track of which words had an incorrect response or other error. This detailed multi-faceted recording allowed for accurate offline evaluation of time intervals between word presentation and first response.

Subjective experience. After each condition, subjective experience was assessed using raw TLX (Hart & Staveland, 1988).

6.2.3 Design

We used a 2 (Driving speed: 80 km/h, 130 km/h) x 3 (Audio task: No audio, Repeat, Generate) within-subjects design. Conditions were blocked by speed level. Half the participants started with 80 km/h, the other with 130 km/h. Within each speed condition, participants completed all audio conditions in a blocked fashion. All participants started with the *no audio condition*. This was followed by the *repeat* and *generate condition*, of which the order was counterbalanced. The order of audio conditions that was used for the first speed level, was also used for the second speed level.

6.2.4 Procedure

Upon arrival, the procedure was explained and an informed consent form was read and signed. Next, all tasks were practiced to ensure that participants were familiar with them. First, the repeat and generate audio task were practiced with ten samples each. Second, participants practiced two single-task driving trials for the 130 km/h speed level and followed by one dual task trial.

Participants then performed six experimental blocks. After each block, participants filled out the TLX questionnaire. Finally, participants completed 20 single-task audio trials (10 repeat and 10 generate trials), to assess single-task audio performance. The total procedure took approximately 70 minutes. Participants were allowed to take short breaks in between blocks.

6.2.5 Measurements

We were interested in how lane change times change as a function of driving speed (80 km/h vs 130 km/h) and the amount of distraction (no audio, repeat, generate). To this end, we analyzed four facets of behavior:

1. *Initial reaction time* was defined as the time it took before the first steering movement exceeding 1 degree was made (T1 in Figure 6.4) relative to the onset of the in-car warning (T0). The initial reaction time is a proxy of how long it takes drivers to initiate a lane change after first stimulus onset. For each condition, we calculated drivers' mean initial reaction time.

2. *Lane change distance* was defined as the distance that was traveled between when the in-car warning showed (T0) and the moment at which the car was fully in the target lane (T2). The criterion for considering the car to be fully in the target lane was that the full body of the car passed the center of the lane markings. Note that the simulator only logged the timestamp of the lane change (i.e., T2), not the distance. The distance was calculated based on the known constant speed and the measured lane change time.
3. *Audio reaction time* was measured as the time it took to respond after an audio cue was played, as a proxy of cognitive distraction. We used the two audio streams for this analysis: a stream that recorded the audio stimuli and a stream that recorded the participants' vocal responses. Both streams were stored in one file and thereby synced in time. Audacity (audacity.sourceforge.net) was used to automate the search of audio responses. A threshold of 20dB was used to discriminate silence (including slight noise) from audio responses. The automated procedure only detected audio responses that were at least 1.5 seconds apart within the same audio stream. After the automatic detection, each response stream was checked and corrected manually for artifacts such as "Uhhh", loud breathing, and coughs. In cases where there was no audio response to a stimulus, this stimulus-response pair was removed from the analysis. For three participants, this analysis was hindered by continuous loud breathing (i.e., which connected multiple sound fragments). These participants were removed from the audio analysis.
4. *Subjective workload* was measured using the TLX questionnaire, where participants scored their workload on a scale from 1 to 20 on various subscales (Hart & Staveland, 1988). We analyzed the score on the mental demand subscale.

Unless otherwise indicated, all metrics were analyzed using a 2 (Driving speed: 80, 130 km/h) x 3 (audio condition: none, repeat, generate) analysis of variance (ANOVA) with an alpha of .05 for significance. Holm-Bonferroni-corrected post-hoc tests were applied in the case of pairwise comparisons.

For some trials, the simulator output indicated that participants failed to change lanes within the maximum measurement time (10 seconds). For the analysis of initial reaction time (T1-T0), these trials were removed. However, for the metric of lane change distance, we replayed video clips of individual trials to verify the cause of the miss. In cases where participants simply did not change lanes (7 trials) or had not fully changed lanes before the measurement ended (11 trials), total lane change times (T2-T0 in Figure 6.4) were set to the total measurement time: 10 seconds. This can be considered a minimum estimate of how slow drivers respond to an in-car warning, as the longest reaction time is hard-coded to 10 seconds (i.e., if the trial finished later, then reaction time might have taken even longer).

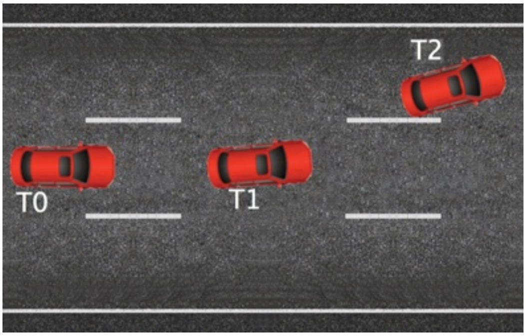


Figure 6.4. Measurements overview showing the critical timestamps in the experiment. At **T0** the visual warning is first presented. At **T1**, the participant makes the first (bigger) steering action. At **T2** the car is fully in the target lane.

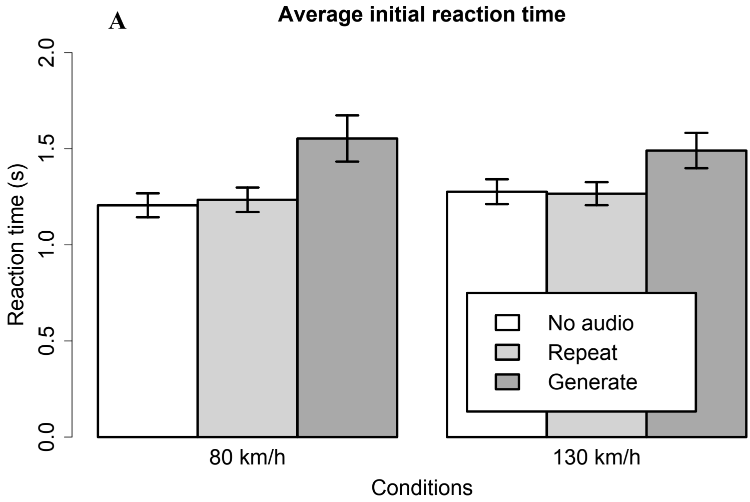
6.3 Results

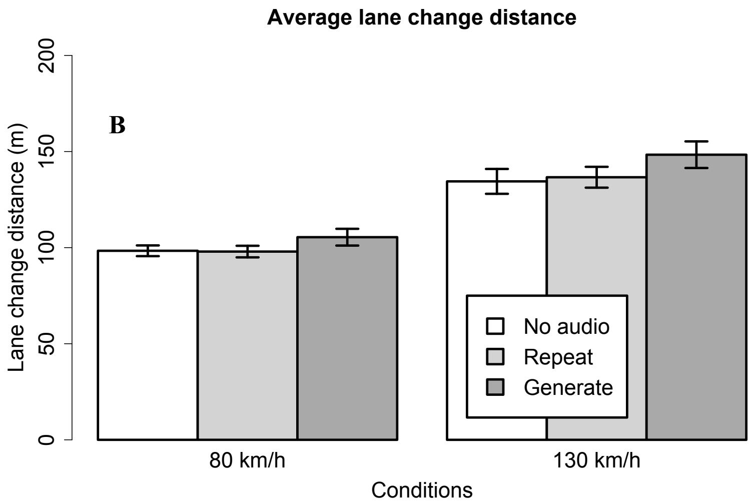
6.3.1 Initial reaction time (T1-T0)

Figure 6.5A presents the initial reaction time in all conditions (i.e. time interval T1 – T0 in Figure 6.4). There is a significant effect of the generation task on reaction time, $F(2,46) = 8.398$, $p = < .001$, $\eta_p^2 = .27$. A post-hoc test confirmed that reaction times were longer in the *generate* condition ($M = 1.52$ s, $SD = 0.52$ s), compared to the *repeat* ($M = 1.25$ s, $SD = 0.30$ s, $p = .006$) and *the no-audio* condition ($M = 1.24$ s, $SD = 0.30$ s, $p = .006$). There was no significant difference between the *no-audio* and *repeat* condition ($p = .90$). The initial reaction time at 80 km/h ($M = 1.33$ s, $SD = 0.45$ s) was not significantly different from that at 130 km/h ($M = 1.35$ s, $SD = 0.37$ s), $F(1,23) = .042$, $p > .1$. There was no significant interaction between speed and audio condition, $F(2,46) = 0.803$, $p > .1$.

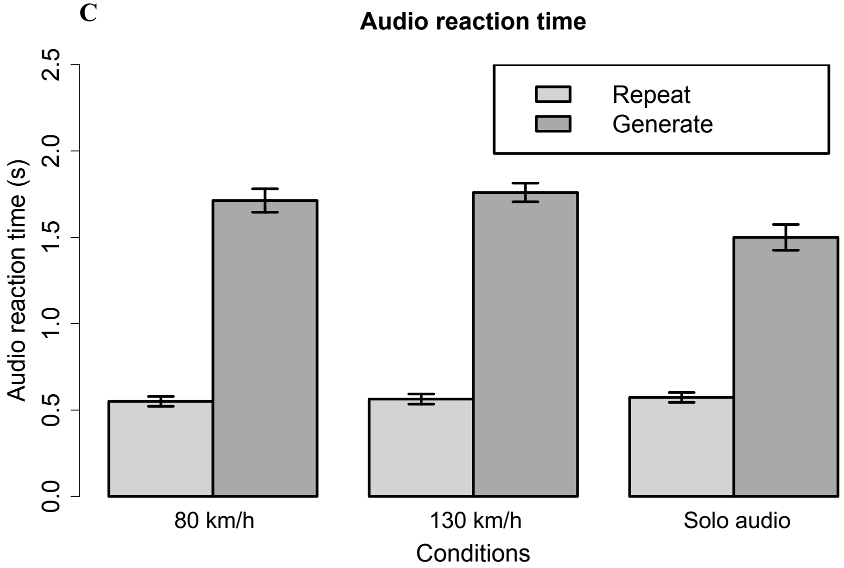
6.3.2 Lane change distance (distance covered during T0 until T2)

Figure 6.5B shows average lane change distance. Overall, the lane change distance was significantly shorter at 80 km/h ($M = 101$ m, $SD = 17$ m) than at 130 km/h ($M = 140$ m, $SD = 31$ m), $F(1,23) = 80.96$, $p < .001$, $\eta_p^2 = .78$. There was also a significant effect of audio condition, $F(2,46) = 3.434$, $p = .041$, $\eta_p^2 = .13$. Although the mean lane change distance was largest in the *generate* condition, a post-hoc test did not find any significant difference between the generate ($M = 127$ m, $SD = 35$ m), repeat, ($M = 117$ m, $SD = 29$ m) and no audio condition ($M = 116$ m, $SD = 30$ m) (all $p > .1$) There was no significant interaction between speed and audio condition, $F(2,46) = 0.686$, $p > .1$.


