Mediator

Behavioural markers for degraded human performance

Deliverable D1.2 – WP1 – Public

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 814735.

Behavioural markers for degraded human performance

Work package 1, Deliverable D1.2

Please refer to this report as follows:

Borowsky, A., Oron-Gilad, T., Chasidim, H., Ahlström, C., Karlsson J.G., Bakker, B., … Christoph, M. (2020). Behavioural Markers for Degraded Human Performance. Deliverable D1.2 of the H2020 project MEDIATOR.

Project details:

Project start date:	01/05/2019
Duration:	48 months
Project name:	MEDIATOR – MEdiating between Driver and Intelligent Automated Transport systems on Our Roads
Coordinator:	Prof. dr. Nicole van Nes | SWOV – Institute for Road Safety Research Bezuidenhoutseweg 62, 2594 AW, The Hague, The Netherlands
	This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 814735.

Deliverable details:

Version:	Final
Dissemination level:	Public (PU)
Due date:	31/08/2020
Submission date:	28/12/2020

Lead contractor for this deliverable:

Avinoam Borowsky – Ben-Gurion University of the Negev, Israel

Report Author(s):

Borowsky, A., Oron-Gilad, T., Chassidim, H. - Ben-Gurion University of the Negev, Israel
Ahlström, C. - VTI National Road and Transport Research Institute, Sweden
Karlsson, J.G. - Autoliv, Sweden
Bakker, B. - Cygnify, The Netherlands
Beggiato, M., Rauh N. - Chemnitz University of Technology, Germany
Christoph, M., Cleij, D., van der Kint, S., Tinga, A. - SWOV Institute for Road Safety Research, Netherlands

Revision history

Date	Version	Reviewer	Description
07/12/2020	Integrated draft 1	Zon, Rolf – Netherlands Aerospace Centre, the Netherlands	Internal review
07/12/2020	Integrated draft 1	Prof. Linda Ng Boyle - University of Washington	External review
20/12/2020	Final draft	Avinoam Borowsky, Tal Oron-Gilad (Task leader) – Ben Gurion University of the Negev, Israel Michiel Christoph (WP leader) - SWOV Institute for Road Safety Research	Final revisions and checks
22/12/2020	Final report	Ingrid van Schagen (QA Officer) – SWOV Institute for Road Safety Research	Final editorial revisions
28/12/2020	Final deliverable	Prof. dr Nicole van Nes (Project Coordinator) - SWOV Institute for Road Safety Research	Approved

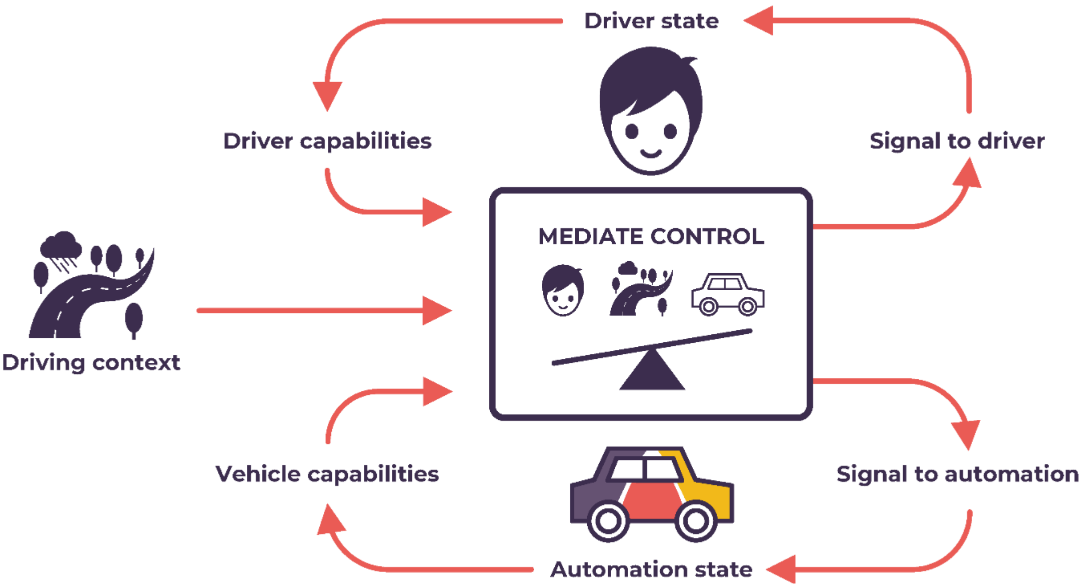
Table of contents

About MEDIATOR

MEDIATOR is a 4-year project led by SWOV. It started in May 2019. MEDIATOR will develop a mediating system for drivers in semi-automated and highly automated vehicles, resulting in safe, real-time switching between the human driver and automated system based on who is most fit to drive. MEDIATOR pursues a paradigm shift away from a view that prioritises either the driver or the automation, instead integrating the best of both.

Vision

Automated transport technology is developing rapidly for all transport modes, with huge safety potential. The transition to full automation, however, brings new risks, such as mode confusion, overreliance, reduced situational awareness and misuse. The driving task changes to a more supervisory role, reducing the task load and potentially leading to degraded human performance. Similarly, the automated system may not (yet) function in all situations. The objective of the Mediator system is to intelligently assess the strengths and weaknesses of both the driver and the automation and mediate between them, while also taking into account the driving context.



The Mediator system will constantly weigh driving context, driver state and vehicle automation status, while personalising its technology to the drivers' general competence, characteristics, and preferences.

MEDIATOR will optimise the safety potential of vehicle automation during the transition to full (level 5) automation. It will reduce risks, such as those caused by driver fatigue or inattention, or on the automation side imperfect automated driving technology. MEDIATOR will facilitate market exploitation by actively involving the automotive industry during the development process.
To accomplish the development of this support system MEDIATOR will integrate and enhance existing knowledge of human factors and HMI, taking advantage of the of expertise in other transport modes (aviation, rail and maritime). It will develop and adapt available technologies for real-time data collection, storage and analysis and incorporate the latest artificial intelligence techniques, such as deep learning.

Partners

MEDIATOR will be carried out by a consortium of highly qualified research and industry experts, representing a balanced mix of top universities and research organisations as well as several OEMs and suppliers. The consortium, supported by an international Industrial Advisory Board and a Scientific Advisory Board, will also represent all transport modes, maximising input from, and transferring results to, aviation, maritime and rail (with mode-specific adaptations).

Executive summary

Vehicle automation has the potential to improve driving safety and driver comfort. The Mediator system aims to aid the realization of this potential by mediating between the driver and the automation on who is fittest to drive. Make this trade-off in a timely and safe manner, requires both driver and automation fitness to be detected and predicted for the near future. Within the MEDIATOR project, numeral representations of driver and automation fitness, i.e., *time to driver fitness* and *time to driver unfitness* and the equivalents for automation fitness, were therefore defined. To take the first steps in estimating these variables in real time, the research described in this deliverable, focuses on detection and prediction of the driver state and finding relationships between driver state and driving performance. These relations were then used for setting behavioural markers that indicate degraded performance, which form the scientific basis for the estimates of *time to driver fitness* and *time to driver unfitness*. This research leads to the main outcome of this deliverable, i.e., the functional requirements for the driver module of the Mediator system.

The driver state detection algorithms for both fatigue and distraction have shown promising results and they will be further improved during the next phase of the project. Specifically, the preliminary prototype of the fatigue detection algorithm showed a Mean Absolute Error (MAE) of about 0.5 points on the Karolinska Sleepiness Scale (KSS), or a binary classification accuracy of about 90%. The accuracy for eyes-off-road classification measured with the F-score was 92.75%. Finally, the real-time comfort detection algorithm, which uses face tracking methods, was found to have potential in detecting users' satisfaction with the current operations of the automated system.

The evolvement of fatigue and distraction while driving with L2 automation and their effects on driving performance, was investigated in one driving simulator study and one on-road study. These studies revealed that fatigue was induced faster under L2 driving conditions than under manual driving conditions. Furthermore, the results from the driving simulator study revealed that drivers who played a game of Simon under L2 driving conditions reported lower KSS scores than drivers who did not play it. This finding is encouraging as it demonstrates the importance of engaging with a secondary task under L2 driving conditions as means of stalling fatigue development. This also emphasizes the complex trade-off between being distracted versus being fatigued under L2 driving conditions.
In terms of driving performance, the driving simulator study showed that under L2 driving conditions, regardless of whether drivers engaged with a secondary task, drivers had fewer number of glances on road hazards, which may indicate poorer situation awareness (SA) as compared to manual driving.

To evaluate the effect of driver distraction during assisted driving, a second driving simulator study investigated driving performance under both manual and L2 driving conditions. The results showed that drivers who were asked to engage with a secondary task during L2 driving conditions had a much smaller probability of identifying a hazard (~0.54) and deactivated the automation due to hazardous situations less often (probability of deactivating the automation =~0.116) than drivers who were not engaged with a secondary task (0.91, 0.22 respectively).

These findings indicate that, when distracted under L2 driving conditions, drivers' SA was poorer than drivers who were not distracted. Although drivers in this study were instructed to engage with a secondary task at specific time windows, the results have shown that drivers do regulate their behaviour and allocate more attention to the road in urban environments than in highway environments under L2 driving conditions.

Finally, long-term effects of drivers' comfort under L3 automated driving conditions were investigated via a literature overview. It was found that offline / long-term prediction of drivers' comfort based on prior knowledge seems to be a promising approach to detect potentially uncomfortable driving situations in advance, to prevent a decrease in drivers' comfort and to consider drivers' preferences for automated vs. manual driving in future driving situations.

Ultimately it can be concluded that, to make a safe and comfortable trade-off between driver and automation fitness, driver states should be monitored and predicted while driving with and without automation functions. Based on the research described in this deliverable, as well as on the knowledge gained during the analysis and experimentation phase of the MEDIATOR project, a first set of functional requirements for the driver module was drafted (Table 1). These functional requirements provide input to guide the further design and development of the Mediator driver state module.

Table 1 Functional requirements for the Mediator driver module

Functional requirements

The system shall estimate worst, likely and best-case *time to driver (un)fitness* based on driver fatigue and distraction estimates for the current driving context
• The system shall *estimate* driver *fatigue* and *distraction*
  ○ The system shall detect heart rate, respiratory and facial features
  ○ The system shall estimate KSS score
  ○ The system shall detect eyes on/off road
  ○ The system shall estimate distraction severity based on eyes off/on road
  ○ The system shall detect non-related driving task
• The system shall *predict* driver *fatigue* and *distraction*
  ○ The system shall predict fatigue progression based on current KSS score
  ○ The system shall predict loss of situation awareness based on distraction severity
  ○ The system shall predict time to driver fitness based on non-related driving task involvement
• The system shall estimate when the driver is *unfit to drive*
  ○ The driver is deemed unfit to drive if it can no longer execute the manual driving task in a sufficiently safe manner due to degraded cognitive abilities (fatigue) or loss of situation awareness (distraction).
• The system shall estimate *time to driver fitness* as the longest estimate based on distraction or fatigue
• The system shall estimate *time to driver unfitness* as the shortest estimate based on distraction or fatigue
• The system shall estimate *worst*, *likely* and *best* cases using reliability of the underlying estimates
• The system shall request *context relevant information* from the context module
  ○ Possible information can be time of day, type of road and situation complexity
• The system shall *personalise* these estimations to improve accuracy
  ○ Detect *driver ID*
    ▪ Possibly request sleep/driving history and driver age via HMI
  ○ Estimate driver fitness with *individualized algorithms* tuned to specific drivers
  ○ Estimate comfort based on *historical data* on user acceptance or rejection of the takeover suggestions

The system shall determine the *driver state class* as fit, distracted or fatigued
The system shall estimate *worst*, *likely* and *best*-case *time to driver discomfort* • The system shall compare upcoming *driving situations* with the identified uncomfortable driving situations

    - The system shall request relevant information from the *context module*
    - The system shall request information from the driver on uncomfortable driving situations
- The system shall estimate the *time* until the uncomfortable situation will occur
- The system shall estimate *worst*, *likely* and *best* case using the probability that comfort will be increased
- The system shall *personalise* the probability of a situation being uncomfortable to improve accuracy

1 Introduction

Vehicle automation has the potential to improve driving safety and driver comfort. To this aim, the Mediator system mediates between the driver and the automation on who is fittest to drive and supports the driver during his or her driving task. The research described in this deliverable is focused on the quantification of driver fitness and the associated markers of degraded driver performance. The main outcomes of this deliverable are the functional requirements for the driver state module of the Mediator system that will be implemented in the MEDIATOR project.