A
Average initial reaction time
Reaction time (s)
0.0
0.5
1.0
1.5
2.0
No audio
Repeat
Generate
80 km/h
130 km/h
Conditions


Average lane change distance
B
Lane change distance (m)
0
50
100
150
200
No audio
Repeat
Generate
80 km/h
130 km/h
Conditions

C



D

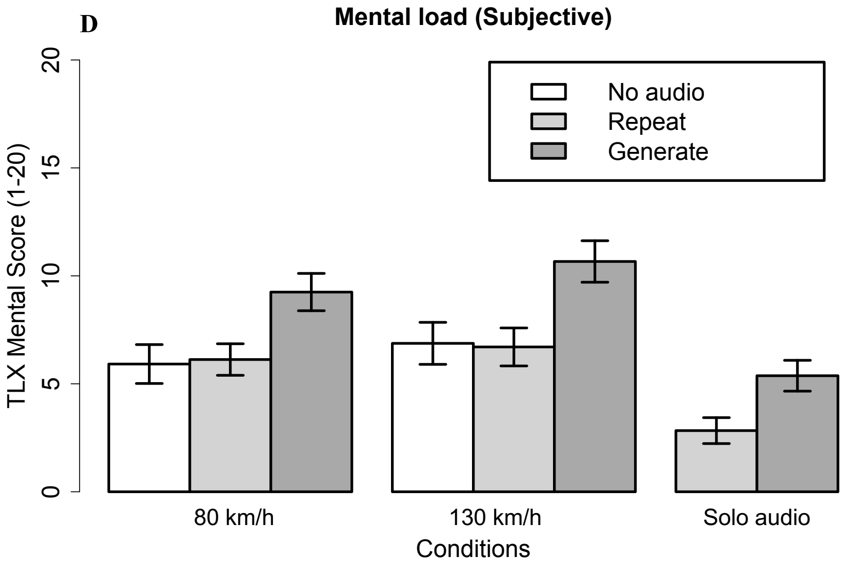


Figure 6.5. A: Bar graph of the average initial reaction time (T1). The graph shows that participants react significantly slower in the generate condition compared to no audio and repeat. **B:** Bar graph of total average lane change distance (at T2) in different conditions. **C:** Bar graph of the average audio reaction time to the different audio conditions. Responses took longer in the generate condition compared to the repeat condition. **D:** Bar graph of the average indicated perceived mental load by participants. The generate condition was perceived significantly higher on mental load compared to the repeat and the no audio condition. All error bars indicate standardized error of the mean.

6.3.3 Audio reaction time

Figure 6.5C shows the average reaction time on the audio task. The audio data was analyzed using a 2 (audio condition: repeat or generate) x 2 (driving speed: 80, 130 km/h) ANOVA. The average audio reaction time was significantly higher for the generate condition (M = 1.7 s, SD = 0.3 s) compared to the repeat condition (M = 0.6 s, SD = 0.1 s), $F(1,20) = 440.9$, $p < .001$, $\eta_p^2 = .96$. There was no significant difference between the two driving speeds $F(1,20) = 1.04$, $p > .1$. There was also no significant interaction between audio condition and driving speed, $F(1,20) = 0.33$, $p > .1$.

We performed an additional 2 (audio condition: repeat or generate) x 2 (number of tasks: single- or dual-task) ANOVA to test whether dual-task audio reaction time differed from single-task audio reaction time. Again, audio reaction time was significantly higher for the generate condition (M = 1.7 s, SD = 0.3 s) compared to the repeat condition (M = 0.6 s, SD = 0.1 s), $F(1,20) = 333.4$, $p < .001$, η_p^2 = .94. We also found a significant effect of the number of tasks, $F(1,20) = 11.96$, $p = .002$, $\eta_p^2 = .37$. However, there was also a significant interaction between audio condition and number of tasks, $F(1,20) = 21.22$, $p < .001$, $\eta_p^2 = .51$. Post-hoc tests show that for the generate condition, performance is significantly slower in the dual-task compared to the single-task condition ($p < .001$). However, for the repeat condition there was no significant difference between single- and dual-task ($p > .1$). Taken together, these differences are consistent with the literature that generating new words is more demanding than repeating a word and therefore causes more dual-task interference (Kunar et al., 2008; Strayer & Johnston, 2001).

6.3.4 Perceived mental load (subjective)

Figure 6.5D shows the perceived mental load indicated on the raw TLX questionnaire. There was a significant difference between the three audio conditions on perceived mental load, $F(2,46) = 37.71$, $p < .001$, $\eta_p^2 = .62$. A post-hoc test showed that perceived load was significantly higher in the generate (M = 8.4, SD = 4.7) compared to the repeat condition (M = 5.2, SD = 4.0; $p < .001$), load was also significantly higher in the generate condition compared to the no-audio condition (M = 6.4, SD = 4.6; $p < .001$). However, there was no significant difference between repeat and no-audio ($p > .1$). There was also a significant effect of driving speed on perceived mental load, $F(1,23) = 5.189$, $p = .003$, $\eta_p^2 = .18$. Mental load while driving at 130 km/h is significantly higher (M = 8.08, SD = 4.89) compared load at 80 km/h (M = 7.09, SD = 4.31). There was no significant interaction between speed and audio condition, $F(2,46) = 0.722$, $p > .1$. As can be seen in Figure 6.5D, in general the perceived mental load was lowest when the participant was not driving at all (solo audio condition – the two right-most bars).

6.4 Discussion of results

Taken together, the results present a coherent pattern. When drivers are distracted by a secondary task, the impact on lane changing depends on the nature of the secondary task. In particular, the generate task created more subjective workload (Figure 6.5D), increased reaction times on the audio task (Figure 6.5C), and in effect also delayed the initiation of a lane change (Figure 6.5A). This pattern is consistent with previous work that demonstrated in other dual-task settings that a word generate task is more distracting than a repeat task and a no-audio task (Kunar et al., 2008; Strayer and Johnston, 2001).

The distance that was traversed to complete a lane change (Figure 6.5B) was influenced by the speed of the car, but not by the type of distracting secondary task. An effect of secondary task might be lacking on this metric, because the effect that the secondary tasks have on initial reaction time is relatively small (around 300 ms) compared to the time it takes to complete a lane change (e.g., 3900 ms for 130 km/h, 4500 ms for 80 km/h). Therefore, by the time the lane change is completed the initial difference might have been compensated for. The average lane change distance was well within the required distance of our context of interest (i.e., within 350 m). However, the traffic circumstances in the simulator were relatively ideal (i.e., no other traffic). In addition, as we will see next, there were some important individual variations.

In our experimental study we had to control for other sources of variation such as mirror and shoulder checks, to allow more accurate measurement of the impact that auditory distraction might have on a lane change. Although this allowed for better detection of effects, we are aware that actual lane changes involve more steps. We will address these issues next using a simulation model where we integrate data from previous studies into our findings, this way we are able to get an idea of the effects of these additions to a more complete lane change task. The simulation model uses previous empirical data to account for these other known sources of variability, to make a projection of the expected distribution of lane change distances in less controlled (non-lab) environments.

6.4.1 Analysis and model prediction of distribution of distances

Analysis of distribution of human data

Our preceding analyses have focused on average reaction times and average lane change distances. For everyday life, however, it is not the average that matters, but the distribution of data – as the whole spectrum of drivers from slowest to fastest needs to be anticipated in the design of roads and technology. In particular, even if on average most drivers make a lane change on time, this is not a safe situation if one or more drivers does not make it and ends up in a crash.

To understand the distribution of lane change distance data, Figure 6.6 shows a histogram of individual trial data, with one figure for all trials (i.e., showing all audio conditions) from the 80 km/h condition (top) and one for the 130 km/h condition (bottom). In this figure, the horizontal axis shows what distance is traveled before a lane change was completed, with 0 being the starting point at which the signal was given on the in-car device. In the scenario that we introduced in the introduction, a roadwork safety trailer would be standing at 500 meters, and this would be preceded by rumble strips at 350 meters (i.e., 150 meters before the roadwork safety trailer) (CROW, 2013). Both distances are highlighted with a vertical bar. Our measurement lasted for a total of 10 seconds after the warning was shown. The bump at the end of each distribution represents the measurements where people had not fully changed to the target lane at that point and where we took the full 10 seconds as value. These measurements (18 trials in total) would normally be part of the tails of the distributions, which is now absent due to the hard 10 second cutoff.

As the figure shows, even in our relatively simple set-up with 'ideal circumstances' (e.g., no other traffic), in some trials drivers did not react in a timely manner to prevent driving over the rumble strips. Taken together, these data suggest that implementing a last resort signal at 500 meters before roadworks start might not be timely for at least some drivers.

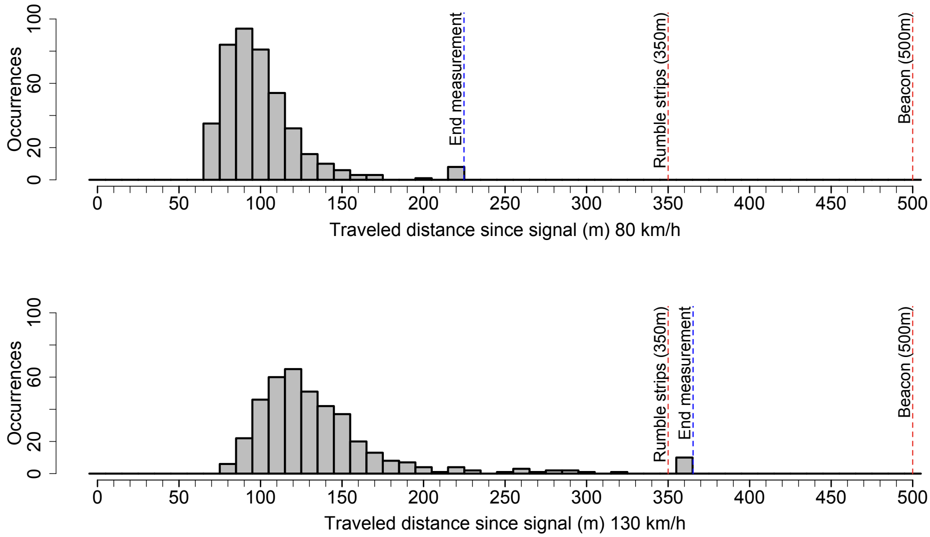


Figure 6.6. Distribution of lane change distances since onset of the in-car signal (0 m). For comparison, we indicated the point at which a beacon is placed (500 m), the start of rumble strips (350 m) and the point at which our measurement stopped (361 m for 130 km/h, 222 m for 80 km/h; this is 10s after the onset of the lane change signal). The top figure shows distances when driving at 80 km/h, the bottom figure shows distances for 130 km/h.

6.4.2 Model Prediction of distance in real-traffic

The above scenario is relatively simple. For example, there is no other traffic and therefore no need to look in the mirrors and to signal a lane change before changing. In real traffic, these steps would be necessary. We therefore investigated what the impact of these steps is using a simulation model that takes these steps into account. The use of simulation models to predict human behavior in traffic situations is valuable (e.g., Brumby, Janssen, Kujala, & Salvucci, 2018; Horrey & Wickens, 2006; Janssen & Brumby, 2010; Janssen, Brumby, & Garnett, 2012; Salvucci, 2001; Salvucci et al., 2007). One advantage of such models is that they can give new insights about situations that are challenging to observe in everyday life, for example because they concern rare events. Moreover, models can give an indication of how likely such scenarios or rare events are to occur

The starting point of our model is the article by Robinson and colleagues (1972), who measured lane-change times on the highway for eight drivers that drove at a slow speed (48 km/h). The researchers explicitly assessed the time that each aspect of visual search, related to the lane change, took: looking in the mirrors, making a head-movement to check for other traffic, and remaining visual search time (see Table 3 in Robinson et al., 1972). For our analysis, we took the data for a scenario with other traffic when merging to the left (M = 6.10 s, SD = 0.56 s) or to the right (M = 4.53 s, SD = 0.30 s).