In order to determine who is fittest to drive, the Mediator system needs to estimate and predict driver fitness and automation fitness in a comparable manner. To this end, MEDIATOR has defined numeral representations of driver and automation fitness: *time to driver fitness* and *time to driver unfitness* and the equivalents for automation fitness.

Time to driver fitness is defined as the estimated time before a driver is able to safely perform the manual driving task. The *time to driver unfitness* is defined as the estimated time until the driver is no longer able to safely perform the manual driving task. Driver fitness is a complex concept that touches on many concepts, theoretical constructs, models and theories within human factors research. In order to reduce the scope and related complexity, MEDIATOR driver fitness will focus on two main driver states that are known to have a major impact on driver fitness: distraction and fatigue. Next to the driver fitness, also the driver comfort will be estimated. The main focus in the MEDIATOR project related to comfort is to initiate takeovers to improve driver comfort.

On the topic of driver distraction, Chapter 2 describes a simulator study focused on the effects of driver distraction on partially automated driving performance. In addition, a validation study on eyes-of-road detection algorithms is being described. Finally, Chapter 2 covers an initial study and describes a follow-up study on the relation between drivers' eyes-of-road behaviour and the level of situational awareness in specific driving contexts.

Chapter 3 is focussed on the research related to driver fatigue and boredom. The results of a large-scale field study in MEDIATOR were used for two purposes. First, an analysis was performed, focussing on the transition from an alert to a sleepy state, while driving on real roads in real traffic, both during manual driving and during partially automated driving with SAE level 2. Second, two fatigue detection algorithms, based on video data and physiological data respectively, were developed using this data set. Chapter 3 also describes a driving simulator study on evaluating the effects of task induced fatigue during SAE level 2 driving with and without a mitigating secondary task on hazard perception ability.

Optimizing driver comfort is an important aspect of the Mediator system and is covered in Chapter 4 on driver competence and comfort. The chapter describes studies aimed at identifying behavioural indicators of discomfort in automated driving using face tracking and facial expression analysis. In addition, offline or long-term comfort and competence research is discussed. This research is aimed at collecting and identifying situations from previous studies, in which drivers would prefer automation instead of driving manually or vice versa.

To assess the partial and full system effectiveness Key Performance Indicators (KPIs) are generally used. While such KPIs will be fully defined and used at a later stage of the project, for the purpose of defining data collection parameters, initial thoughts on relevant KPI(S) are already discussed in this report in Chapter 5.

Finally, Chapter 6 summarises the main conclusions of this deliverable and describes the resulting functional requirements for the driver module of the Mediator system.

In some of the chapters of this deliverable, references are made to 'the MEDIATOR use cases'. A total of ten use cases were developed to define the scope of the MEDIATOR project. These use cases are described in MEDIATOR deliverable D1.4[1]. This deliverable describes the research performed related to estimating and predicting driver fitness. As the Mediator system mediates between the driver and the automation, similar efforts to estimate and predict automation fitness will be described in D1.3 (in preparation).

1. Cleij, D., Bakker, B., Borowsky, A., Christoph, M., Fiorentino, A., van Grondelle, E., van Nes, N. (2020). Mediator System and Functional Requirements, Deliverable D1.4 of the H2020 project MEDIATOR.

2 Driver distraction

2.1 Examining the effects of drivers' distraction on partially automated driving performance: A driving simulator study

2.1.1 Background

2.1.1.1 Automation in vehicles

Over the past decade, there have been rapid developments in automobile automation. Automation in vehicles brings many benefits such as improvement in traffic flow, increased road safety and reduction of greenhouse gases. It can also prove of particular benefit to special populations such as younger and older drivers and mobility impaired drivers. In addition, it can be the basic platform of smart cities and can increase the detection of driver state in cases s/he is under the influence of alcohol, fatigue and drugs (Fisher et al., 2016). The SAE (Society of Automotive Engineers) defined six levels of driving automation. One of them is Level 2- Partial Automation in which at least two functions of the driving tasks are automated (typically longitudinal control and lateral motion control, that is, steering and acceleration/ brake). In this level it is expected that the human driver will monitor the road environment at all times and will take over control whenever is necessary.

2.1.1.2 Driver performance when interacting with secondary tasks in manual driving

According to the National Highway Traffic Safety Administration (NHTSA, 2009) it is estimated that about 16 percent of fatal crashes is contributed by driver inattention in various forms. One form of driver inattention is driver distraction, which is claimed to be a main factor in more than a half of inattention crashes (Stutts et al., 2001). Since wireless communication, driver assistance, and entertainment systems are becoming popular in vehicles, it is expected that the incidence of distraction-related crashes will increase (Young et al., 2007; Qin et al. 2019).

Engagement with secondary tasks while driving has different effects on driving performance and safety, depending on the task's type. Since the driving task is mainly visual and requires visual attention, secondary tasks that involve visual attention are known as distractions that have significant effect on driving (e.g., Borowsky et al., 2016). In other words, because driving is considered as a visual-manual task, performing a visual secondary task while driving will compete over limited resources of visual attention and may thus interfere with the driving task. Therefore, in order to keep a sufficient level of driving performance while concurrently engage with a secondary task, drivers must **regulate their engagement** with the secondary task in accordance with the driving context and driving requirements.

Another factor influencing driving performance is the driver's **personal interest** in the secondary task. Horrey et al. (2017) found that engaging in an interesting secondary task (e.g. listening to an interesting audio) causes longer response times to braking events and larger car following distances, compared to a boring secondary task. In addition, distraction has an impact on detecting hazardous situations (i.e., hazard perception). Burge & Chaparro (2018) examined how texting affects a driver’s response to hazards. Their results showed that texting while driving reduces drivers' ability to identify and respond to hazards.

Next, due to penetration of in-vehicle automation systems that take control over some of the driving tasks, the driving demands imposed on the driver are reduced which in turn may encourage drivers to engage with driving unrelated tasks even more than in manual driving. The next section presents evidence of the effects of distraction on driving performance in low levels of automation.

2.1.1.3 Driver performance when interacting with secondary tasks in low levels of driving automation (levels 1-2)

In partially automated driving of L2 the driver's role has changed from a vehicle controller to a supervisor where the driver is required to monitor both the position and speed of her/his vehicle that are being controlled by the automation. Under these conditions, the drivers' main task is to monitor for hazards, which depends on various factors such as the road conditions and the automation function (Lin et al., 2019). Under this level of automation, the driver has to remain vigilant in order to notice changes in the system (e.g., loss of lane tracking due to poor road markings) and has to remain situationally aware of keeping track of the changing roadway environment (Fisher et al., 2016). In addition, in cases that the automation fails or reaches its functional limits the driver has to take over control (De Winter et al., 2016). Relieving the driver from controlling the vehicle's position and speed results in a higher tendency to engage in non-driving tasks (Solís-Marcos et al., 2018). This increased tendency of engagement may lead to the conclusion that drivers will spend less time in monitoring the automation and more time in making off road glances toward the secondary task. The outcomes of this trade-off between engagement with a secondary task vs. monitoring the road and the automation are yet to be explored. For example, in a driving simulator study (Miller et al., 2015) participants were asked to drive under partially automated driving conditions and at several points the automation was transferred back to the human driver. During the automation mode some participants were asked to monitor the road and the automation and others were able to engage in Non-Driving Related Tasks (NDRT) such as watching a movie. The results have shown that 27% of the participants who were asked to monitor the automation exhibited behaviour of drowsiness compared to 6% of the drivers who engaged with a NDRT. Furthermore, the results have shown that there were no driving performance differences between the two groups after the control transfer to the human. Although these results imply that it is better to engage with a secondary task rather than to monitor the automation, it is not clear whether drivers who are engaged with a secondary task will be able to perceive hazards during periods of automated conditions. This issue is studied here.

In fact, the problem of driver distraction in partially automated driving (i.e., L2) has changed into a phenomenon of secondary task involvement. Engagement with non-driving related tasks (NDRT) while driving, affects driver's vigilance, situation awareness and workload. On the positive side, engagement with a secondary task can relieve driver's drowsiness (Miller et al., 2015; Lin et al., 2019) and help drivers to maintain alertness (Naujoks et al., 2018). However, engaging with secondary tasks may simultaneously compete with various process that are important for safe driving such as monitoring the automation performance and the process of hazard perception over the limited resource of visual attention, and may thus interfere with drivers' ability to read the road and anticipate upcoming hazardous situations, which may lead to difficulties in mitigating hazardous situations in a safely manner.

2.1.1.4 Hazard Perception – situation awareness of hazardous situations

Hazard perception (HP), i.e., hazard anticipation or hazard awareness, may be defined as drivers' ability to read the road and identify dangerous traffic situations (Yamani et al., 2016). It can be considered as driver's situation awareness of potentially dangerous situations in the traffic environment (Horswill & McKenna, 2004). HP not only includes scanning and perceiving hazard on time, but also correctly estimating it and knowing how to act in order to prevent it. It is the perception in time of partly or completely hidden situations which might become in the future hazardous (SWOV Institute for Road Safety Research, 2014). Research has consistently showed that hazard perception correlates with traffic crashes (Horswill and McKenna, 2004), especially among novice drivers (Horswill et al., 2015), and therefore it has received considerable attention over the years.

Measuring HP performance is typically done by analysing various eye movement parameters (e.g. fixations location, fixations duration on area of interest like mirrors or risky areas on the road) and by analysing drivers' response time and/or response rate to actual or potential road hazards. In driving simulator studies, a hazard is marked as identified if a drive has at least one fixation on the hazard. The time from the moment the hazard first appeared in the scene until the first fixation on the hazard is typical used to measure the amount of time it took drivers to identify the hazard. Hazard mitigation (HM) is the driver's action to respond to actual or potential hazards in the road.

There are various types of hazards and research has shown that latent hazards (i.e., hidden hazards) are the types of hazards that young-inexperienced drivers typically fail to identify (Borowsky and Oron-Gilad, 2013). Latent hazard is a potential hazard which is blocked by another object that has not yet materialized (e.g. a pedestrian in a crosswalk that is blocked by a parking vehicle right before the crosswalk). A cue such as road sign or other information can help the driver to anticipate a latent hazard (Krishnan et al., 2019). Latent hazards can only materialize in the future, therefore successful latent Hazards Anticipation demands the highest level of situation awareness (Krishnan et al., 2019). Glances towards areas from where a latent hazard may show up are highly important since they allow drivers to anticipate them and proactively apply a safe manoeuvre.

In conclusion, although drivers of partially automated vehicles at L2, are obligated to continuously monitor the driving task at all times, these drivers are ironically relieved from parts of the driving tasks that are now being operated by the automated driver. This reduction in the driving task demands may encourage drivers to engage with NDRTs. The current study is aimed at examining the ways by which engagement with secondary tasks affect drivers' performance under L2 driving conditions.

2.1.2 Goals and objectives

The main goal of the study is to examine how secondary task engagement at Level 2 affects various measures of driving performance, specifically, vehicle control, hazard perception and on-road and off-road glance behaviour.

A sub-goal of this study is to find a set of psycho-physiological parameters and performance measures that will help in the estimation of time to driver fitness and time to driver unfitness to drive. Time to driver fitness and time to driver unfitness are yet to be determined calculated measures that will provide the MEDIATOR decision logic component an estimation of the amount of time before the driver is no longer capable of driving (Time to driver unfitness) or the amount of time it will take the driver gain back his capability of driving (time to driver fitness).

2.1.3 Research hypotheses

- **H1**- Drivers under L2 driving conditions who are engaged with a secondary task will be less likely to identify potential hazards than drivers under both manual driving conditions and under L2 driving conditions without a secondary task.
- **H2**- Under L2 driving conditions drivers who are engaged with a non-driving related secondary task will be less likely to deactivate the automation mode than drivers who are not engaged with a secondary task. This hypothesis follows the assumption, that engagement with a secondary task take attentional resources that were otherwise allocated to the driving task.
- **H3**- Under L2 driving conditions drivers who are engaged with a non-driving related secondary task will be less likely to apply force on the brake pedal than drivers who are not engaged with a secondary task.

- **H4**- Drivers under L2 driving conditions who are engaged with a secondary task will experience higher mental workload (higher score in NASA- TLX questionnaire) than drivers under both manual driving conditions and under L2 driving conditions without a secondary task.

2.1.4 Method

2.1.4.1 Participants

Thirty participants, 13 females (mean age=25.3, SD=1.7) and 17 males (mean age= 26.23, SD=1.71) participated in the study and were randomly assigned into one of three experimental groups (n=10 for each group). All participants were undergraduate students from Ben-Gurion University of the Negev (BGU) in Beer Sheva city in Israel and received monetary compensation for their participation. All participants were right-handed and have normal or corrected to normal visual acuity (eye contact-lenses were allowed), normal contrast sensitivity, normal colour vision and no background of heart problems. Participants had a valid driver license for at least 5 years and reported to have no previous experience with a driving assistance system that provides both lateral and longitudinal support. The average number of years of driving was 8.45 (SD=1.87) and the weekly average of driving hours was 4.45 (SD=2.96). Three participants were involved in minor car crashes. Seventeen of the participants were car owners.

2.1.4.2 Materials and Apparatus

Driving Simulator

The study was conducted using RTI Driving Simulator (Realtime Technologies, Inc.) which consisted of an engineless Cadillac-STS family vehicle and a 7m diameter curved screen (2.4m X 6.1m) creating a visual angle of the virtual world of 165 degrees which was located at about 1m in front of the Cadillac (*Figure 1* left). Three laser projectors were used to display the virtual world on the curved screen and a designated software (Wrapalizer, Inc.) was used to do the edge blending. A designated 3D Perception Inc. was responsible for merging the three video channels into a single view. A screen behind the vehicle allowed a rear-view presentation of the virtual environment through the rear-view mirror inside the car. In addition, two 7" LCDs which were located within each one of the side mirrors. An in-vehicle display included a 10" touchscreen connected to a PC was used to display the secondary task (see Section 2.1.4.3.). The display was located on the central stack on the dashboard (see Figure 1 right).

Figure 1. Real-time technologies high fidelity driving simulator and an in-vehicle touch screen display

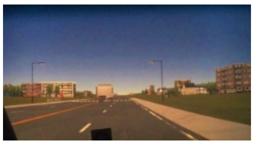
The RTI software product included three components: SimCreator, SimVista and Altia Designer. The SimCreator is a graphical, hierarchical, real time simulation and modelling software. The second component is SimVista which was responsible for creating and designing the simulation scenarios. The user's interface allows to drag objects into the virtual reality and to give them desirable behaviour, such as moving pedestrians and vehicles at a predefined time. In addition, there were stationary objects such as buildings, trees, vehicles, traffic signs and more. The third component is the Altia Designer software, a graphical tool that allowed designing the visuals of the in-vehicle display. The driving simulator provides various driving measures such as driving speed, steering angles, acceleration and braking at a rate of 30Hz.

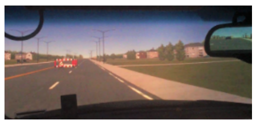
Driving Scenarios

The simulator test drive was a trip of approximately 25 min. The virtual drive included both highways environments and urban environments in daylight. The roads in the urban environment included a 2-lanes bidirectional road except for one scenario that had 1 lane in each travel direction. During the drive participants navigated through six different unmaterialized hazard scenarios and one materialized hazard. The seven hazards were distributed along the drive with an average of 3.5 minutes (average distance of 4 km) between 2 consecutive hazardous scenarios. The materialized hazard scenario always appeared as the seventh scenario at the end of the drive. Five different combinations of the six hazardous scenarios including the seventh hazard at the end were built and were randomly assigned between participants in order to reduce learning effects. Table 2 describes in detail the seven hazardous scenarios that were designed in the study. The "cue" field describes the hazard related visual cues that preceded each hazardous situation and indicated the participant of the hazard that is likely to occur in the near future. Furthermore, since these scenarios (except one) were used for the fatigue study (see Section 3.2) **Fout! Verwijzingsbron niet gevonden.** includes an additional scenario that was used only in the fatigue study.

For each of the six latent hazards a time window was defined during which the participant should identify the hazard in order to avoid a potential crash. The start of the time window is called a launch zone, a pre-defined area of each driving environment where the hazard becomes visible and the driver must begin glancing toward the target zone. Target zone is a visual Area of Interest (AOI) of a latent hazard (Krishnan et al., 2019). The time window ended when detection of the latent hazard would have been too late to avert a crash (Vlakveld et al., 2018). However, none of the six latent hazards materialized. The six-time windows ranged from 18 to 38 seconds and the seventh hazard time window was 22 seconds.

Table 2. A detailed description of the seven hazardous scenarios for both studies of distraction and fatigue

#	Scenario Name	Description	Scenario picture
1 Distraction & Fatigue	Pulling out from a line of parked vehicles	**Description:** The participant's vehicle drives on a road in industrial area (same type of road as in the urban environment). There are vehicles parked on the right side of the road. One of the vehicles uses its left signal to indicate of its intention to merge into the driver's lane. A truck is on a side street on the left waiting to enter traffic, serving as a distraction as the driver approaches the parked vehicle. **Hazard:** The vehicle which intends to pull out. **Road type:** Urban **Speed limit:** 60 km/h **AOI:** The parked vehicle with the signal light. **Cue:** Parked vehicle's signal. **Expected response:** The driver should glance at the car when he or she approaches to it.	
2 Distraction & Fatigue	A truck approaches an intersection from the left-hand side.	**Description:** The participant's vehicle is driving on the right lane in an urban environment. Approximately, 500 meters before the participant approaches a stop-controlled intersection (the participant has the right of way) a truck is approaching on the left lane from the left-hand side of the intersection. A bus that is stopped on the right lane on the left-hand side of the intersection obscures the truck that is driving on the left lane when it reaches the intersection. **Hazard:** The obscured truck on the left-hand side of the intersection that can turn left and merge onto the participant's path. **Road type:** Highway **Speed limit:** 90 km/h **AOI:** <u>Cue</u>: The approaching truck on the left-hand side of the intersection (before it becomes obscured). <u>Hazard</u>: Area in front of the stopping bus. **Cue:** Truck approaching from the left side of the intersection **Expected response:** When the participant is about to cross the intersection he or she should glance towards the front of the bus and make sure he notices the obscured truck. to make sure it stopped and does not burst into the intersection.	
3 Distraction & Fatigue	SUV obscures the view of a midblock crosswalk	**Description:** The participant's vehicle drives on the right lane of a road. An SUV is stopping on the left lane right in front of the crosswalk. **Hazard:** The front of the SUV obscures possible pedestrians that may be crossing the road in front of the SUV. **Road type:** Urban **Speed limit:** 60 km/h **AOI:** Area in front of the van at the crosswalk. **Cue:** crosswalk, stopping SUV. **Expected response:** The driver should notice the crosswalk and slow down. When passing by the crosswalk the participant should	

		look left in front of the SUV and search for a possible pedestrian. Note that the automation will not slowdown in this scenario**.**	
4 Distraction & Fatigue	Exit from a gas station	**Description:** The participant's vehicle drives on the right lane on a 4-lane road in a highway environment. The participant's Partially Autonomous Vehicle (PAV) passes near a gas station that has an entrance and an exit from and to the main road. Approximately 500 meters prior to the gas station, there is a vehicle that enters the gas station. When the participant's PAV passes there is a vehicle at the exit, waiting to merge into the main road. **Hazard:** The vehicle on the right is waiting to merge onto traffic from the exit area of the gas station. **Road type:** Highway **Speed limit:** 90 km/h **AOI:** Cue: The vehicle ahead enters the gas station Hazard: The gas station's exit. **Cue:** Vehicle ahead enters the entrance of the gas stations driving in the gas station's parking and may exit. **Expected response:** The driver should look at the cat at the gas station's exit.	
5 Distraction & Fatigue	Sudden traffic slowing cascade	**Description:** The participant's vehicle drives on the right lane of a 4-lane unidirectional curved road in a highway environment and approaches a traffic jam consisted of several lines of slowing vehicles. **Hazard**: Decelerating vehicles in front. The driver may collide with these vehicles in case the automation suddenly fails. **Road type:** Highway **Speed limit:** 90 km/h **AOI:** Cue: Area beyond the curve's apex. Hazard: Braking vehicles' back lights. Cascade at the end of the curve. **Cue:** The participant can make glances beyond the curve's apex and can notice lines of slowing cars. This can notify him of the potential hazard. **Expected response:** The car will automatically slow because of the ACC but we would like to see if the driver notices and looks at the braking lights of the lead vehicles. Other drivers might decide to disengage from the automation and apply their brake in order to prevent a crash.	
6 Distraction & Fatigue	Vehicle from behind approaches fast near construction zone	**Description:** The participant's vehicle drives on the right lane of a two-lane bidirectional road in an urban environment. There is a construction zone ahead on the left lane approximately 500 meters in front of the participant's car. While driving on the right lane there is an approaching car on the left lane behind the participant that approaches the construction zone relatively fast. **Hazard:** The approaching vehicle may pass the ego car and then cut in the simulator car in order not to enter the construction zone. **Road type:** Urban **Speed limit:** 60 km/h **AOI:** Cue: working zone. Hazard: Rear and left mirrors.	

		Cue: Working zone **Expected response:** The driver should notice to the vehicle in the mirrors and the work zone, disengage from the automation and slow down and slow down.	
7 Distraction only	Give a way to the vehicle in right in a four-way intersection	**Description:** The participant's vehicle approaches a four-way intersection ahead of a second vehicle. The second vehicle slows before entering to the intersection and then accelerate since there is no vehicle in the right. The participant's PAV's ACC will follow the second vehicle's velocity but when it will get close to the intersection a third vehicle will approach to the intersection from the right (which has a right of way). **Hazard:** Vehicle in the right that has a right of way. **Road type:** Urban **Speed limit:** 60 km/h **AOI:** The right side of the intersection. **Cue:** Right of way sign and the vehicle approaching from the right side of the intersection. **Expected response:** The driver should notice to the right of way sign and the approaching vehicle from the right side and stop in order to give him a right way.	
7 Fatigue only	Bus obscuring crosswalk on the opposing lane	**Description**: Bus stopping on a bus station on the opposing lane blocks a crosswalk **Hazard**: The bus blocks the view on possible pedestrians that may cross the road on the crosswalk behind the bus. **Road type**: Urban **AOI**: Area behind the bus at the crosswalk. **Cue:** Bus at crosswalk **Expected response**: The drive should glance at the area behind the bus at the crosswalk to search for possible pedestrians.	

.

Electro Cardio Gram (ECG)A BioPac ECG system (MP150) was utilized for measuring participants' Heart Rate at a rate of 2000Hz (see Figure 2- panel (a)). The system consists of a matched transmitter and receiver module which emulated a “wired” connection from subject to computer. The ECG was synchronized with the driving simulator. In this way we could easily define the start and the end of each scenario.

Eye tracker

Participants’ eye movements during driving were recorded via Tobii Pro Glasses 2 head mounted Eye Tracking System (Figure 2- panel (b)). This device weighs 45g, and includes four eye cameras, full HD wide angle scene camera, a gyroscope, and accelerometer sensors (Tobii, 2018). The eye tracker samples gaze position at a rate of 50Hz with accuracy of 0.5 degrees of visual angle. The Tobii pro Lab software (ver. 1.142) was utilized to analyse glance patterns. The main output of the data includes videos with gaze positions superimposed. The eye tracker was synchronized with the simulator.

(a) Dual Wireless ECG (b) Tobii Pro Glasses

Figure 2. Biopac heart rate monitoring and TOBII pro glasses eye tracker.

<u>Questionnaires</u>

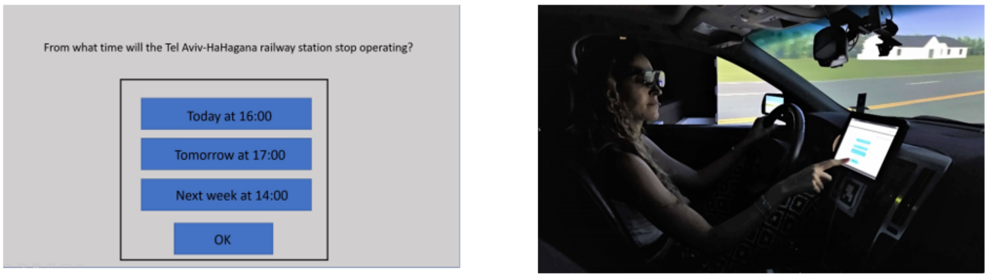
The participants were asked to complete three different questionnaires: (1) Demographic questionnaire. This questionnaire included 7 questions about the participant background and driving experience with regard to driving assistance systems (DAS). (2) Multidimensional Driving Style Inventory questionnaire (MDSI; Taubman-Ben-Ari et al., 2004). This questionnaire included 44 questions that are related to the driver's driving style. Each question contained a 6-points Likert scale (“1” represents no agreement; “6” represents full agreement). This questionnaire distinguishes between eight driving styles: Dissociative, Anxious, Risky, Angry, High-velocity, distress-reduction, Patient and Careful. (3) User’s adoption and trust aspects of autonomous vehicles questionnaire consists of 12 questions with a 7-point Likert scale (Choi & Ji, 2015). (4) After each experimental drive, participants completed a NASA-TLX questionnaire (Hart & Staveland, 1988) for mental load assessment. All the questionnaires were filled by using Google Forms application.