We then ran Monte Carlo simulations to estimate the resulting lane change times when our human data of 80 km/h and 130 km/h would be combined with the estimated glance times from Robinson and colleagues. For each speed condition and each turn direction (left or right), we ran 100,000 simulations. For each simulation, we took a random data point from our experimental data of that speed condition, and combined this with a sample that was drawn from the distribution of the corresponding turn direction as measured by Robinson and colleagues (e.g., a normal distribution with M = 6.10 s, SD = 0.56 s for left turns). This resulted in four datasets with 100,000 duration estimates each. Using the driving speed, these lane change times were then transformed in lane change distances (expressed in meters).

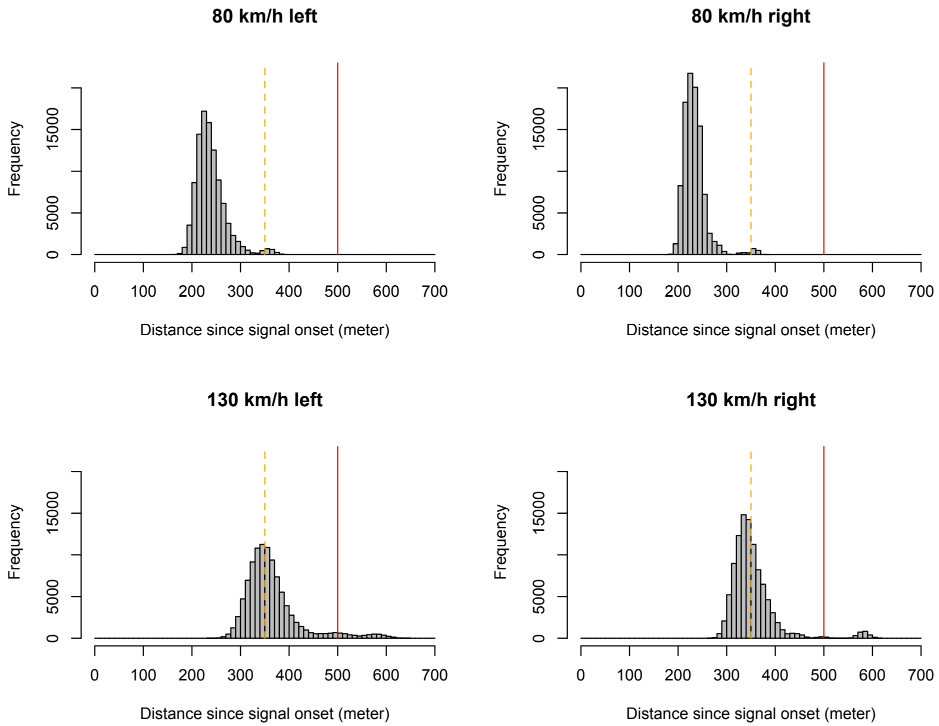


Figure 6.7. Histogram of model predictions of lane change distances when making a turn to the left (left plots) or right (right plots) at 80 km/h (top plots) and at 130 km/h (bottom plots). A red solid bar indicates the location of the roadworks (500 m after signal onset) and dashed orange line indicates the location of the rumble strips (350 m).

Figure 6.7 shows the resulting distributions. The top figure shows data for the 80 km/h condition, the bottom row for the 130 km/h condition. The left plots show data for changing lanes to the left, the right plots show data for changing to the right. Each figure again plots a line for the location of the roadworks (500 m; red solid line) and of the rumble strips (350 m; orange dashed line).

When comparing Figure 6.7 to Figure 6.6, it becomes apparent that the added glance times have shifted all distributions further to the right. At 80 km/h (top), between 1% (right lane change) and 2% (left) of the simulated data points crossed the 350 m border (rumble strips), and none crossed the 500 m border.

For the 130 km/h condition, results are more dramatic. The 350m border is crossed by between 42% (right lane change) and 53% (left) of the simulated data points. The simulation predicts that the roadwork safety trailer (500 m) would even have been hit by between 3% (right lane change) and 5% (left lane change) of the instances.

Using these simulations, we can also get a better estimate of the minimum distance at which one would need to place the rumble strips such that no driver (in this case: no simulation) might hit them. This is the furthest distance after which the model might have changed in time. In our simulation, this was 618 m. Assuming that a roadwork safety trailer is placed 150 m further down the road (compared to where the rumble strips starts), an in-car signal would need to be received at least 770 m before the position of the roadwork safety trailer.

As these data are based on simulations, the exact values are less informative than the general trend in the data. The trend is very consistent: in situations where drivers rely on the last-minute warning, sending this warning 500 m before the roadwork safety trailer will not be early enough for over 50% of the drivers. In order to make a safe situation, the distance between the signal and the roadwork safety trailer needs to be increased with over 50% of the current technological estimated capacity (Gozálvez et al., 2012; Paier et al., 2010): from 500 m to 770 m.

Even this estimate might still be a minimum for three reasons. First, whether one can change lanes will also depend on how much traffic is present. Changing lanes on a busy road can take more time. This is included implicitly in our model, as the model sampled from a distribution of data that included other traffic (Robinson et al., 1972). This same article showed that changing lanes took at least 1 second longer when other traffic is present (the data we used) compared to when no traffic was present. However, as traffic volume increases, this time might increase further. Second, in our dataset some participants did not change lanes within the given time frame. We assigned a lane change time of 10 seconds to these cases, but in reality, this might be more. In Figure 6.7 these data points lead to the slightly higher bump at the far end of the tail of the distribution. In reality, this tail would be distributed over points to the right of it. Third, and finally, in our experiment participants knew that they would have to change lanes in every trial. This is not always the case in real driving.

6.5 General discussion

We investigated how fast drivers respond to visual in-car warnings. Our experimental results demonstrate that although participants in general were able to change lanes in time (i.e., within 350 m after receiving an in-car warning), there were also participants that missed these warnings (see Figure 6.6). In practice, such participants would have hit the rumble strips. In our experimental results, the misses happened both in conditions where drivers were performing a secondary audio task, and in the condition where they were not.

A model simulation suggests that under real traffic conditions, the estimated number might be even higher: over 50% of drivers might not have enough time to make a lane change before they hit rumble strips, and around 3 to 5% of drivers might even hit the road works (see Figure 6.7). Although the exact number might be slightly different when measured on the road, as compared to the prediction of our model, the consistency of the findings in the model give us high confidence in our main conclusion that 500 meter is not sufficient time to change lanes.

To prevent any accidents, the distance between the presentation of the in-car alert and the rumble strips needs to be around at least 620 meters. Hence the warning distance needs to be 770 meters (i.e., 620 + 150 after the rumble strips), since the warning is transmitted from the trailer, not from the rumble strip location. This contrasts with the way in-car signals can currently be implemented (cf. Gozálvez et al., 2012; Paier et al., 2010): on top of the first roadwork safety trailer (see introduction). Therefore, other alternatives need to be explored.

Different levels of cognitive load were used to see how lane change times varied with distraction (cf. Kunar et al., 2008; Strayer & Johnston, 2001). We found that the type of distraction significantly affected the initial reaction time to an in-car warning. These results are in line with previous work (Kunar et al., 2008; Strayer & Johnston, 2001). However, this previous work did not investigate how the impact of distraction continues after an initial reaction (i.e., after T1 in our study). We found that on average this difference did not persist in the total lane change time. This is most likely because the effect size of the differences between initial reaction times is small (300 ms) compared to the time that the entire lane change takes (around 4.2 s). The expected impact on everyday driving of this effect is hard to anticipate: being distracted mostly seems to impact initial reaction to a signal (in our study: time between T0 and T1, see Figure 6.4). In cases where these delays are small (as in our study), later actions (e.g., between T1 and T2) might be able to compensate for some delays. However, this might not be the case if the initial reaction time is delayed severely.

6.5.1 Limitations and future work

Our study has investigated reaction time in the context of a visual warning and auditory distraction. In real traffic, in-car (visual) warnings might have been preceded by other external or internal warnings (e.g., road signs, or early warnings such as described in Van der Heiden, Iqbal, & Janssen, 2017). Our work is meant to investigate what happens if the in-car warning is truly a last resort, which is one of the reasons such alerts are being designed. As the empirical and simulation results suggest, it might not be sufficient when a last resort warning is provided at 500 m distance from the road works. Future (empirical) work can study at how this changes when the signal is combined with other (on road or in-car) warnings. Moreover, such research could consider whether in these cases an additional signal might be confusing, conflicting, or distracting — for example, in cases where the driver has already changed lanes.

The in-car alert in our study was only visual. We chose this modality, as it is relatively less invasive compared to for example an audio alert (e.g., Cao, Castronovo, Mahr, & Müller, 2009). This is important for our context, as an alert by itself might also distract the user. Specifically, if a driver has already changed lanes, then a strong (audio) alert that indicates a lane change might confuse them. Future work should look into how different modalities of alerting affect lane change capabilities (e.g., cf. Spence & Ho, 2008). In general, reaction times to an audio signal are expected to be relatively short, as it is hard to overlook because audio signals are omni-present. Moreover, in contrast to a visual stimulus, auditory stimuli do not overlap with the visual requirements of the driving task and therefore interfere less with the driving task (cf. Wickens, 2002; 2008).

We found that there was no systematic effect of different levels of cognitive load on eventual lane change distance. This is due to the relatively small effect of initial reaction time. If a bigger sample of participants was used, or if the cognitive tasks were even more demanding, a stronger lasting effect of the interfering task might have been found. Future work can explore how further methods of varying load affect performance.

6.5.2 Conclusion

In-car warnings as a last resort to prevent a crash might be helpful. However, our work shows that such warnings should be given timely. Even in a simple lane change set-up, a visual warning at 500 m before the incident is not sufficient for all drivers. It is therefore important to further consider whether and how technology can be used to provide such warnings in an even more timely fashion.

References

Abdullaev, Y.G., Posner, M.I., 1998. Event-Related Brain Potential Imaging of Semantic Encoding during Processing Single Words. *NeuroImage* 7, 1–13. doi:10.1006/nimg.1997.0309

Birth, S., Pflaumbaum, M., Potzel, D., Sieber, G., 2009. *Human Factors Guideline for Safer Road Infrastructure*. Paris: World Road Association.

Brumby, D. P., Janssen, C. P., Kujala, T., & Salvucci, D. D. (2018). Computational models of user multitasking. In A. Oulasvirta, P. Kristensson, X. Bi, & A. Howes (Eds.), *Computational interaction design*. Oxford: Oxford University Press.

Cao, Y., Castronovo, S., Mahr, A. & Müller, C., 2009. On timing and modality choice with local danger warnings for drivers. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* 75-78. ACM.

CBS, 2013. *Mobiliteit per regio naar vervoerwijzen en algemene kenmerken, 1985-2007*. CBS StatLine. http://statline.cbs.nl/Statweb/search/?Q=verplaatsingsgedrag+in+Nederland

CROW, 2013. *Maatregelen op autosnelwegen 2013*, 514 ed. CROW. http://www.crow.nl/publicaties/maatregelen-op-autosnelwegen-__-werk-in-uitvoering.

Dingus, T.A., Guo, F., Lee, S., Antin, J.F., Perez, M., Buchanan-King, M., & Hankey, J., 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences* 113, 2636–2641. doi:10.1073/pnas.1513271113

Dubbeldam, R., 2007. *NOA: nieuwe ontwerprichtlijn autosnelwegen*. Rijkswaterstaat, [Adviesdienst Verkeer en Vervoer].