<u>The participant's Simulated Vehicle</u>

The participant's simulated vehicle was designed in two modes: manual driving mode and partially automated driving mode. In the manual driving mode, the participant was in control of all aspects of the driving task. In the partially automated driving mode, the vehicle simulated a L2 PAV (SAE, 2018). Under this mode, the PAV controlled the speed of the car including acceleration, maintained safe time headways from lead vehicles and maintained lane position (i.e., longitudinal and lateral control). Activating the PAV was done by the driver using a designated button located inside the car. Disengagement from the PAV and switching back to manual driving was done by the participant via pressing the brakes pedal.

2.1.4.3 Secondary task

The participants were asked to perform a visual distracting task while driving (see Figure 3- panel (a) for the task's interface). This secondary task included 14 text messages popped up in the in-vehicle display for 10 seconds each. Immediately afterwards, participants were asked to answer 28 multiple-choice questions: two questions for each message (see Figure 3- panel (b)). The topics of the text messages were diverse (e.g., public transportation, travelling, shows etc.). The secondary task was designed in-house with a designated software that came with the driving simulator. In order to assure that all the participants will engage with the secondary task at the same locations along the drive, the instances of the secondary task were always triggered at the locations along the drive for all drivers. Seven instances of the secondary task were triggered at the beginning of the launch zone of hazardous situations (one instance per scenario) and seven instances of the secondary task were triggered along the other parts of the drive that are not linked to any hazard. The latter seven instances of the secondary task were aimed at preventing drivers from associating between secondary task engagement and hazardous situations. The messages and the order of the questions were randomly assigned for each participant.



(a) Task's interface

(b) Engagement with the Secondary task

Figure 3 The experimental setup of the secondary *in-vehicle task*

2.1.4.4 Experimental Design

The experimental design is a mixed within-between-subjects design. The study included two, between-subjects, experimental conditions: (1) manual driving without a secondary task vs. L2 driving without a secondary task (i.e., baseline). This condition included 10 participants, and (2) L2 driving without a secondary task vs. L2 driving with a secondary task. This condition included 20 participants. In the first experimental condition participants were asked to complete two drives (one week separated between the drives) such that one drive was driven under manual driving conditions and the other drive under L2 driving conditions. Each drive lasted for 24 minutes and consisted of seven hazardous scenarios that were pseudo-randomized across drives (see Table 2 for a detailed description of the hazards). In the second experimental condition, participants were asked to complete the same drives as in the first condition except that in this condition the driving mode was L2 for both drives were only one of these drives included engagement with a visual-manual non-driving related task. The order of the drives was counter-balanced across participants. Participants' eye scanning behaviour and heart rate were recorded throughout the experiment. The vehicle dynamics were recorded as well. Demographic details, driving style and trust in automation data were collected via questionnaires prior to the beginning of the experiment and workload data were collected via a NASA-TLX questionnaire at the end of each drive.

<u>Independent variables</u>
The independent variables in the experiment were: (1) the level of automation which was a within-subjects variable (but only for those who were assigned to experimental condition 1) and included 2 levels: manual (L0), L2. (2) Secondary task existence which included 2 levels: yes, no. (3) Drive number which included 2 levels: drive 1, drive 2. The scenario's chronological order (between 1 and 7) as it appeared in each experimental drive. The road type of the specific scenario which included 2 levels: Highway, Urban.

Dependent variables
Table 3 below provides an overview of the dependent variables in this study as well as their meaning.

Table 3. A summary of the dependent variables and their meanings

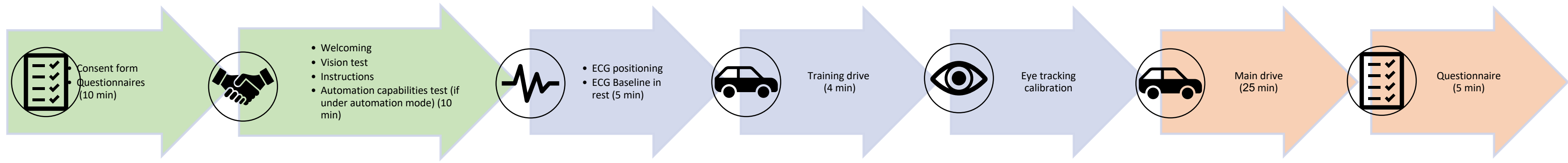
Measure	Definition	Name of variable
Did glance at the hazard?	Whether a participant had at least one glance of at least 100ms at the hazard (regardless of cue detection	GlanceYesNoNum
Did glance at the hazard given cue detection?	Whether a participant who detected the cue had at least one glance of at least 100ms at the hazard	GlanceYesNoNum (with a different data set)
Did deactivate the automation mode?	Whether a participant deactivated the automation mode in response to a hazard and drove manually	IsAutomationChangedNum

2.1.4.5 Experiment Procedure

At first, participants were requested to digitally sign an online informed consent form and filled in all questionnaires online except for the NASA-TLX. Upon their arrival at the human performance evaluation lab (HPEL) the participants underwent a visual acuity test (Snellen Chart) and a functional acuity contrast test (FACT) (Ginsburg, 2003). Participants who qualified, were given a written explanation about the simulator vehicle, the devices they are about to use and the experimental procedure in accordance with the driving condition they were assigned to. They were instructed to follow the Israeli traffic laws and drive as they would have normally done in similar real-world situations. Participants who began their first drive under level 2 of automation, were given an explanation about the automated vehicle and its capabilities (Adaptive Cruise Control ACC and Lane Keeping System LKS). Specifically, participants were informed that they are fully responsible for driving safety even when automation is active (i.e., they should keep monitoring the driving environment all the time). Deactivating the PAV driving mode and shifting to manual mode was done by breaking the brake pedal and returning to partially automated driving mode was done by pressing a button inside the vehicle. Switching between the two driving modes was possible at all times. After, they answered a short quiz about the automation capabilities in order to make sure they understand how it works (For the L2 driving where a secondary task was included, participants received an explanation on how it works and were told to engage with it (reading the messages and answering the questions after) as soon as possible as long as they felt they are not compromising their safety and the safety of other road users. These instructions were aimed at locating drivers' engagement with a secondary task in the "control" level of the model of Schömig & Metz (2013).

Next, the ECG electrodes were positioned on the participants' chest. Prior to their first drive they were given an IKEA magazine to browse while resting for 5 minutes. Their heart rate and Respiration Rate (R-R)- interval baseline was measured to serve as a baseline in the analysis phase. The participants drove a 4-minute training drive to become familiarized with the simulated driving environment and learnt about the sensitivity of the steering system and the car pedals. In this training session, participants experienced driving on straight roads, curves, intersections, and a transition from a two-lane road to a four-lane road. Each training session matched the driving conditions that were used in the experimental drive respectively (manual drive/ level 2 with secondary task/ level 2 without secondary task). After training, the participants were asked to wear the eye-tracker glasses and their gaze position was calibrated. Then, the first experimental drive began depending on the experimental condition. Once the 25 minutes' experimental drive ended, participants were asked to complete the NASA-TLX questionnaire regarding the drive they have just completed. A week later each participant came back to the lab approximately at the same hour as in the first session, for the second session of the experiment where he or she underwent the second experimental drive. On their second session, participants underwent the same procedure as in session one except for the 5 minutes rest period (the baseline at rest was only measured once in session 1 as mentioned). At the end of the two sessions, the participant was thanked and received monetary compensation (100 NIS) for his/her participation. Figure 4 presents a schematic description of the experimental procedure.

Week 1:



Week 2:

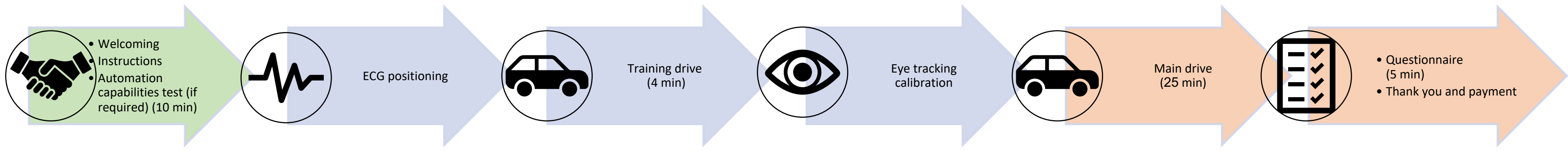


Figure 4 Schematic description of the experimental procedure

2.1.4.6 Statistical analysis procedure

For dependent variables that have a binary outcome, we used a logistic regression model with a logit link function within the Generalized Linear Mixed Models (GLMM) framework. These variables included hazard identification, hazard identification given cue detection and deactivate the automation mode. Statistical significance was assessed at α = 0.05.

The initial model in each analysis included the scenario chronological order, drive number, task existence (yes/no), road type (highway/urban), the second order interaction between road type and task existence, the second order interaction between road type and drive number, the second order interaction between road type and scenario chronological order, the second order interaction between task existence and drive number, the second order interaction between task existence and scenario chronological order, the second order interaction between drive number and scenario chronological order, the third order interaction between road type, task existence and drive number and the third order interaction between road type, task existence and scenario chronological order as fixed effects. The random effects were scenario and subject. The final model of each analysis was achieved via a backwards elimination procedure where all non-significant fixed interaction effects were removed from the model. For significant fixed effects with more than two levels, post hoc pairwise contrasts comparisons was applied and the Tukey HSD/ LSD procedure was used to correct for multiple comparisons (all statistical analysis computed in R software).

2.1.5 Results

The presented results are related to the second experimental condition (L2 and L2 with secondary task). The analyses reported here include all scenarios except for the seventh scenario. This latter scenario included a materialized hazard and its analysis is still under progress. The data of four drives was not valid and therefore was removed from the analysis. To summarize, the analyses is based on 18 drives under L2 and 18 drives under L2 with secondary task.

2.1.5.1 Hazard perception ability

<u>Hazard identification</u>

The initial Logistic regression model included a binary dependent variable representing hazard identification performance. This variable describes whether a participant had at least one glance of at least 100ms at the hazard ("1") or not ("0"). The final logistic regression model included 3 significant main effects of task existence (X_1^2=5.8001, p<0.05), drive number (X_1^2=6.4263, p<0.05) and cue detection (X_1^2=6.0402, p<0.05). Also, there were two significant second-order interactions, one between task existence and road type (X_1^2=4.7805, p<0.05) and second between task existence and scenario chronological order (X_5^2=11.9468, p<0.05).

First, with respect to the main effect of task existence, in the presence of the secondary task participants were less likely to identify a hazard (Estimated Mean = 0.548, SE=0.2439) than in the absence of the secondary task (0.913, 0.0847). Second, with respect to the main effect of drive number, in the first drive participants were more likely to identify a hazard (0.87, 10.114) than in the second drive (0.652, 0.228). Third, with respect to the main effect of cue detection, participants who identified the cue that preceded the hazard were more likely to identify the hazard (0.894, 0.0991) than participants who did not identify the cue (0.601, 0.2485). Also, with respect to the interaction between task existence and road type, while driving in highways participants were more likely to identify a hazard when a secondary task was absent (0.966, 0.0526) than when a secondary task was present (0.480, 0.3674) (P adjusted < 0.01). For the urban environment there was no effect of the secondary task and participants were equally likely to identify the hazard regardless of whether they engaged in a secondary task or not (P adjusted = 0.1547). The interaction between task exist and road type is presented by Figure 5. Finally, with respect to the interaction between task existence and scenario chronological order, it was found that under the condition of the forth place in the scenario chronological order, in the absence of secondary task the estimated probability to identify hazards was higher (EM = 0.9762, SE = 0.0346) compared to the presence of secondary task (EM = 0.0983, SE=0.1129). Moreover, it was found that the fifth (p_ = 0.0586) place in the scenario chronological order was marginally significant.

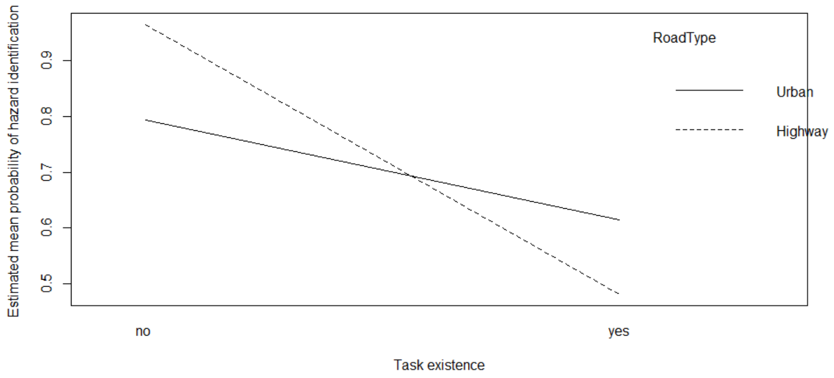


Figure 5. The interaction between task existence and road type

<u>Hazard identification given cue detection</u>

The data for this analysis included only the scenarios in which the participant did recognize the cue. The initial Logistic regression model included a binary dependent variable representing hazard identification performance. This variable describes whether a participant had at least one glance of at least 100ms at the hazard ("1") or not ("0"). The final model included one significant main effect of task existence (X_1^2=6.9696 p<0.05). Given that a participant identified the cue that preceded a hazard he or she was less likely to identify the hazard in the presence of a secondary task (0.738, 0.1974) than in the absence of a secondary task (0.933, 0.0666).

2.1.5.2 Mental workload

Figure 6 presents the average workload rate for each drive condition. In condition 1 (panel (a)) under L2 without non-driving related secondary task 10 participants rated lower workload (2.48) compared to the manual drive in which 9 participants rated 3.57. In condition 2 (panel (b)) 18 drivers under L2 driving condition who engaged with a non-driving related secondary task rated higher workload (3.99) compared to 18 drivers under L2 without non-driving related secondary task (2.89).

The initial Logistic regression model included a dependent variable representing the reported workload at the end of the drive. The initial model in included the task existence (yes\no) and automation level (0\2). The final model included two significant main effect of task existence (X_1^2 = 18.0106, p < 0.05) and automation level (X_1^2 = 6.7855, p < 0.05). First, with respect to the main effect of task existence, in the presence of the secondary task participants rated higher workload (Estimated Mean = 4.31, SE = 0.288) than in the absence of the secondary task workload (Estimated Mean = 3.21, SE = 0.229). Second, with respect to the main effect of automation level, under level 0 of automation (i.e., manual drive) participants rated higher workload (EM = 4.20, SE=0.351) than under level 2 pf automation (EM = 3.32, SE = 0.194).

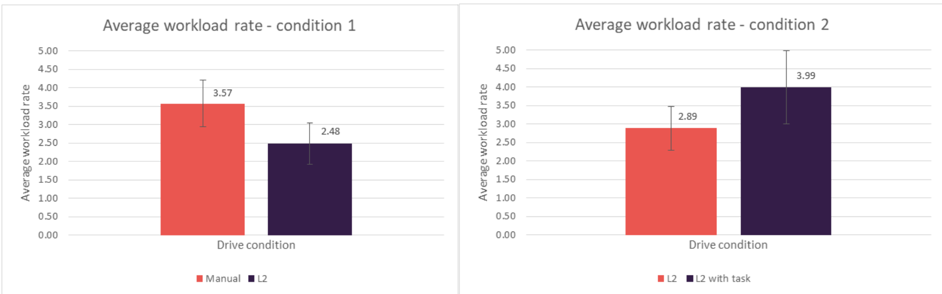


Figure 6. The average reported workload rate for each experimental condition. Condition 1. Manual vs. L2 driving without a secondary task. Condition 2 – L2 driving with and without a secondary task

2.1.5.3 Vehicle control- Deactivate the automation mode

The initial Logistic regression model included a binary dependent variable that describes whether a participant deactivate the automation mode during scenario ("1") or not ("0"). The final logistic regression model included one significant main effect task existence ($X_1^2 = 5.2681$, $p < 0.05$). In the presence of the secondary task the participants deactivate the automation mode less often (0.116, 0.0662) than in the absence of the secondary task (0.232, SE = 0.1101).

2.1.6 Discussion

This study aimed to investigate the effect of secondary task engagement at Level 2 on various measures of driving performance, specifically, vehicle control and hazard perception. The discussion is organized around the four hypotheses that followed the literature overview and discusses the findings considering the literature.

In accordance with our first hypothesis H1, the results demonstrated that under L2 driving conditions participants were significantly less likely to identify a hazard when a secondary task was present than when a secondary task was absent. These differences remained significant when the analysis was restricted to cases where the cue that preceded the hazard was identified. In fact, the cue preceding the hazard served as a mediating variable which dampened the effect of the existence of the secondary task. Nevertheless, the near ceiling effect of hazard identification when the secondary task was absent (~0.92 probability to identify a hazard) emphasize the deleterious effects of engaging with a secondary task during L2 driving conditions. Furthermore, the differential effects of hazard identification with respect to the road environment demonstrate the effects of the context on the way drivers regulate their attention allocation towards the road and the in-vehicle interface. It may be that in Urban environments drivers are in general more attuned to the driving environment and therefore are less likely to engage with a secondary task which may affect the likelihood of detecting a hazard. On highways on the other hand, drivers may feel they have more spare time to engage with secondary tasks and thus the likelihood that they will detect a hazard decrease due to their engagement with the secondary task.

The negative effects of the secondary task engagement under L2 condition are also emphasized by the confirmation of H2. It was found that under L2 driving conditions drivers who were engaged with a non-driving related secondary task were less likely to deactivate the automation mode than drivers who are not engaged with a secondary task. This finding provides further support to our argument that drivers who are engaged with a secondary task under L2 driving conditions are less attuned to the driving environment and are less aware of the driving context that require drivers' engagement in cases where hazards occur. Of course, we did not use materialized hazards in this experiment and thus failing to take over control over the vehicle in hazardous situations did not result in crashes.

Nevertheless, we argue that driver distraction under L2 driving should be minimized to avoid such risky driving behaviours. With regard to H3, this is still a work in progress.

Finally, consistent with the fourth hypothesis H4, drivers under L2 driving conditions who were engaged with a secondary task experienced higher mental workload (higher score in NASA- TLX scale) than drivers under both manual driving conditions and under L2 driving conditions without a secondary task. This finding may be explained by the fact that drivers under L2 have to keep monitoring the driving environment as they would do during manual driving. Engaging with secondary task in addition to the monitoring task causes a high mental workload that can be explained by drivers' impoverished ability to keep monitoring the road and the automation performance and may lead to higher uncertainty.

Ultimately, our findings provide initial evidence that drivers' ability to regulate their engagement with a secondary task under L2 driving conditions is impeded compared to manual driving. In other words, while drivers of manual driving are performing reasonably well when engaging with a secondary task (Schömig & Metz, 2013) by regulating their driving behaviour, under L2 conditions with a secondary task, drivers are poorer at identifying hazards and take over control of the vehicle less often than when not performing a secondary task. This conclusion emphasizes the need to incorporate corrective and preventive mediation in the Mediator system to make sure the driver is able to cope with instances of increased road demands.

2.2 Exploring the effectiveness of an eyes-off-road detection algorithm using existing video databases of naturalistic driving

2.2.1 Eyes-off-road detection with associated estimation of loss of situation awareness

Driver distraction can be defined as “diversion of attention away from activities critical for safe driving toward a competing activity” (Lee, 2008). Different types of distraction are distinguished (Ranney, 2001): physical (biomechanical) distraction, auditory distraction, visual distraction, and cognitive distraction. Several types of distraction may occur at the same time when a driver engages in a secondary task. For example, texting while driving may evoke visual, physical and cognitive distraction. All these different types of distraction may potentially impact safety but, visual distraction is often referred to as especially dangerous for safe driving, as driving is primarily a visual task (Klauer et al., 2006). As visual distraction can especially be considered as important for safety, MEDIATOR will mainly focus on visual distraction.

Driver monitoring systems that are able to adequately detect distracted driving could potentially mitigate the negative impact of distraction (Lee, 2013). When it comes to detecting visual distraction specifically, one way to operationalise visual distraction could be to consider a secondary activity while driving as visual distraction when it causes increased crash risk (thus impacting safe driving). Indeed, some evidence on crash risk related to specific visual behaviour when engaging in secondary tasks is available (Victor, 2015; Klauer et al., 2010; Simons-Morton, 2014). Yet, because there is no clear-cut ground truth to establish whether a driver is distracted or not distracted, testing the accuracy of distraction detection algorithms has always been debatable (Kircher and Ahlström, 2013).

Many researchers have focused on distraction detection algorithms that use eye glance direction as input. In Kircher and Ahlström (2013) a clear overview of such algorithms is presented. This overview includes a classification that describes the applied thresholds for classifying distraction. One frequently applied algorithm that relates eyes-off-road measures to distraction is the AttenD algorithm proposed by Kircher and Ahlström (2013). The outcomes of the AttenD algorithm have been demonstrated to correlate significantly with crash events from a large-scale naturalistic driving study (Seaman, 2017). In addition to algorithms being explored in the scientific literature, there are currently several commercial driver distraction monitoring systems available. Examples of these monitoring systems are: The Guardian (Seeing Machines)[2], SmartEye[3], LytxDriveCAM[4], Nauto[5] and Streamax[6]. Some of these commercial algorithms are applied by industry such as logistics and transport. However, information on the performance of these systems is limited. It is therefore impossible to evaluate these systems based on performance indicators.

The current section on eyes-off-road detection covers three main topics:
- Validation of eyes-off-road detection with associated estimation of loss of situation awareness using lab setting data (section 2.2.2)
- Setting up validation study for of eyes-off-road detection with associated estimation of loss of situation awareness using naturalistic driving (UDRIVE) (section 2.2.3)
- Setting up analysis for including driving context factors in the estimation of loss of situation awareness (section 2.2.4).

2.2.2 Validation of eyes-off-road detection with associated estimation of loss of situation awareness using lab setting data

The first approach is based on detecting eyes-off-road, from cameras aimed at the driver face, and using facial analysis algorithms developed by Cygnify (CYG). **Fout! Verwijzingsbron niet gevonden.** illustrates this eyes-off-road detection, using CYG's algorithms. The core of the approach is that eye gaze and eye closure can be estimated reliably.

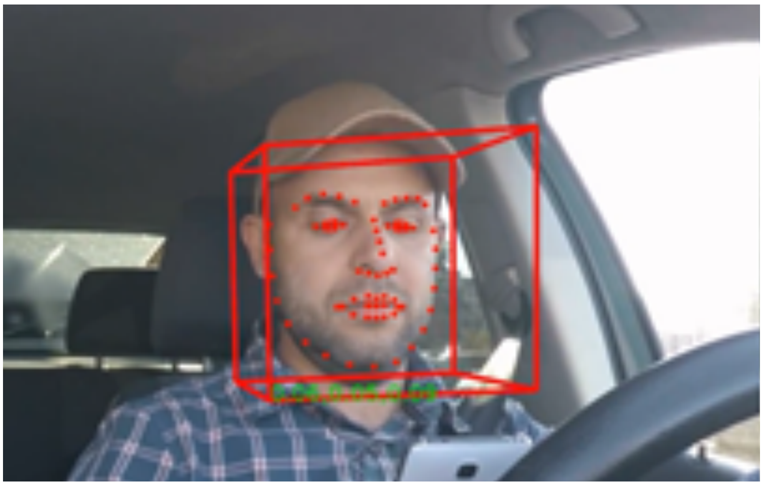


Figure 7. detection of eyes-off-road

[2] https://www.seeingmachines.com/guardian/
[3] https://smarteye.se/automotive-solutions/
[4] https://www.lytx.com/en-us/fleet-management/drivecam
[5] https://www.nauto.com/product/driver-behavior-alerts
[6] http://www.en.streamax.com/index.php?m=Fujian&a=show&id=286

Three main research questions were addressed to determine the performance of CYG's technology in eyes-off-road detection with the associated estimation of loss of situation awareness when using lab-setting data:

- How well can CYG's technology identify if drivers' eyes are directed off the forward road?
- How well can CYG's technology identify loss of situation awareness in 40-seconds episodes?
- How well can CYG's technology identify the severity of visual distraction in 40-seconds episodes?

Using CYG's frame-by-frame eyes-off-road detection as a basis, loss of situation awareness was estimated by looking at the *duration* and *frequency* of such glances off-road. In our preliminary work, we have used an existing algorithm in the literature, called the AttenD algorithm (Kircher and Ahlström, 2013) to assess, numerically, loss of situation awareness using such duration and frequency estimates.