Finnegan, P. & Green, P., 1990. *Time to Change Lanes: A Literature Review.* Ann Arbor, MI: University of Michigan. Report No. IVHS-TR-90-13.

Fitch, G.M., Lee, S.E., Klauer, S., Hankey, J., & Sudweeks, J., 2009. *Analysis of lane-change crashes and near-crashes.* Springfield, VA: US Department of Transportation, National Highway Traffic Safety Administration. Report nr: DOT HS 811 147

Gozálvez, J., Sepulcre, M., & Bauza, R., 2012. IEEE 802.11p vehicle to infrastructure communications in urban environments. *IEEE Communications Magazine*, 50, 176–183. doi:10.1109/MCOM.2012.6194400

Hart, S.G., & Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Human Mental Workload*, 1, 139–183. doi:10.1016/S0166-4115(08)62386-9

Hetrick, S., 1997. Examination of driver lane change behavior and the potential effectiveness of warning onset rules for lane change or "side" crash avoidance systems. Doctoral dissertation, Virginia Polytechnic Institute & State University, 1–82.

Horrey, W.J., Wickens, C.D., 7 Consalus, K.P., 2006. Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied* 12, 67–78. doi:10.1037/1076-898X.12.2.67

IEEE, 2010. Ieee standard for information technology– local and metropolitan area networks– specific requirements– part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 6: Wireless access in vehicular environments. IEEE Std 802.11p-2010 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, IEEE Std 802.11n-2009, and IEEE Std 802.11w-2009), 1–51.

Iqbal, S.T., Ju, Y.-C., & Horvitz, E., 2010. Cars, calls, and cognition: Investigating driving and divided attention, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1281–1290. doi:10.1145/1753326.1753518

Janssen, C.P., & Brumby, D.P., 2010. Strategic Adaptation to Performance Objectives in a Dual-Task Setting. *Cognitive Science* 34, 1548–1560. doi:10.1111/j.1551-6709.2010.01124.x

Janssen, C.P., Brumby, D.P., & Garnett, R., 2012. Natural Break Points: The Influence of Priorities and Cognitive and Motor Cues on Dual-Task Interleaving. *Journal of Cognitive Engineering and Decision Making* 6, 5–29. doi:10.1177/1555343411432339

Janssen, C.P., Gould, S.J.J., Li, S.Y.W., Brumby, D.P., & Cox, A.L., 2015. Integrating knowledge of multitasking and interruptions across different perspectives and research methods. *Journal of Human Computer Studies* 79, 1–5. doi:10.1016/j.ijhcs.2015.03.002

Janssen, C.P., Iqbal, S.T., & Ju, Y.-C., 2014. Sharing a driver's context with a caller via continuous audio cues to increase awareness about driver state. *Journal of Experimental Psychology: Applied* 20, 270–284. doi:10.1037/xap0000020

Khattak, A.J., Khattak, A.J., & Council, F.M., 2002. Effects of work zone presence on injury and non-injury crashes. *Accident Analysis & Prevention* 34, 19–29. doi:10.1016/S0001-4575(00)00099-3

Klauer, S.G., Guo, F., Simons-Morton, B.G., Ouimet, M.C., Lee, S.E., & Dingus, T.A., 2014. Distracted Driving and Risk of Road Crashes among Novice and Experienced Drivers. *New England Journal of Medicine* 370, 54–59. doi:10.1056/NEJMsa1204142

Kroon, E., Brookhuis, K., Hagenzieker, M., & Martens, M., 2014. *Human factor guidelines for the design of safe in-car traffic information services*. Den Haag, The Netherlands: Ministerie van Infrastructuur en Milieu (Ministry of Infrastructure and Environment), 2014, 33 p., 47

Kunar, M.A., Carter, R., Cohen, M., & Horowitz, T.S., 2008. Telephone conversation impairs sustained visual attention via a central bottleneck. *Psychonomic Bulletin & Review* 15, 1135–1140. doi:10.3758/PBR.15.6.1135

Math, R., Mahr, A., Moniri, M.M., & Müller, C., 2013. OpenDS: A new open-source driving simulator for research. *GMM-Fachbericht-AmE 2013*.

Paier, A., Faetani, D., & Mecklenbrauker, C. F., 2010. Performance evaluation of ieee 802.11 p physical layer infrastructure-to-vehicle real-world measurements. In *Applied Sciences in Biomedical and Communication Technologies (ISABEL)*, pages 1–5. IEEE.

Spence, C. & Ho, C. (2008). Multisensory warning signals for event perception and safe driving. *Theoretical Issues in Ergonomics Science*, 9, 523–554.

Robinson, G.H., Erickson, D.J., Thurston, G.L., & Clark, R.L., 1972. Visual Search by Automobile Drivers. *Human Factors*, 14, 315–323. doi:10.1177/001872087201400404

Salvucci, D.D., 2001. Predicting the effects of in-car interface use on driver performance: an integrated model approach. *International Journal of Human-Computer Studies* 55, 85–107. doi:10.1006/ijhc.2001.0472

Salvucci, D.D., Mandalia, H.M., Kuge, N., & Yamamura, T., 2007. Lane-Change Detection Using a Computational Driver Model. *Human Factors* 49, 532–542. doi:10.1518/001872007X200157

Snyder, A.Z., Abdullaev, Y.G., Posner, M.I., & Raichle, M.E., 1995. Scalp electrical potentials reflect regional cerebral blood flow responses during processing of written words. *Proceedings of the National Academy of Sciences* 92, 1689–1693. doi:10.1073/pnas.92.5.1689

Strayer, D.L. & Johnston, W.A., 2001. Driven to distraction: Dual-Task Studies of Simulated Driving and Conversing on a Cellular Telephone. *Psychological Science* 12, 462-466.

Van Gent, A.L., 2007. *Verkeersonveiligheid bij werk in uitvoering: een literatuurstudie*. Den Haag, The Netherlands: SWOV. Report nr: R-2007-5

Van der Heiden, R. M., Iqbal, S. T., & Janssen, C. P. (2017). Priming drivers before handover in semi-autonomous cars. In P*roceedings of the SIGCHI conference on human factors in computing systems* (pp. 392–404). New York, NY, USA: ACM. http://doi.acm.org/10.1145/3025453.3025507

Wakasugi, T., 2005. A Study on Warning Timing for Lane Change Decision Aid Systems Based on Driver's Lane Change Maneuver. In *Proceedings of the 19th international technical conference on the Enhanced Safety of Vehicles*. Paper number 05-0290*,* 1–7.

Wickens, C.D., 2008. Multiple Resources and Mental Workload. *Human Factors*, 50, 449–455. doi:10.1518/001872008X288394

Wickens, C.D., 2002. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science* 3, 159–177. doi:10.1080/14639220210123806

Chapter 7

General discussion

This thesis investigated susceptibility of the human brain to auditory signals under different (driving) conditions **(Part I: Chapters 2-4)**, and the ability of humans to act on alerts in specific contexts **(Part II: Chapters 5,6)**. Table 7.1 shows our original research question for each chapter and the main findings.

In Part I, three experiments were conducted to answer the main question: "How susceptible are humans to auditory stimuli?" To answer this question, we measured electroencephalogram (EEG) Event-Related Potential (ERP) responses in an oddball paradigm in (automated) driving tasks and cognitive load inducing tasks. The main finding was that automated driving (Chapter 2, 4, see also Van der Heiden et al., 2019) and cognitive load inducing tasks (Chapter 3, 4) can both reduce susceptibility to auditory alerts.

Given this observation that auditory susceptibility is sometimes reduced and, more generally, to explore ways to support accurate human behavior while driving, in Part II, we explored interventions that might alleviate the associated potential problems. In the first experiment (Chapter 5; Van der Heiden, Iqbal, Janssen, 2017), we introduced early warnings, or pre-alerts, to allow drivers to anticipate a later acute imperative alert for an upcoming transition of control from the car to the human driver. The results showed that pre-alerts can support safer transitions of control. In the second experiment (Chapter 6; Van der Heiden, Janssen, Donker, & Merkx, 2019), we tested an intervention where visual in-car warnings were presented to indicate an upcoming, necessary lane change. Results showed that, on average, drivers were able to change lanes in time. However, at the same time, there are occasionally some drivers that did not respond in a timely manner.

In the remainder of this chapter, the general findings and theoretical implications of Part I and Part II will be discussed, followed by a discussion of practical implications.

	Chapter	Main research question	Answer & main findings
Part I Measuring auditory susceptibility	2	Does automated driving reduce auditory susceptibility?	**Yes.** Susceptibility is reduced during automated driving, but not as strong as during manual driving. A response requirement to an unrelated stimulus can increase susceptibility.
	3	Does cognitive load reduce auditory susceptibility?	**Yes.** Cognitive load inducing tasks, even without visual/manual components, reduce auditory susceptibility.
	4	Does cognitive load during automated driving reduce auditory susceptibility?	**Yes.** Auditory susceptibility is reduced by concurrent cognitive load and automated driving.
Part II Testing interventions	5	Can pre-alerts improve the transfer of control in semi-automated vehicles?	**Yes.** Pre-alerts make drivers gaze at the road earlier and allow for an earlier initiation of speed reduction to cope with challenging traffic demands.
	6	Do last-minute visual in-car warnings accommodate a timely lane change?	**No.** On average, drivers are able to change within safe margins. However, outliers and computer simulations show that these signals might not always be sufficient on their own.

Table 7.1. Main research questions and findings

7.1 Measuring susceptibility (Part I): Comparison of levels of auditory susceptibility across chapters

Susceptibility to auditory alerts was investigated in three separate experiments (Chapters 2-4). Some conditions were relatively comparable (i.e., all studies have a baseline control, some have automated driving conditions, some cognitive tasks). Therefore, we can now reflect on the common divider and the differences between the results across chapters. In all three chapters EEG was used to measure auditory susceptibility. In Chapter 2 (Van der Heiden et al., 2018) we studied how auditory susceptibility is influenced by manual driving and automated driving. Following the design of Wester, Böcker, Volkerts, Verster, & Kenemans (2008), we also manipulated whether participants had a response requirement to a deviant tone ("active") or no response requirement ("passive"). Chapter 3 studies the effect of cognitive load (verb generation task) on auditory susceptibility, as well as how this is influenced by the inter-stimulus interval between a stimulus of the cognitive task and the probe of the oddball task (0, 200, or 400 ms). Chapter 4 studies the combined effect of automated driving and concurrent cognitive load. Figure 7.1 and 7.2 show the results of Part I (Chapters 2-4) side by side. Figure 7.1 presents all fP3 levels over the three different studies. The barplot in Figure 7.2 shows one bar for each condition from each study.

As is clear from Figure 7.2, the different studies show a consistent pattern driven by the induced cognitive load. That is, auditory susceptibility, as operationalized by fP3 amplitude, is consistently the highest under baseline conditions without any other task. The trend in the data is that under automated driving conditions, there is a slight reduction in fP3 amplitude. Although this only reached significance in Chapter 2, the trend is the same in Chapter 4. Moreover, the mean value of the fP3 response under automated driving in Chapter 4 (M = 9.9 μV) is comparable to that observed for automated driving in the active condition in Chapter 2 (M = 9.5 μV). The strongest reduction in fP3 amplitude is observed during cognitive load inducing tasks (e.g., verb generation tasks in Chapters 3 and 4). Across the various versions that were used in Chapter 3 (manipulation of timing of stimulus) and Chapter 4 (repeat or generate version of the verb task), a mean value of around 5 to 6 μV is observed. This is comparable to the values observed for manual driving under passive conditions in Chapter 2. In other words: cognitive tasks can reduce the susceptibility to auditory alerts as strongly as visual-manual tasks such as driving.

Admittedly such comparisons across experiments should be carried out cautiously, because different experiments involve different samples (although from largely the same population), were conducted in different time period, and with partly different equipment.

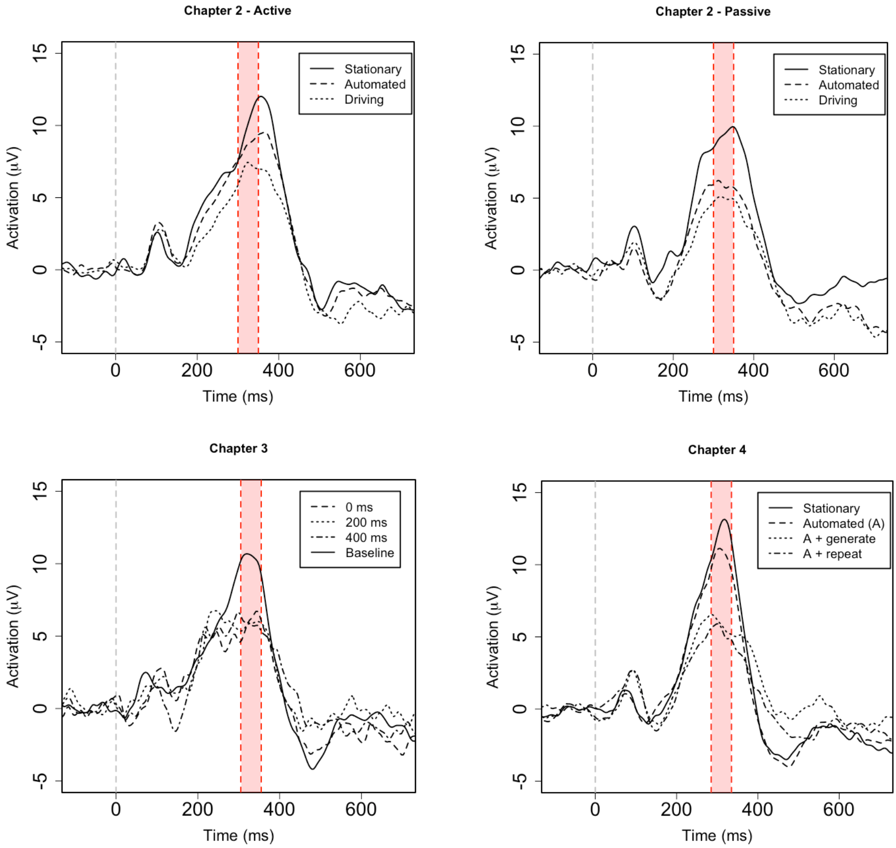


Figure 7.1. fP3 results from Part I (Chapters 2 - 4) of the thesis. Each plot shows the average fP3 difference wave from electrode FCz for each corresponding condition. The gray line at time point 0 indicates the onset of the oddball probe. The fP3 peak location, which slightly differs per study, is indicated with a ribbon. The data within this 50 ms epoch are used for statistical analysis within the study. The average values per condition within the ribbon are also shown as barplot in Figure 7.2. For details on the different conditions, see the corresponding chapters.

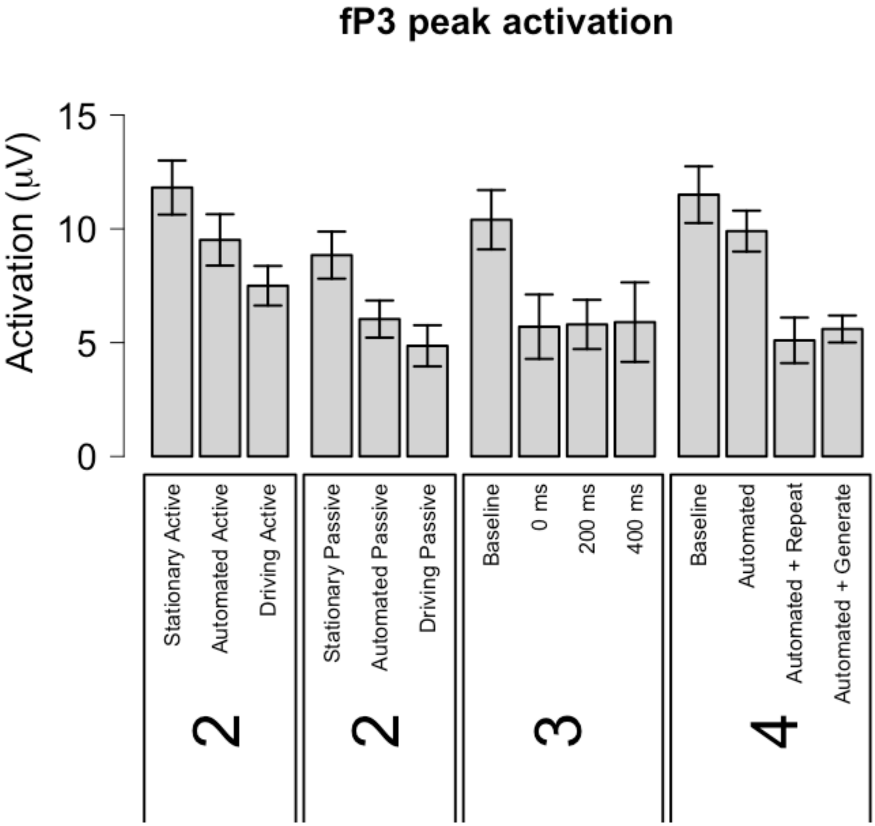


Figure 7.2. fP3 peak activation for each condition in each of the three chapters (2-4) from Part I. Error bars show standard error of the mean. Chapter 2 is represented in two sets of bars. The first set of three bars represents the results for the active response requirement group (i.e., pressing a button after deviant sound presentation). The second set represents the results for the passive group, that did not have a response requirement. The third set of bars represent the results from Chapter 3, where apart from baseline, a verb generation task is presented either 0, 200 or 400 ms prior oddball probe onset. The fourth set of bars represent the results from Chapter 4. Apart from single task baseline (stationary) and automated, automated driving is separately combined with a noun repetition task and a verb generation task. For details on the different conditions, see the corresponding chapters. Clearly noticeable is the consistent pattern where baseline shows highest fP3 activation and fP3 decreases when more cognitive load is added.

7.2 Measuring susceptibility (Part I): Implications for theory

The pattern in Figure 7.2 is consistent with observations in the literature, but also expands it. Specifically, in Chapter 2 we replicate the finding that fP3 response is reduced under visual-manual conditions (e.g., Allison & Polich, 2008; Massar, Wester, Volkerts, & Kenemans, 2010; Miller, Rietschel, McDonald, & Hatfield, 2011; Scheer, Bülthoff, & Chuang, 2016, 2018; Ullsperger, Freude, & Erdmann, 2001; Wester, Böcker, Volkerts, Verster, & Kenemans, 2008; Wester, 2009). We expand the existing knowledge as reported in the literature by demonstrating that fP3 is also reduced under automated driving conditions.

Our novel finding in Chapters 3 and 4 is that the fP3 is reduced during cognitive load, without visual or manual components. This is consistent with previous literature that consistently has found a reduced fP3 during cognitive load (e.g., Allison & Polich, 2008; Miller et al., 2011; Ullsperger et al., 2001). However, those previous studies all used cognitive load inducing tasks that involved manual interaction or visual stimuli that went on continuously while fP3 was elicited by probes. Chapter 3 used a cognitive load inducing task to measure the effects of the thought process of generating a word. fP3 is reduced during this process compared to baseline. This shows that fP3 is reduced by cognitive load in the absence of manual interaction or visual stimuli. This effect was replicated in Chapter 4.

In other words, the combined results from Chapters 2-4 suggests that cognitive load might be an important factor that influences fP3 response. This hypothesis has been suggested before in overview articles of fP3 studies (e.g., Polich, 2007), but required further evidence. The hypothesis that cognitive load reduces fP3 amplitude might also explain why there is a reduction of the fP3 under automated driving conditions without additional task (Chapter 2, and condition "automated" in Chapter 4): Although no manual action is required of the participant under automated driving conditions, the visual stimulation seems to induce some cognitive load. Future work can explore further what this process might entail exactly. Whatever this will yield, the present results are consistent with perceptual-load theory (Murphy, Spence, & Dalton, 2017): Processing of task-irrelevant stimuli, however salient (e.g., novels), is reduced with increasing task load, at least in conditions of simulated (automated) driving and additional tasks (generating or repeating words). This also holds in on-the-road driving conditions (Wester, 2009), but it remains to be seen how this works out when such unexpected and/ or salient stimuli inherently convey information relevant to driving (e.g., sirens, honking, brake lights in front of you). One clue in this respect is the fact that stimuli presented outside the focus of attention, but nevertheless relevant to performance, elicit an fP3-like response that is actually stronger than that elicited by the same stimulus within the focus of attention (Logemann, Böcker, Deschamps, Kemner, & Kenemans 2014).

In a similar vein, future work should explore what the effect is of different levels of cognitive load. Behavioral studies (including our own versions in Chapter 4 and 6) have consistently shown that the generate condition takes longer to respond to than the repeat condition (Iqbal, Ju, & Horvitz, 2010; Kunar, Carter, Cohen, & Horowitz, 2008; Strayer & Johnston, 2001). Moreover, the generate task also creates more dual-task interference (Current Chapter 6; Iqbal et al., 2010; Kunar et al., 2008; Strayer & Johnston, 2001) and produces more frontal brain activity (Abdullaev & Posner, 1998; Bijl, De Bruin, Böcker, Kenemans, & Verbaten, 2007). This is commonly interpreted as that generating a verb creates more cognitive load compared to repeating a word. Despite this difference in load level, we did not find an effect on fP3 response.

As discussed in Chapter 4, there are at least two scenarios to explain this lack of effect. First, the lack of differential fP3 could reflect equal cognitive load in repeat and generate, but induced by response production in the former and by semantic search (preceding response production) in the latter. Second, as visible in Figure 4.5, the verbal response in the repeat condition overlaps, and therefore might mask the probe stimulus, which in effect this might result in a similar fP3 reduction. As also discussed in Chapter 4, one way to elucidate this in future research is to use delayed intervals between word stimulus and probe, also in the repeat condition.

Given the reduction in susceptibility as indexed by fP3, can we also determine potential correlates at the behavioral or subjective level? One clue comes from the active condition in Chapter 2, in which occasional deviant tones had to be detected (and detection was indicated by button press). Although both the mean time needed to press the button and the mean within-subject variability of this reaction time (RT) were increased under driving relative to autonomous and stationary conditions, these differences were not statistically significant. This can be considered as being consistent with the pattern of RT results reported earlier: As for driving relative to stationary, Wester et al. (2008) found greater RT variability but no difference for mean RT; Wester (2009) found the reverse pattern; and Wester, Verster, Volkerts, Böcker, & Kenemans (2010) found no difference for either variable. Note that the mere introduction of the active condition enhanced fP3 in both driving and stationary conditions (present Chapter 2), or specifically in the driving condition (Wester et al., 2008; Wester, 2009). Future studies are needed to explore his further, for example, by examining the correlation between driving-induced modulations of behavioral deviant detection and driving-induced reduction of fP3.