The AttenD algorithm was simplified in order to deal with the lack of reliable ground truth data for eye glances directed at other locations than on/off road. Eyes off road is interpreted as a general, aggregated class where the eyes are clearly (according to human annotation) not on the road, independent of where he/she appears to look exactly. The simplified algorithm applies a time buffer of 2 seconds which runs empty each timestep that off-road glances are detected with a value equal to that timestep. The value of 2 seconds is derived from the well-known finding in driver safety literature that eyes off road for 2 seconds or more is correlated with significantly higher crash risk. For example, the time buffer is reduced with three times 0.1 seconds if the timestep between frames is 0.1 seconds and the eyes were off road for each of these three consecutive frames. When glances return to the road, a 0.1-second delay is applied in which the buffer does not change, to account for physiological adaptation to long distance focusing, before the buffer increases again with a value equivalent to the duration of one timestep for each consecutive frame where eyes on road are detected. When the eyes are not detected, glances on road are assumed. It is illustrated how the outcomes of the AttenD algorithm can be applied to classify 40-second episodes as episodes in which distraction occurs. For this analysis, the dataset is split into 40-second episodes. The (annotated) ground truth for eyes-off-road for each episode is fed into the simplified version of the AttenD algorithm. If the AttenD buffer runs empty at least once during a 40-second episode, the ground truth for this episode is classified as 'Distracted'.

An illustration of how this algorithm works is shown in Figure 8.

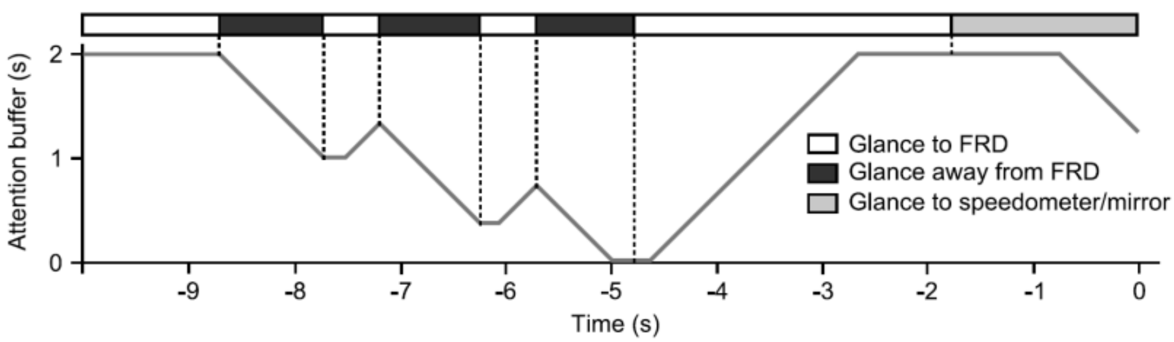


Figure 8 Illustration of attention buffer and glance behaviour

The key idea is that there is a "time buffer", which normally starts at 2 seconds. When the eyes are off-road, the buffer starts to deplete. When the buffer has reached 0, situation awareness is determined to be ‘lost’ which is associated with increased crash risk. Glances back to the road fill up the time buffer again, gradually regaining situation awareness, but only after some delay. This attention buffer is illustrated in Figure 8. This is of course an oversimplification of how situation awareness really works, for the purpose of our work a reasonable operationalisation.

2.2.2.1 Method

Eyes-off-road detection

For the main validation of the eyes-on/off-road detection accuracy a frame by frame annotated eyes-off-road signal was compared to the output of stage 2 of Cygnify’s algorithm.

To this aim, an overview of the performance of the algorithm will be presented in a table of confusion. The table of confusion includes all false and true detections of eyes on-road and eyes off-road by the algorithm (compared to the ground truth annotations). Additionally, the performance of the algorithm is analysed using the F-score (Equation 1). This indicator was chosen as the dataset is unbalanced and contains many more samples where eyes were on-road (78%) than samples where eyes were off-road (22%).

The F1-score is calculated as follows:

$$F1 = 2 \cdot \frac{Precision \times Recall}{Precision + Recall}$$

Equation 1 F1-Score

As can be seen in the formula above, the F1-score is based on precision and recall. A score based on these metrics has been recommended by scholars to assess how well an algorithm can detect the minority class (i.e. eyes off road) (Davis & Goadrich, 2006).

Precision is the proportion of correctly predicted eyes off road out of all eyes off-road predictions made by the algorithm. This value therefore reflects how often a false warning would have been given if the predicted eyes off road values would be used in a warning system.

Precision is calculated as follows:

$$Precision = \frac{\sum Correctly\ Predicted\ Eyes\ Off\ Road}{\sum All\ Predicted\ Eyes\ Off\ Road}$$

Equation 2 Precision

Recall indicates how many of the actual eyes-off-road events were predicted correctly. This metric provides insight into how many true eyes-off-road events would have been missed with this algorithm.

Recall is calculated as follows:

$$Recall = \frac{\sum Correctly\ Predicted\ \ Eyes\ Off\ Road}{\sum\ All\ Ground\ Truth\ Eyes\ Off\ Road}$$

Equation 3 Recall

In addition to computing an F1-score for the complete dataset, the F1-score was also computed for each participant separately. This allowed for exploring potential reasons for relatively lower F1-scores for specific participants. For this exploration, correlations were computed between the individual F1-scores and 1) the percentage of samples in which the eyes were not visible, 2) in which data quality problems occurred, and 3) in which unusual events occurred. Spearman's rho statistics were applied to compute the correlations as these statistics are reliable nonparametric measures for data that is independent of any known distribution.

Distraction detection

The distraction detection analysis focused on 40-second episodes instead of frame-by-frame data. For each episode, both the ground truth and predicted eyes-off-road measures were used as input for the simplified AttenD algorithm to classify the episode as either an episode in which distraction occurred or in which no distraction occurred. In line with the analyses for the eyes-off-road detection validation, the table of confusion, precision, recall and F1-score were calculated for the validation of distraction detection. Again, the F1-score was computed for each participant separately and potential reasons for relatively lower F1-scores were explored. Differences between data quality of participants with an F1-score of 100% and <100% were compared using Mann-Whitney U tests as these tests are reliable to apply to non-normal data. This exploration was also applied to episodes that were misclassified.

Distraction severity

The validation of distraction severity prediction was assessed in two steps: 1) assessing any bias or inconsistency in the distraction severity prediction and 2) determining the similarity between the ground truth and predicted values.

In the first step a linear model was fitted to the ground truth severity versus the predicted severity, which is a commonly applied approach to evaluate ground truth versus predicted values (Piñeiro, Perelman, Guerschman, & Paruelo, 2008; Kobayashi & Salam, 2000). A robust linear model was fitted in order to ensure the outcomes are less affected by extreme values. The model was evaluated based on its slope and its intercept, as these model properties describe the consistency and model bias respectively. A consistent model without any bias would have a slope of 1 and an intercept of 0 (Smith & Rose, 1995).

In the second step, the root mean squared error (RMSE) was calculated to assess the similarity between the ground truth and predicted values. The RMSE represents the mean distance between prediction and ground truth (Kobayashi & Salam, 2000) and it has the same unit as the quantity of the data, which facilitates interpretation. A measure related to RMSE is the mean squared deviation (MSD) which is the square of the RMSE. Compared to RMSE, MSD is better suited to directly compare the prediction and the ground truth values across assessment of models (Kobayashi & Salam, 2000; Gauch, Hwang, & Fick, 2003). The lower the value of RMSE and MSD, the better the prediction. As RMSE is frequently reported and MSD might in addition contribute to the assessment of the model in relationship to current and future models, we report both RMSE and MSD. These values will give a good indication about the accuracy of the Cygnify algorithm output for distraction severity.

2.2.2.2 Results

Eyes-off-road detection

Preliminary results by CYG on an open driver distraction in simulator dataset (https://osf.io/c42cn/, https://www.nature.com/articles/sdata2017110), with support and evaluation by SWOV, indicate that the accuracy for eyes-off-road classification the precision of the classification was 91.18%, recall was 94.38% and the F1-score was 92.75%.

<u>Distraction detection</u>
In our preliminary experiments, using “episodes” of 40 seconds long in that same open dataset described above, our methods are able to detect instances of situation awareness loss according to this criterion (AttenD time buffer has run out) with a precision of the classification of 93.69%, recall of 98.11% and an F-score of 95.85%.

<u>Distraction severity</u>
Furthermore, using our face analysis algorithms, our regression-like continuous estimate of distraction “severity”, defined as the number of seconds that the AttenD time buffer has run out within such a 40 second episode, the prediction for distraction severity had an RMSE of 0.06 and MSD was 0.00.

2.2.2.3 Discussion

The proof of concept of Cygnify’s algorithms looks promising when looking at the results on the three main research questions:

- How well can CYG’s technology identify if drivers’ eyes are directed off the forward road? The accuracy for eyes-off-road classification measured with the F-score was 92.75%.
- How well can CYG’s technology identify loss of situation awareness in 40-second episodes? The accuracy of classifying a 40-second episode as distracted based on the simplified AttenD algorithm was 95.85%.
- How well can CYG’s technology identify the severity of visual distraction in 40-second episodes? The prediction for distraction severity had an RMSE of 0.06 and MSD was 0.00.

Overall, the performance of the model was high for all evaluated parameters.

The current evaluation is based on data from a controlled experiment, further work is undertaken and described in the following section to evaluate how this technology performs on real road data with uncontrolled (lighting) conditions.

2.2.3 Validation of eyes-off-road detection with associated estimation of loss of situation awareness using naturalistic driving (UDRIVE)

As described in the previous section, CYG’s technology was evaluated using open-source videos of drivers in a driving simulator in a controlled experiment. CYG’s technology proved to be very promising. This evaluation, however, was performed using video data recorded in a lab setting whereas data from an actual field trial could be more challenging for the eyes-off-road detection technology. This section describers the preparations that have been made to evaluate CYG’s technology with video data from a naturalistic driving study (UDRIVE). It will start with a description of the UDRIVE data. Next, the selection of video segments that will be used will be described. The selected UDRIVE segments were ‘annotated’ manually. This process will be described a well in this section.

<u>UDRIVE naturalistic driving data</u>
For this study Naturalistic driving data of Dutch drivers in the UDRIVE dataset (van Nes et al, 2019) will be used. This data was collected between 2015 and 2017. In the current study, only data collected by Dutch car drivers will be used. Thirty-three drivers participated in the Dutch field trial; 3727 hours of data were collected and in total 230,842 kilometres were driven. Participants drove in

the vehicle for a period of 6 months. All participants were provided a leased Renault Clio IV which they drove in during their participation. The cars were equipped with seven cameras and a data acquisition system able to log CAN-bus data, GPS-data, Mobileye data and video footage (more information on the UDRIVE data collection see Bärgman et al., 2017 and Van Nes et al., 2019).

Selection of segments

In order to evaluate the eyes-of-road detection algorithms, a train and test dataset had to be created from the UDRIVE data. In order to have a ground truth for the evaluation, video data had to be manually annotated for the glance behaviour. Manual annotation is a labour (and therefore budget) intensive process restricting the amount of data that could be processed. In order to have a representative sample from the UDRIVE Dutch car drivers (representative in the sense that is sufficiently covers the factors that challenge eyes-of-road detection), 30-second segments were sampled from the UDRIVE Dutch subsample. This segment length is in line with several other related studies (Tivesten & Dozza, 2014; Seaman et al., 2017; Bärgman et al., 2017; Morando et al., 2019)

In a previous study (Christoph et al., 2019) a sample of the data collected by Dutch car drivers has been annotated for different categories of mobile phone use. This annotated subsample contains 656 trips from 28 participants and amounts to 225 hours of video data and 14.159 kilometers of driving. For the evaluation of CYG's technology it was important to have a homogenous distribution of eyes-off-road and eyes-on-road behaviour. Therefore, we sampled segments from the dataset described above. More specifically, half of the selected segments (per participant) were selected from episodes where drivers were engaged in a visual-manual task with their mobile phone. In these segments a higher proportion of eyes-off-road behaviour was expected. We will refer to these segments as *visual-manual (VM) task segments*. The other half of the selected segments were selected from episodes where drivers were not engaged with their mobile phone. In these segments a low proportion of eyes-off-road behaviour was expected. We will refer to these segments as *baseline segments.*

Additionally, several other selection criteria were setup to select the segments. Overall, these criteria were chosen to challenge the algorithm with factors that can occur during everyday driving. The following additional criteria were used:

- Equal distribution of different lightning conditions (day versus night)
- Equal distribution of different road types (urban, rural and highway)
- Exclude segments where the vehicle was standing still for the major part of the 30 second *baseline segment* or for the major part of the visual-manual task episode in the *VM-task segment*
- Only include participants where at least 10 *VM-task segments* were available

Annotation

Two researchers annotated the selected segments to capture variables that we will refer to as 'signals'. **Fout! Verwijzingsbron niet gevonden.** presents an overview of the signals that were annotated.