In addition, future work can explore the link between auditory susceptibility levels and other behavioral metrics of performance. This pertains to, for example, initial reaction time and brake response time (as in Chapter 5), and lane change time or distance (as in Chapter 6). Apart from variability and mean response time, studies can also look at outliers (see current Chapter 6), or missed responses and false alarms (in addition to RT and RT variability). Similarly, future work can explore how susceptibility relates to more subjective experience, such as experienced workload (Hart & Staveland, 1988) and trust (e.g., Mühl et al., 2019), and how such factors in turn influence user's use of automation (Parasuraman & Riley, 1997). Another interesting venue concerns assessing the brain's global activation level, as reflected in the power of EEG alpha waves. Alpha power has been shown to increase with prolonged driving (indicating decreasing activation), even when subjective activation increases (Schmidt et al., 2007). Alpha power has also been shown to predict failed detection of infrequent external signals up to about 15 seconds before the external signal is actually presented (O'Connell et al., 2009). An open question is therefore how alpha levels change under automated driving conditions.

7.3 Testing interventions (Part II): Implications for theory

In Part II, we explored behavioral metrics other than those examined in Part I. Chapter 5 (Van der Heiden et al., 2017) showed that drivers of semi-automated vehicles have better transitions of control, if they are forewarned of the transition by a so-called pre-alert. In trials where a pre-alert is present, human drivers tend to look at the road earlier and reduce their speed earlier and more gradually. Human drivers also tend to look at the road more regularly during conditions where a pre-alert is absent, which possibly impacts their performance on non-driving related tasks that they are performing. This shows that pre-alerts can contribute to safety when control is transferred from the system to the driver. Another recent independent study replicated the positive effects of pre-alerts, and additionally showed that pre-alerts worked better when they additionally cued the appropriate action (steering vs. braking) on part of the driver (Borojeni, Weber, Heuten, & Boll, 2018) Specifically, they found that priming drivers for their required action results in drivers acting faster and safer.

The notion of pre-alerts builds on the theoretical concept that mediated, negotiated, or scheduled interruptions are better for human performance compared to sudden interruptions (McFarlane, 2002; McFarlane & Latorella, 2002). These ideas have been widely explored in desktop based (office) settings (e.g., Bailey & Iqbal, 2008; Mark, 2015). Our work demonstrates the value of such theories to dynamic, safety-critical environments such as driving. While our analysis focused on the processes preceding, during, and immediately following the pre-alert, future work can explore more widely how pre-alerts affect transitions of control from the vehicle to the human and from the human back to the car (see also Janssen, Iqbal, Kun, & Donker, 2019).

Chapter 6 (Van der Heiden et al., 2019), measured human reaction time to visual alerts. Our set-up was a relatively controlled lab environment in which participants had to change lanes based on a visual (in-car) signal. Participants on average changed lanes swiftly, and in time given current technological interventions (i.e., beacons along the road that warn 500 m before a required lane change). However, on some trials, participants responded too late. These outliers are important to consider in traffic safety, as the percentage of errors that might lead to a crash needs to be minimized.

To further explore the potential impact under real traffic situations, we turned to the use of computer simulations. The motivation was that some tasks are difficult to implement in a simulator (Kemeny & Panerai, 2003). As visual cues for depth are reduced in a driving simulator, a task such as mirror checking might be invalid in that context. To accommodate this limitation, we resorted to published data (Robinson et al., 1972) on speed of lane-change initiation during on-the-road driving. The lane change operations included mirror checking and sideways head movements. These data were combined using computer simulation models with our driving-simulator data, to enrich the latter. The eventual results show that responses to in-car alerts may be even more delayed than already observed directly in the lab, perhaps even compromising the performance of 50% of the drivers. More generally, this illustrates that using computer simulation models to combine lab- with real-world behavioral data can be used to predict human behavior under a wider set of conditions than those observed in the lab, including situations that might be dangerous for participants and therefore impossible to study in the lab (e.g., Brumby, Janssen, Kujala, & Salvucci, 2018; Horrey, Wickens, & Consalus, 2006; Janssen & Brumby, 2010; Janssen, Brumby & Garnett, 2012; Salvucci, 2001; Salvucci, Mandalia, Kuge, & Yamamura, 2007).

7.4 Implications for practice

During automated driving task load is often assumed to be reduced compared to manual driving. Therefore, one would perhaps suspect to observe high levels of auditory susceptibility under automated driving. However, as shown in Chapter 2, even during automated driving auditory susceptibility may be reduced compared to when the simulated car is stationary. This suggests that automated driving should still be considered as a potential source of cognitive load. The implication for practice, including the design and evaluation of cars and in-car equipment, is that designs should not assume a human driver's full attention to the environment under automated driving conditions, although their engagement in driving related tasks is reduced.

Considering automated driving as a source of cognitive load is in line with the notion that automation can radically change human behavior and, in irony, lead to worse performance (Bainbridge, 1983), and not only because humans engage in non-driving related tasks, as they did not do that in chapter 2 but nevertheless susceptibility was reduced relative to a stationary condition. That said, further studies are needed to test if and how strongly a reduction in auditory susceptibility can impact for example situational awareness (Endsley & Garland, 2000) or behavioral responses (i.e., in a similar fashion as we did in Chapters 5 and 6).

If reduced human susceptibility (Chapters 2-4) is indeed associated with negative consequences for behavior (e.g., reduced situational awareness, slower response time), then safety-critical systems should not rely on a single, uni-modal alert. Manufacturers of cars and in-car devices should design their systems in a way that can accommodate (although not necessarily actively facilitate) the distracted driver who might possibly miss an alert. In Chapter 2 we showed that introducing a secondary auditory task may in fact enhance susceptibility. In a way, this suggests that multi-tasking may be a solution. However, this goes only so far: it may hold for listening and manual responses, but not for talking or vocally responding otherwise (Chapters 3, 4, 6). Part I of this thesis further shows that auditory susceptibility is in general reduced with cognitive load, in particular when the cognitive load is induced by a task that strongly resembles a hands-free telephone conversation, which may be considered worrisome. However, the link between auditory susceptibility and response to warning signals is not clear and therefore needs more research. An example of the effect of early warnings on behavioral results is shown in Chapter 5 (Van der Heiden et al., 2017) where the use of a pre-alert allows for earlier gaze to the road and more smooth speed reduction after transfer of control, hence improved safety.

The literature on driving and driver distraction largely focuses on average responses and differences in responses between conditions (e.g., Burns, Parkes, Burton, Smith, & Burch, 2002; Hancock, Simmons, Hashemi, Howarth, & Ranney, 1999; Horberry, Anderson, Regan, Triggs, & Brown, 2006; Strayer & Drews, 2007), for example, driving with distraction versus driving without distraction. However, Chapter 6 shows empirically that in situations where the average participant responds in-time, the complete distribution of all trials from all drivers contains a wider set of response times, of which some might be too slow. Interventions should, therefore, consider such extreme behavior. This is reiterated in Chapter 5, where our intervention that provides an early warning can accommodate people that occasionally miss an alert, or that need more time to respond (see Chapter 6).

We propose early warnings (pre-alerts, Chapter 5) as a potential solution where the human driver that acts as a backup during automated driving can engage in other tasks without the need for immediate response after a sudden intervention request. Pre-alerts enable the human driver to become aware of the upcoming request, which allows orienting ahead of time by gazing at the road. This effect is consistent with more theoretical notions of interruption handling (McFarlane, 2002; McFarlane & Latorella, 2002), see also (Janssen, Iqbal, Kun, & Donker, 2019).

The combined work in this thesis also has implications for road authorities and lawmakers. We found that human susceptibility is reduced under automated driving (Chapter 2, 4), and that responses to environment-triggered alerts might not always be in time (Chapter 6). In particular, some of these negative effects seem to be bigger under distracted conditions (Chapter 3, 4, 5), and can also occur under distraction that is not visual. The implication is therefore that, although many technological innovations can bring society progress, road authorities and lawmakers should be critical. Consistent with Bainbridge's notion of the "irony of automation" (1983), automation interventions that bring us progress might change human behavior in unanticipated ways, and have occasional unexpected negative consequences. Careful, principled study of human behavior for theoretical (Part I) and practical (Part II) scenarios allows us to be prepared for the future, and better predicting human cognitive processing and human behavior.

References

Allison, B. Z., & Polich, J. (2008). Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biological psychology*, 77(3), 277-283.

Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 14(4), 1-28.

Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man–machine systems* (pp. 129-135). Pergamon.

Bijl, S., De Bruin, E. A., Böcker, K. E., Kenemans, J. L., & Verbaten, M. N. (2007). Effects of chronic drinking on verb generation: an event related potential study. *Human Psychopharmacology: Clinical and Experimental*, 22(3), 157-166. https://doi.org/10.1002/hup.835

Boothroyd, G. (2005). Assembly automation and product design. CRC Press.

Borojeni, S. S., Weber, L., Heuten, W., & Boll, S. (2018). From reading to driving: priming mobile users for take-over situations in highly automated driving. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1-12).

Brumby, D. P., Janssen, C. P., Kujala, T., & Salvucci, D. D. (2018). Computational models of user multitasking. In A. Oulasvirta, P. Kristensson, X. Bi, & A. Howes (Eds.), *Computational interaction design*. Oxford: Oxford University Press.

Burns, P. C., Parkes, A., Burton, S., Smith, R. K., & Burch, D. (2002). *How Dangerous is Driving with a Mobile Phone?: Benchmarking the Impairment to Alcohol* (Vol. 547). TRL.

Endsley, M. R., & Garland, D. J. (2000). Theoretical underpinnings of situation awareness: A critical review. *Situation awareness analysis and measurement*, *1*, 24.

Hancock, P. A., Simmons, L., Hashemi, L., Howarth, H., & Ranney, T. (1999). The effects of in-vehicle distraction on driver response during a crucial driving maneuver. *Transportation Human Factors*, *1*(4), 295-309.

Hart, S.G., & Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Human Mental Workload*, 1, 139–183. doi:10.1016/S0166-4115(08)62386-9

Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention*, *38*(1), 185-191.

Horrey, W.J., Wickens, C.D., Consalus, K.P., 2006. Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied* 12, 67–78. doi:10.1037/1076-898X.12.2.67

Iqbal, S.T., Ju, Y.-C., & Horvitz, E., 2010. Cars, calls, and cognition: Investigating driving and divided attention, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1281–1290. doi:10.1145/1753326.1753518

Janssen, C.P., & Brumby, D.P., 2010. Strategic Adaptation to Performance Objectives in a Dual-Task Setting. *Cognitive Science* 34, 1548–1560. doi:10.1111/j.1551-6709.2010.01124.x

Janssen, C.P., Brumby, D.P., & Garnett, R., 2012. Natural Break Points: The Influence of Priorities and Cognitive and Motor Cues on Dual-Task Interleaving. *Journal of Cognitive Engineering and Decision Making* 6, 5–29. doi:10.1177/1555343411432339

Janssen, C. P., Iqbal, S. T., Kun, A. L., & Donker, S. F. (2019). Interrupted by my car? Implications of interruption and interleaving research for automated vehicles. *International Journal of Human-Computer Studies*, *130*, 221-233.

Kemeny, A., & Panerai, F. (2003). Evaluating perception in driving simulation experiments. *Trends in cognitive sciences*, 7(1), 31-37.

Kunar, M.A., Carter, R., Cohen, M., & Horowitz, T.S., 2008. Telephone conversation impairs sustained visual attention via a central bottleneck. *Psychonomic Bulletin & Review* 15, 1135–1140. doi:10.3758/PBR.15.6.1135

Logemann, H. A., Böcker, K. B. E., Deschamps, P. K. H., Kemner, C., & Kenemans, J. L. (2014). The effect of the augmentation of cholinergic neurotransmission by nicotine on EEG indices of visuospatial attention. *Behavioural brain research*, *260*, 67-73.

Mark, G. (2015). Multitasking in the digital age. *Synthesis Lectures On Human-Centered Informatics*, 8(3), 1-113.

Massar, S. A., Wester, A. E., Volkerts, E. R., & Kenemans, J. L. (2010). Manipulation specific effects of mental fatigue: evidence from novelty processing and simulated driving. *Psychophysiology*, 47(6), 1119-1126.

McFarlane, D. C. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction*, *17*(1), 63-139.

McFarlane, D. C., & Latorella, K. A. (2002). The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction*, *17*(1), 1-61.

Miller, M. W., Rietschel, J. C., McDonald, C. G., & Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. *International Journal of Psychophysiology*, 80(1), 75-78.

Mühl, K., Strauch, C., Grabmaier, C., Reithinger, S., Huckauf, A., & Baumann, M. (2019). Get Ready for Being Chauffeured: Passenger’s Preferences and Trust While Being Driven by Human and Automation. *Human factors*, 0018720819872893.

Murphy, S., Spence, C., & Dalton, P. (2017). Auditory perceptual load: A review. *Hearing Research*, 352, 40-48.

O'Connell, R. G., Dockree, P. M., Robertson, I. H., Bellgrove, M. A., Foxe, J. J., & Kelly, S. P. (2009). Uncovering the neural signature of lapsing attention: electrophysiological signals predict errors up to 20 s before they occur. *Journal of Neuroscience*, 29(26), 8604-8611.

Parasuraman, R., & Riley, V. (1997) Humans and automation: use, misuse, disuse, abuse *Human Factors*, 39 (2). 230-253

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148. https://doi.org/10.1016/j.clinph.2007.04.019

Robinson, G.H., Erickson, D.J., Thurston, G.L., & Clark, R.L., 1972. Visual Search by Automobile Drivers. *Human Factors*, 14, 315–323. doi:10.1177/001872087201400404

Salvucci, D.D., 2001. Predicting the effects of in-car interface use on driver performance: an integrated model approach. *International Journal of Human-Computer Studies* 55, 85–107. doi:10.1006/ijhc.2001.0472

Salvucci, D.D., Mandalia, H.M., Kuge, N., & Yamamura, T., 2007. Lane-Change Detection Using a Computational Driver Model. *Human Factors* 49, 532–542. doi:10.1518/001872007X200157

Scheer, M., Bülthoff, H. H., & Chuang, L. L. (2016). Steering demands diminish the early-P3, late-P3 and RON components of the event-related potential of task-irrelevant environmental sounds. *Frontiers in human neuroscience*, 10, 73.

Scheer, M., Bülthoff, H. H., & Chuang, L. L. (2018). Auditory task irrelevance: a basis for inattentional deafness. *Human factors*, 60(3), 428-440.

Schmidt, E. A., Kincses, W. E., Scharuf, M., Haufe, S., Schubert, R., & Curio, G. (2007). *Assessing drivers’ vigilance state during monotonous driving.*

Strayer, D. L. & Drews, F. A. (2007). Cell-phone–induced driver distraction. *Current Directions in Psychological Science*, *16*(3), 128-131.

Strayer, D. L. & Johnston, W.A., 2001. Driven to distraction: Dual-Task Studies of Simulated Driving and Conversing on a Cellular Telephone. *Psychological Science* 12, 462-466.

Van der Heiden, R. M., Iqbal, S. T., & Janssen, C. P. (2017). Priming drivers before handover in semi-autonomous cars. *In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 392-404). ACM.

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., Hardeman, L. E., Mans, K., & Kenemans, J. L. (2018). Susceptibility to audio signals during autonomous driving. *PloS one*, 13(8), e0201963. https://doi.org/10.1371/journal.pone.0201963

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., & Merkx, C. L. (2019). Visual in-car warnings: How fast do drivers respond?. *Transportation research part F: traffic psychology and behaviour*, *65*, 748-759.

Ullsperger, P., Freude, G., & Erdmann, U. (2001). Auditory probe sensitivity to mental workload changes–an event-related potential study. *International Journal of Psychophysiology*, 40(3), 201-209.

Wester A.E. Attention Deficit and Impulsivity: Driving, Drugs and Electrophysiology. 2009. Utrecht: Utrecht University. PhD thesis.

Wester, A. E., Böcker, K. B. E., Volkerts, E. R., Verster, J. C., & Kenemans, J. L. (2008). Event-related potentials and secondary task performance during simulated driving. *Accident Analysis & Prevention*, 40(1), 1-7. https://doi.org/10.1016/j.aap.2007.02.014

Wester, A. E., Verster, J. C., Volkerts, E. R., Böcker, K. B., & Kenemans, J. L. (2010). Effects of alcohol on attention orienting and dual-task performance during simulated driving: An event-related potential study. *Journal of psychopharmacology*, *24*(9), 1333-1348.

Appendix I

Nederlandse samenvatting

Automatisering neemt in de huidige samenleving veel taken over van de mens. Niet alleen in een professionele omgeving, zoals in de besturing van vliegtuigen en productieprocessen in fabrieken, maar ook voor consumenten neemt de automatisering van vroegere handmatige taken snel toe, bijvoorbeeld in de vorm van robots voor huishoudelijke taken. Dit is ook zichtbaar in auto's, waar al jaren steeds meer processen geautomatiseerd worden en waarbij in de laatste jaren ook steeds meer kritische processen met betrekking tot veiligheid (deels) worden geautomatiseerd. Voorbeelden hiervan zijn systemen voor het automatisch afstand houden tot de voorligger, voor het voorkomen van een botsing en voor vermoeidheidswaarschuwingen. In de toekomst worden deze en andere systemen samengevoegd waardoor auto's grotendeels zelfstandig (autonoom) zullen kunnen rijden. Onderzoek laat zien dat er bij dit soort automatisering niet alleen taken voor de machine zijn, maar dat er juist samenwerking tussen mens en machine nodig is. Voor een goede samenwerking tussen mens en machine is het daarom belangrijk om te weten wat de capaciteiten van de mens zijn tijdens de omgang met geautomatiseerde systemen.

Wanneer een auto automatisch bestuurd wordt, is het van belang dat kan worden gewaarschuwd wanneer overname van de besturing of een andere actie door de mens nodig is. Hiervoor worden vaak auditieve waarschuwingen gebruikt. Het is echter onbekend wat de invloed is van rijden en autonoom rijden op de ontvankelijkheid voor auditieve informatie. Dit wordt in dit proefschrift onderzocht. Om de verkeersveiligheid te kunnen waarborgen is het daarnaast belangrijk om te weten wat het vermogen van de mens is om adequaat te reageren op zowel auditief als visueel aangeboden informatie.

Het proefschrift is opgedeeld in twee delen. Het eerste deel beschrijft een aantal studies die de ontvankelijkheid van het menselijk brein voor auditieve informatie in verschillende (rij-) omstandigheden en niveaus van afleiding meet. Het tweede deel beschrijft een tweetal studies dat het vermogen van mensen om waarschuwingen op te volgen onderzoekt.

In het eerste deel (hoofdstukken 2 - 4) behandel ik de volgende vraag: Hoe gevoelig is het menselijk brein voor auditieve informatie? Om deze vraag te beantwoorden maak ik gebruik van een rijsimulator en een meting van breinactiviteit (elektro-encefalografie, of EEG). Tijdens (autonoom) rijden en taken die zorgen voor mentale belasting wordt een zogenaamd oddball paradigma (een experimentele taak) gebruikt. In dit paradigma wordt de reactie van het brein op auditieve informatie gemeten. Een frequent afgespeelde pieptoon wordt af en toe afgewisseld met verschillende omgevingsgeluiden, bijvoorbeeld een blaffende hond. Als maat voor ontvankelijkheid voor nieuwe auditieve informatie wordt het verschil tussen de hersenactiviteit die wordt veroorzaakt door de verschillende omgevingsgeluiden en de hersenactiviteit die wordt veroorzaakt door de pieptoon gebruikt. De belangrijkste uitkomst van deze metingen is dat zowel autonoom rijden als mentaal belastende taken de ontvankelijkheid voor nieuwe auditieve informatie verminderen.

De lagere ontvankelijkheid vraagt om (een betere) ondersteuning van de bestuurder. In deel 2 (hoofdstukken 5 en 6) wordt in twee experimenten gekeken naar mogelijke manieren om de (afgeleide) bestuurder te ondersteunen. In het eerste experiment (hoofdstuk 5) onderzoek ik de effectiviteit van een vroege voorwaarschuwing voor de overdracht van besturing van machine naar de mens. Onze voorspelling was dat een voorwaarschuwing (in onze studie: 20 seconden voor een kritieke gebeurtenis op de weg) kan bijdragen aan een veiligere transitie van de besturing tussen machine en mens. Dit geeft meer tijd om andere niet-rijtaken tijdig af te ronden en om op het verkeer te oriënteren voordat het besturen van de auto overgenomen moet worden. Het onderzoek laat zien dat dergelijke voorwaarschuwingen inderdaad leiden tot veiliger gedrag: mensen kijken eerder naar de weg en reageren beter op kritieke situaties door bijvoorbeeld rustig te remmen in plaats van plotseling. In het tweede experiment (hoofdstuk 6) onderzoek ik hoeveel tijd er nodig is om van rijstrook te wisselen na een visuele waarschuwing in de auto voor wegafsluiting. Over het algemeen maken bestuurders een tijdige, veilige wissel van rijstrook. Tegelijkertijd zijn er verschillende proefpersonen die af en toe niet op tijd wisselen en daarmee een gevaarlijke situatie zouden kunnen veroorzaken. Om de verkeersveiligheid te verbeteren is het van belang om het aantal optredende gevaarlijke situaties te minimaliseren. Wanneer dergelijke situaties al bij een studie zoals deze met een rijsimulator gemeten worden, is de kans groot dat vergelijkbare omstandigheden op de weg ongelukken voorkomen.

Resumerend: mijn proefschrift laat zien dat de ontvankelijkheid voor nieuwe auditieve informatie verminderd is tijdens zowel rijden als autonoom gereden worden. Daarnaast zorgt ook mentale belasting voor een afname van de ontvankelijkheid. Het tijdig bewust maken van de bestuurder wanneer overname van de besturing nodig is (door middel van een voorwaarschuwing) zorgt over het algemeen voor een veiligere overdracht van de besturing van machine naar de mens. Deze kennis is relevant voor de ontwikkeling van autonome auto’s, omdat het aangeeft wat we kunnen verwachten van de capaciteiten van de menselijke bestuurder bij het gebruik van autonome auto’s.

Appendix II

Acknowledgments

In de vier jaar dat ik aan dit proefschrift heb gewerkt heb ik veel mensen leren kennen, zowel binnen de wetenschappelijke wereld als daarbuiten. Ik ben enorm dankbaar voor al de interacties die ik met iedereen heb gehad. Zonder jullie had dit proefschrift er zeker niet in deze vorm gelegen en was het überhaupt nog lang niet af geweest.