Table 4. Annotated state change signals

Signal	State
Glance direction	Eyes on road
	Eyes off road
	Unable to determine
Certainty glance direction	Sure about glance direction
	Unsure about glance direction
Direction	Driver looks to his left side
	Driver looks to the front (dashboard, front window)
	Driver looks to his right side
The number of eyes visible	Two eyes visible
	One eye visible
	No eyes visible
Face obstructed by object	True
	False
Face obstructed by accessory	True
	False
Video quality issues	No video quality issues
	Video quality issues
Passenger presence (front seat)	True
	False

A codebook was set-up as a reference document to serve as a guideline for annotators. It included a description of each variable with specific instructions on how to annotate. For example, considering the variable "Glance direction", annotators had to determine whether the driver had *eyes-on-road, eyes-off-road* or that *it was not possible to determine* the glance direction. The codebook provided the annotators with information on when to code the categories included in "Glance direction" as *eyes-on-road*, *eyes-of-road* and *unable to determine* while also providing an image captured from UDRIVE video's as an example. An illustration of the face camera footage is given in **Fout! Verwijzingsbron niet gevonden.**.

Figure 9. Illustrations of face camera footage from UDRIVE non-participant data. From left to right; eyes -on-road example (left), face obstructed by object example (middle) and video quality issues (right)

At the start of the annotation, the annotators were trained for half a day. This training consisted of a discussion about the codebook, a briefing about the annotation tools and practicing with the actual annotation. At the end of the training both annotators annotated eight segments allowing to determine the inter-rater reliability for all signals annotated. For all annotated signals the inter-rater reliability, measured with Cohen's Kappa was sufficient (k>.61) to proceed with the annotation.

Annotators were provided with an Excel sheet that contained information about ‘their’ segments, as well as a “link” to the segment that would automatically show the video. The annotation tool did not allow navigation to a timestamp before the start of the segment or after the end of the segment to minimise errors. The two annotators viewed the 30-second face video segments (most commonly on a frame-by-frame basis) and clicked buttons when any of the states for the different signals changed. During annotation, annotators were instructed to write down difficult segments or segments that they were unsure how to annotate correctly so they could be discussed and viewed during daily meetings that took place. Any changes or decisions that were made were noted in the codebook for future reference. About ten percent of all segments annotated were annotated by both annotators to allow for continuous inter-rater reliably monitoring.

After two weeks, 893 segments were annotated: 445 segments where drivers performed VM tasks and 448 segments where drivers did not. This dataset forms the basis for the follow-up activity where the accuracy of the eyes-of-road detection technology and the related estimates for loss of situation awareness will be evaluated. At the moment of writing this activity is starting up and will be reported in future deliverables of MEDIATOR.

2.2.4 Setting up analysis for including driving context factors in the estimation of loss of situation awareness

In Section 2.2.2 we described how the level or loss of situation awareness due to off-road glances was estimated by a simplified version of the AttenD algorithm. There might be arguments however, that the loss and regaining of situation awareness is different for different driving contexts and that the AttenD algorithm averages the influence of context on loss/regain of situation awareness. Generally speaking, workload increases when the complexity of the driving context increases (Törnros, & Bolling, 2006; Cantin, Lavallière, Simoneau, & Teasdale, 2009). Therefore, the time it takes for a driver to lose and regain situational awareness might vary as well. In a complex/high workload situation, more information about ‘the situation’ needs to be processed compared to a low complex/low workload situation potentially requiring more visual attention. There is evidence in literature that drivers adapt their glance behaviour according to the driving context. Wierwille (1993), for example, reported longer glances back to the road in a driving context that was more demanding. In addition, Christoph et al (2019) found that drivers regulate visual-manual phone tasks in relation to several driving context factors. Similarly, Schömig and Metz (2013) found in a simulator study that the choice of starting a secondary task as well as the duration of the secondary task is based on the demands of the driving task. More visual attention (expressed in a higher glance frequency) was given to driving when the situation became more complex compared to an undemanding driving context.

The rationale underlying this follow-up study, is that we can learn from how drivers distribute their visual attention between the road and non-driving related tasks in different contexts. With the assumption that in the majority of time drivers perform quite well in distributing their visual attention safely, this knowledge could be used to optimize to AttenD algorithm for different driving contexts. The UDRIVE database is a very suitable source of information to learn how drivers distribute their visual attention between the road and non-driving related tasks in different contexts. With the efforts of evaluating eyes-of-road detection technology described in Section 2.2.3, the entire UDRIVE databased can be enriched with information on glance behaviour (eyes on/off road). At the moment of writing this activity is starting up and will be reported in future deliverables of MEDIATOR.

2.3 References

Bärgman, J., van Nes, N., Christoph, M., Janssen, R., Heijne, V., Carsten, O., ... & Forcolin, F. (2017). *The UDrive dataset and key analysis results.* Deliverable 41.1 of the EU FP7 project UDRIVE Consortium.

Borowsky, A., Horrey, W. J., Liang, Y., Garabet, A., Simmons, L., & Fisher, D. L. (2016). The effects of brief visual interruption tasks on drivers' ability to resume their visual search for a pre-cued hazard. *Accident Analysis and Prevention*, *93*, 207–216. https://doi.org/10.1016/j.aap.2016.04.028

Borowsky, A., Shinar, D., & Oron-Gilad, T. (2010). Age, skill, and hazard perception in driving. *Accident Analysis and Prevention*, *42(4)*, 1240–1249. https://doi.org/10.1016/j.aap.2010.02.001

Burge, R., & Chaparro, A. (2018). An investigation of the effect of texting on hazard perception using fuzzy signal detection theory (fSDT). *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 123–132.

Cantin, V., Lavallière, M., Simoneau, M., and Teasdale, N. (2009). Mental workload when driving in a simulator: Effects of age and driving complexity. *Accident Analysis, & Prevention, 41(4)*, 763-771.

Choi, J. K., & Ji, Y. G. (2015). Investigating the Importance of Trust on Adopting an Autonomous Vehicle. International *Journal of Human-Computer Interaction, 31(10)*, 692–702. https://doi.org/10.1080/10447318.2015.1070549

Christoph, M., Wesseling, S., & van Nes, N. (2019). Self-regulation of drivers' mobile phone use: The influence of driving context. *Transportation Research Part F: traffic psychology and behaviour, 66*, 262-272.

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, 233-240.

De Winter, J. C. F., Stanton, N. A., Price, J. S., & Mistry, H. (2016). The effects of driving with different levels of unreliable automation on self-reported workload and secondary task performance. *International Journal of Vehicle Design*, *70(4)*, 297–324. https://doi.org/10.1504/ijvd.2016.076736

Fisher, D. L., Lohrenz, M., Moore, D., Nadler, E. D., & Pollard, J. K. (2016). Humans and intelligent vehicles: The hope, the help, and the harm. *IEEE Transactions on Intelligent Vehicles, 1(1)*, 56–67. https://doi.org/10.1109/TIV.2016.2555626

Gauch, H. G., Hwang, J. G., & Fick, G. W. (2003). Model evaluation by comparison of model-based predictions and measured values. *Agronomy Journal, 95(6)*, 1442-1446.

Ginsburg, A. P. (2003). Contrast sensitivity and functional vision. *International Ophthalmology Clinics, 43(2)*, 5–15.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology, 52(C)*, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Horrey, W. J., Lesch, M. F., Garabet, A., Simmons, L., & Maikala, R. (2017). Distraction and task engagement: How interesting and boring information impact driving performance and subjective and physiological responses. *Applied Ergonomics, 58*, 342–348. https://doi.org/10.1016/j.apergo.2016.07.011

Horswill, M. S., & McKenna, F. P. (2004). *Drivers' hazard perception ability: Situation awareness on the road. A Cognitive Approach to Situation Awareness*: Theory and Application, 155–175.

Kircher, K., & Ahlström, C. (2009). Issues related to the driver distraction detection algorithm AttenD. In: M. Regan and T. Victor (Eds), *First International Conference on Driver Distraction and Inattention.* Gothenburg, Sweden.

Kircher, K., & Ahlström, C. (2013). The driver distraction detection algorithm AttenD. In: M. A. Regan & John D. Lee (Eds.), *Driver distraction and inattention: Advances in research and countermeasures,* 327–348. Surrey, UK: Ashgate.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data*. NHTSA: Washington DC. p. 226.

Klauer, S. G., Guo, F., Sudweeks, J., & Dingus, T. A. (2010). *An analysis of driver inattention using a case-crossover approach on 100-car data* (No. HS-811 334). NHTSA, Washington DC

Kobayashi, K., & Salam, M. U. (2000). Comparing simulated and measured values using mean squared deviation and its components. *Agronomy Journal, 92(2)*, 345-352.

Krishnan, A., Samuel, S., Yamani, Y., Romoser, M. R. E., & Fisher, D. L. (2019). Effectiveness of a strategic hazard anticipation training intervention in high risk scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour, 67*, 43–56. https://doi.org/10.1016/j.trf.2019.10.004

Lee, J. D., Young, K. L., & Regan, M. A. (2008). Defining driver distraction. *Driver distraction: Theory, effects, and mitigation, 13(4)*, 31-40.

Lin, R., Liu, N., Ma, L., Zhang, T., & Zhang, W. (2019). Exploring the self-regulation of secondary task engagement in the context of partially automated driving: A pilot study. *Transportation Research Part F: Traffic Psychology and Behaviour, 64*, 147–160. https://doi.org/10.1016/j.trf.2019.05.005

Miller, D., Sun, A., Johns, M., Ive, H., Sirkin, D., Aich, S., & Ju, W. (2015, September). Distraction becomes engagement in automated driving. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59(1*),1676-1680. Sage CA: Los Angeles, CA: SAGE Publications.

Naujoks, F., Höfling, S., Purucker, C., & Zeeb, K. (2018). From partial and high automation to manual driving: Relationship between non-driving related tasks, drowsiness and take-over performance. *Accident Analysis and Prevention, 121*, 28–42. https://doi.org/10.1016/j.aap.2018.08.018

NHTSA (2009). *An Examination of Driver Distraction as Recorded in NHTSA Databases*. Report number DOT HS 811 216. NHTSA, Washington DC

Piñeiro, G., Perelman, S., Guerschman, J. P., & Paruelo, J. M. (2008). How to evaluate models: observed vs. predicted or predicted vs. observed? *Ecological Modelling, 216(3-4)*, 316-322.

Qin, L., Li, Z. R., Chen, Z., Bill, M. A., & Noyce, D. A. (2019). Understanding driver distractions in fatal crashes: An exploratory empirical analysis. *Journal of Safety Research, 69*, 23-31.

Ranney, T. A., Garrott, W. R., & Goodman, M. J. (2001). *NHTSA driver distraction research: Past, present, and future* (No. 2001-06-0177). SAE Technical Paper.

SAE. (2018). SAE International Releases Updated Visual Chart for Its “Levels of Driving Automation” Standard for Self-Driving Vehicles. https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-“levels-of-driving-automation”-standard-for-self-driving-vehicles

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

Seaman, S., Lee, J., Seppelt, B., Angell, L., Mehler, B., & Reimer, B. (2017). *It's all in the timing: using the AttenD algorithm to assess texting in the NEST naturalistic driving database*. Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, June 26-29, 2017, 403-409.

Smith, E. P., & Rose, K. A. (1995). Model goodness-of-fit analysis using regression and related techniques. *Ecological Modelling, 77(1)*, 49-64.

Solís-Marcos, I., Ahlström, C., & Kircher, K. (2018). Performance of an Additional Task During Level 2 Automated Driving: An On-Road Study Comparing Drivers With and Without Experience With Partial Automation. *Human Factors, 60(6)*, 778–792. https://doi.org/10.1177/0018720818773636

Stutts, J. C., Reinfurt, D. W., & Rodgman, E. A. (2001). *The role of driver distraction in crashes: an analysis of 1995-1999 Crashworthiness Data System Data*. Annual Proceedings / Association for the Advancement of Automotive Medicine. Association for the Advancement of Automotive Medicine, 45(May), 287–301.

SWOV. (2014). SWOV Fact sheet: *Hazard perception and how to test it*. September, 1–6. The Hague, SWOV Institute for Road Safety Research.

Taubman-Ben-Ari, O., Mikulincer, M., & Gillath, O. (2004). The multidimensional driving style inventory - Scale construct and validation. *Accident Analysis and Prevention*, *36(3)*, 323–332. https://doi.org/10.1016/S0001-4575(03)00010-1

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

Tobii, A. B. (2018). Tobii pro glasses 2.

Törnros, J., & Bolling, A. (2006). Mobile phone use–effects of conversation on mental workload and driving speed in rural and urban environments. *Transportation Research Part F: Traffic Psychology and Behaviour, 9(4)*, 298-306.

van Nes, N., Bärgman, J., Christoph, M., & van Schagen, I. (2019). The potential of naturalistic driving for in-depth understanding of driver behavior: UDRIVE results and beyond. *Safety Science, 119*, 11-20.