Chris, het voelt als de dag van gisteren, ik zocht een rijsimulator voor mijn onderzoek bij masterscriptie bij Rijkswaterstaat, toen bleek dat jij een rijsimulator in de doos op je kantoor had staan begon dit avontuur. Ik zag er wel een uitdaging in om geschikte software werkend te krijgen. De samenwerking was goed. Zo goed dat jij gesprekken met Rijkswaterstaat voortzette om een PhD plek te creëren. Ik ben je enorm dankbaar voor het vertrouwen dat je me hebt gegeven en het geduld dat je met me hebt gehad. Ik ben trots op wat we in de afgelopen jaren hebben bereikt. Wat begon met een computerscherm en een racestuurtje, is nu een aantal jaren later een complete rijsimulator die signalen kan uitwisselen met een systeem dat hersenactiviteit en spraak reactietijd registreert.

Stella, zonder Toegepaste Cognitieve Psychologie was deze kans voor mij nooit voorbijgekomen. Ik dacht terug aan de colleges van TCP1 toen ik al werkend als softwareontwikkelaar op zoek ging naar een Master (gefrustreerd door het gebrek aan aandacht voor de gebruiker). Als onderdeel van de dagelijkse begeleiding van mijn promotieonderzoek wist je vaak dingen van een andere kant te belichten. Soms vond ik bijval bij je, een andere keer lukte het je om me te overtuigen om het toch op een andere manier te doen. Ik vond het fijn om af en toe even mijn verhaal te doen en een kop koffie te halen, bedankt daarvoor!

Leon, onze samenwerking was top. Ondanks dat ik weet dat je een hoop te doen hebt stond je deur altijd voor mij open, bedankt! Je kwam zelfs geregeld bij mij langs, soms ging dat over mijn onderzoek maar misschien nog wel vaker over analysescripts die je zelf nog ooit geschreven had in Fortran en die nog steeds bleken te werken. We probeerden samen grip te krijgen op een aantal analysestappen die allemaal door verschillende programma's werden uitgevoerd: splitsen, exporteren, converteren, berekenen, samenvoegen, what can possibly go wrong?...

Zonder Rijkswaterstaat was dit hele proefschrift niet geschreven, ik ben Rijkswaterstaat dankbaar voor het vertrouwen wat ze in mij hadden. Maar nog belangrijker waren de mensen bij Rijkswaterstaat zelf die oprecht geïnteresseerd waren in wat ik deed. Allereerst Chantal, je zorgde ervoor dat ik me ook thuis voelde bij Rijkswaterstaat, ik kwam graag in Rijswijk en voelde me verbonden met de afdeling WVL. Marleen, Rini, ik vond het erg leuk dat jullie altijd even vroegen hoe het ervoor stond als ik weer een keer in Rijswijk was.

Kirsten, Lidwien, Fred, Hans (2x), Erna, Serge, Erik, Evert, Marga, Willem, Boris, Ilse (hoewel eigenlijk geen RWS) en talloze andere RWS-ers, bedankt voor jullie inzicht, interesse en gezelligheid!

Shamsi, it was amazing to spend a summer at Microsoft Research already in the early stages of my PhD, you gave me (at least the sense of) a great amount of freedom to design an experiment and introduced me to some amazing scientists like Michel, Shane, Ece, Ivan, who came up with great ideas and helped to work out the driving simulator setup with eye-tracking, Microsoft Band and Windows phone measurements. I feel honored that I was able to have several meetings with Eric Horvitz; the conversations we had still motivate me even today.

During my time at MSR I met with some great PhD students from around the world. It was great to meet all these people with different backgrounds. On my first day at MSR I met with Kavosh. We both felt lost but at least we were not alone, which made the process way more relaxed. It was great to team up and go to work-out at the biggest gym in the world (spent most of the time in the hot tub chatting about our projects and life). I really value our great friendship and always look forward to meeting you again for a new adventure, like some amazing trips we did to national parks, San Juan Islands, and what not.

I am also thankful to all wonderful people I met during ACM CHI conferences, and more important the smaller but more interesting AutoUI conference and its community. Thanks Lewis for the teasing chats we had about EEG studies. Thanks Andrew, Duncan, for the discussions on and off (my) topic. Some of these chats acted as true motivational speeches.

On the first day of my PhD I was shocked. Someone started the same day and had binders full of papers with notes on them; I had nothing. Kayla, that freaked me out, but it was great fun having you as office mate. I was not very pleased when I found out that I had to move to another office after returning from Microsoft. However, after some time I realized that this E0.38 office wasn't as bad as I thought. It didn't take long until I felt the other PhD students had accepted me. Since we were all in the same boat it was great to share a lot. It has been a great time with you, Kamerbuddies Febe, Renata, Isabell, we were all going to the same process, slightly out of sync. Too bad this crisis hit, I was looking forward to finally learn "Lili hop" (https://youtu.be/TVQVmqQYMAg). In the end I'm pretty happy that we actually did keep the empty vertical garden, great spacious touch to the room. In the last few month's all my office mates changed, which enabled me to fully concentrate on the final bits of my thesis.

Anne Marieke, Anneloes, 't was kort maar krachtig, gelukkig werd de drukte een beetje over de week verspreid, heel af en toe hadden we op vrijdag full house met Caroline, druk, gezellig, maar ook serieus, bedankt voor jullie wederhoor.

Iris, ook jij was een tijdje kamergenoot maar daarnaast ook erg belangrijk voor mijn onderzoek, je leerde me in sneltreinvaart de basisbegrippen van EEG-metingen, het aansluiten van een heel spinnenweb aan draden (de electroden), de beginselen van experimenten in Presentation en analysebomen in Brainvision, bedankt daarvoor! Gelukkig stond ik er tijdens veel metingen niet alleen voor, bedankt Nina, je ging met super veel energie van start waardoor we binnen no-time een hele berg proefpersonen hadden gemeten en data verzameld hebben voor 2 publicaties.

Son, Jim, Pieter, Humphrey, Dennis, Django, bedankt voor de technische ondersteuning, jullie dachten altijd mee en kwamen vaak met een betere oplossing dan die ik zelf kon bedenken. Eveline en Janet, bedankt voor de geweldige ondersteuning en oprechte interesse, zowel persoonlijk als in mijn onderzoek!

Verder zijn er nog tal van collega's die ik wil bedanken voor de input voor mijn proefschrift en de bijzonder prettige samenwerking tijdens het onderwijs wat we hebben gegeven: Jeroen, Jelle, Chris, Yun, Roy, Ignace, Stefan, Krista, Nico, Martijn, Nathan, David, Peter, Ada, Martijn, Paul, Jasper, Chris, Gijs, Yentl, Matthijs, Linda, Joke, Carlijn, Gabriël, Marnix, Sjoerd, Maarten, Susan, Ben, Jacob.

Martijn, bedankt voor alle interessante en minder interessante gesprekken die we voor een groot deel hadden terwijl we aan het reizen waren van of naar werk, lol om een gezamenlijke aandelenhobby die ineens instortte. Hoewel het allemaal anders liep dan verwacht ben ik blij dat je paranimf wilde zijn.

Waarschijnlijk is het niemand die dit proefschrift in handen krijgt ontgaan dat ik de laatste twee jaar van mijn onderzoek meerdere keren per week de boot in stap om te roeien, voor mij een perfecte manier om even mijn hoofd leeg te maken. Ik wil daarom de heren van BKF (Brute Kracht & Flamingo's), Eric, Harald, Jonne, Fons & coach Caroline, danken voor de welkome afleiding als ik niet aan mijn onderzoek aan het werken was.

Kristel, ik ben onder de indruk hoe goed jij bent met een pen en een tablet, uurtje brainstormen en je hebt hopla een supergave cover voor me getekend. En man wat kun jij roeien zeg! Ik kan er maar moeilijk bij hoeveel energie er van jou afstuitert, maar het werkt aanstekelijk. Supertof om samen gedisciplineerd te trainen en daardoor blik na blik te trekken.

Stella Maris, wat ben ik dankbaar om onderdeel uit te maken van zoiets moois en veel te veel namen om allemaal op te noemen. Het voelt het alsof we met iedereen een onvoorwaardelijke band hebben, waardoor we elkaar overal doorheen helpen. Daarnaast is het vast geen toeval dat dit al het zoveelste proefschrift is van een (ex) lid.

Lars, Paul, Maarten, Tom, soms ging het over mijn onderzoek, maar meestal over 1000 andere dingen, gelukkig maar. Bedankt voor jullie oprechte interesse en vriendschap. Sharon, bedankt voor je steun, geduld en het meedenken, Hans, Rosanne, Paul, bedankt voor het soms lenen van een stukje van jullie brein, ik waardeer het enorm dat jullie er altijd voor open stonden om over mijn onderzoek mee te denken. Aniek, bedankt, soms had ik precies die motiverende woorden van jou even nodig. Martine, Opa, bedankt voor de steun en het bijdragen aan de laatste loodjes!

Jeroen, Mama, Wim, naast een aantal van de mensen van de alinea hierboven hebben jullie het beste kunnen zien dat het niet altijd even makkelijk is om 4 jaar focus te hebben voor een onderwerp. Daarom wil ik jullie en ook de rest van mijn familie bedanken voor de onvoorwaardelijke support, het vertrouwen en de afleiding. Jeroen, bedankt dat je voor mij dat gekke pak aan wilt trekken en paranimf wilt zijn.

Papa, ik weet dat je trots bent en wat zou ik dit proefschrift graag met je gedeeld hebben. Ook al was je er niet, je was er ergens toch altijd wel!

Voor iedereen die dit leest, bedankt! Ik waardeer je interesse!

Appendix III

List of publications

Published peer reviewed

Van der Heiden, R. M., Iqbal, S. T., & Janssen, C. P. (2017). Priming drivers before handover in semi-autonomous cars. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 392-404).

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., Hardeman, L. E., Mans, K., & Kenemans, J. L. (2018). Susceptibility to audio signals during autonomous driving. *PloS one*, *13*(8).

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., & Merkx, C. L. (2019). Visual in-car warnings: How fast do drivers respond?. *Transportation research part F: traffic psychology and behaviour, 65*, 748-759.

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., & Kenemans, J. L. (2020). The influence of cognitive load on susceptibility to audio. *Acta psychologica, 205*, 103058

Under review

Van der Heiden, R. M., Kenemans, J. L., Donker, S. F., & Janssen, C. P. The effect of cognitive load on auditory susceptibility during automated driving, Under review

Appendix IV

About the author

Remo van der Heiden was born in Zürich, Switzerland, on the 20th of September 1987. After graduating from high school, he started studying Computer Science at the University of Applied Sciences in Eindhoven. He decided to continue his studies at Utrecht University and completed a Bachelor Artificial Intelligence. He then worked for some time as a software engineer building intelligent online systems. He quickly found his interest in human computer interaction due to the lack of incorporation of user testing for web applications at that time. He started with the Master Applied Cognitive Psychology at Utrecht University, in which he graduated cum laude after a year. He completed his Master thesis at the Dutch road authority (Rijkswaterstaat). The work was nominated for the Vliegenthart thesis prize. Remo then received funding to obtain a 4-year PhD position at Utrecht University in order to study human behavior in automated driving. In his first year he spent the summer working at Microsoft Research lab in Redmond. This collaboration lead to his first publication.