Victor, T., Dozza, M., Bärgman, J., Boda, C. N., Engström, J., Flannagan, C., ... & Markkula, G. (2015). *Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk.* No. SHRP 2 Report S2-S08A-RW-1.

Vlakveld, W., van Nes, N., de Bruin, J., Vissers, L., & van der Kroft, M. (2018). Situation awareness increases when drivers have more time to take over the wheel in a Level 3 automated car: A simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour, 58*, 917–929. https://doi.org/10.1016/j.trf.2018.07.025

Wierwille, W. W. (1993). Demands on driver resources associated with introducing advanced technology into the vehicle. *Transportation Research Part C: Emerging Technologies*, 1(2), 133-142.

Yamani, Y., Samuel, S., Knodler, M. A., & Fisher, D. L. (2016). Evaluation of the effectiveness of a multi-skill program for training younger drivers on higher cognitive skills. *Applied Ergonomics, 52*, 135–141. https://doi.org/10.1016/j.apergo.2015.07.005

Young, K., Regan, M., & Hammer, M. (2007). *Driver distraction: A review of the literature*. Distracted Driving, 2007, 379–405.

3 Driver fatigue and boredom

3.1 Transition from alert state to sleepy state under manual and L2 driving conditions. An on-road study

3.1.1 Background

Driver fatigue is a well-known contributor to about 20% of all road crashes. Both sleepiness due internal factors, such as the window of circadian low and homeostatic effects of time awake (Åkerstedt, Connor, Gray, & Kecklund, 2008), as well as external factors, such as task related cognitive underload (May & Baldwin, 2009), must be considered to understand crash risk. A sleepy driver has a 2.5–3.5 times higher risk of being involved in a crash (Bioulac et al., 2017; Connor et al., 2002) than a well-rested driver. The reasons for such a crash risk are not only related to insufficient prior sleep and night-time driving (Åkerstedt, Folkard, & Portin, 2004), but also to task related underload or overload (Williamson et al., 2011).

Automated driving is often put forward as the solution to all human error related problems in traffic. This might be true if the car is fully automated and the driver is relieved from the driving task, but this is not foreseen to happen in the nearby future. Until then, the driver will be responsible for monitoring the situation and eventually take over the control if automation disengages or makes a mistake (Kyriakidis et al., 2019).

Due to the lack of active involvement in the driving situation and due to monotonous driving environments, drivers using automation are likely to become fatigued faster than manual drivers (Schömig, Hargutt, Neukum, Petermann-Stock, & Othersen, 2015; Vogelpohl, Kühn, Hummel, & Vollrath, 2019). However, little is known about the progression of fatigue during automated driving and its effects on the ability to take back manual control after a take-over request. Preliminary findings from driving simulator research indicate that drivers are unable to stay alert during extended periods of automated driving (Vogelpohl et al., 2019). Fatigued drivers could therefore pose a serious hazard, both in complex take-over situations where situation awareness is required, as well as in monotonous driving situations where drivers might doze off after long periods of inactivity.

3.1.2 Aim and connection to MEDIATOR Use Situations

The aim of this study was to compare the transition from alert to sleepy while driving on real roads in real traffic, both during manual driving and during partially automated driving on SAE level 2. The study was conducted in a real-life setting due to the clear differences in how sleepiness and fatigue develops in driving simulators compared to on real roads in real traffic (Fors, Ahlström, & Anund, 2018; Hallvig et al., 2013). In a driving simulator, the human operator will not be motivated to stay awake for two reasons: (1) they can quickly become bored in a monotonous driving simulator experiment. Another reason is that there is no real threat in a simulator. Without real consequences of a potential crash, drivers tend to fight less hard to remain awake.

In relation to the MEDIATOR use cases and use situations, the data collection is motivated by the need for new data to facilitate fatigue algorithm detection development in WP2. This algorithm is crucial for most use situations when estimating if the driver is fit enough to drive in takeover situations, or if the driver is unfit and is not suitable to drive at a certain level of automation in his/her present state. The data collection is also relevant for the prediction of Time To Driver Fitness (TTDU) and Time To Driver Unfitness (TTDF) in level 2 driving, which is needed in use situations 7, 8 and 9.

3.1.3 Method

Two study populations were included in this study: (i) a homogenous sample with the aim to reduce intra-individual trait-like differences, and (ii) a heterogenous sample that facilitates a more diverse group of drivers, see Table 5. The first group will be referred to as the XC90 group and the seconds group as the V60 group, based on the experimental vehicles that the two groups drove.

Table 5. Inclusion criteria for real road test

XC90 car	V60 car
Homogenous sample (n=40)	Heterogeneous sample (n=40)
Age: 30-60 years	Age: – 18-30 years old (n=10), 30-45 years old (n=15), 45-60 years old (n=15)
Gender: males (n=20) females (n=20)	Gender: males (n=20) females (n=20)
Driving licence for passenger car	Driving licence for passenger car
Experienced with advance driving assistance systems (ACC, LKA or similar)	Experienced with advance driving assistance systems (ACC, LKA or similar)
No glasses	No glasses
No disability that prevents from driving a passenger car.	No disability that prevents from driving a passenger car.
No problem with motion sickness	No problem with motion sickness
	Facial features: Asian (n=20), Caucasian (n=10 with glasses; n=10 without glasses)

In total eighty-nine experienced drivers (36 women and 53 men) participated in the study (mean age = 38 years, SD = 11 years, range = 20–59 years). For practical reasons, the experience criterion had to be relaxed in the end of the experiment, resulting in 54 drivers who had experience with automatic cruise control, 44 with experience with lane keeping assistance, 48 with parking assistance, and 19 with level 2 assistance. Seventeen drivers did not have experience with any advanced driving assistance systems. Each participant received 4000 SEK (≈ €400) for participation. The study was approved by the Swedish Ethical Review Authority (Dnr 2019-04813) and the Swedish government approved the experiment with sleepy drivers on real roads (N2007/5326/TR).

Preparations before arrival:
Three days before arrival all participants used a sleep and wake diary. This provides a view on how and when the participants slept the days before the experiment, and also a deeper understanding of other factors influencing their performance.

The experimental days: The study had a within-subject (2x2) design. Factors for alertness level (alert vs sleep deprived) and driving mode (manual driving without automation vs driving with level 2 automation) were included in the experimental design. In total 4 drivers participated in parallel each experiment day. Drivers named A and B drove the XC90 and driver C and D drove the V60. Each participant first drove in the afternoon (alert condition) and then during night-time or early morning hours (sleep deprived condition). The afternoon drive started at 15.00h (driver A and C) or 17.00h (driver B and D) and the night-drive started at 01.00h (driver A and C) or 03.00h (driver B and D). The normal driving and the automated driving sessions took place on different days, thus requiring two visits to VTI. The order was balanced between participants so half of the drivers started with the manual driver and the other half started with the automated drive. The same procedure was used for both days, see Figure 10. The difference between the days was only if the drivers drove with or without level 2 automation activated.

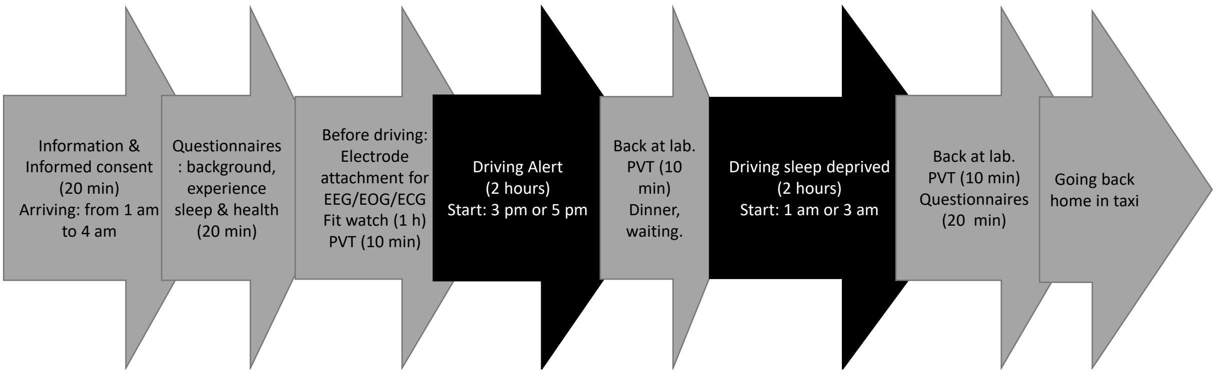


Figure 10. The procedure from arrival to back home.

The route driven was about 90km and took approximately 2h to drive back and forth. The participant left from (and returned) to VTI, but the actual experimental route started and ended at exit 111 on the motorway E4. The participants drove south to exit 104 (Gränna) and turned back towards Linköping, see Figure 11. This is one of the roads that VTI has permission to use for sleepiness tests on real roads (N2007/5326/TR). No additional tasks were performed during the drive.

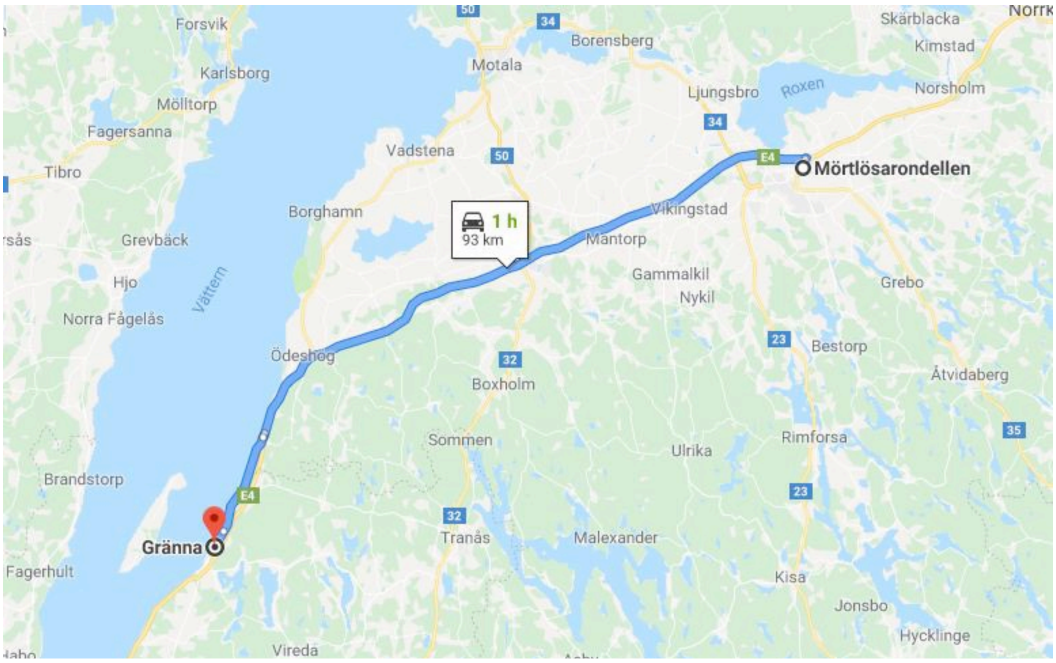


Figure 11. The route driven on E4 to Gränna and back to Linköping.

Test vehicles: The vehicles were so-called SAE level 2 partially automated cars. Such cars have recently been introduced on the market and allows “hands off” the steering wheel (automatic lane keeping) and “feet off” the pedals (automatic speed and distance keeping), but not “eyes off” the road. To ensure that the driver is engaged in the driving task, these vehicles are equipped with a Deadman’s switch that forces the driver to touch the steering wheel at least every 15th second. The brand of the test vehicles was Volvo and both cars were equipped with Pilot Assist version 2, see

Figure 12. One car was a 2015 Volvo XC90 updated to Pilot Assist version 2 and owned by Autoliv AB, and the other car was a 2020 Volvo V60 owned by Smart Eye AB. Two different models were used for practical reasons of ownership and the availability of test vehicles. Both cars were equipped with dual command allowing the test leader to intervene if needed. All safety drivers had undertaken a specific safety driving course. This is mandatory and required in the permission to run experiments with sleepy drivers on real roads.

Figure 12. Pilot assist, when the steering wheel is grey the pilot assist is activated but the steering assist is not active. When the steering wheel is green the steering assist is active.

Measurements: Data were collected from the vehicle (Speed, GPS, etc.) and from the participants (physiology, video, reaction time, self-reported sleepiness and questionnaires). Physiological data (electrocardiography, ECG), eye tracking data, vehicle data, self-reporting's, psychomotor vigilance task data (PVT), and questionnaire data were collected for both groups. Brain activity (electroencephalography (EEG)) and eye blinks (electrooculography (EOG)) were only collected for the XC90 group. The reason for not measuring brain activity and eye blinks in both groups was that unobstructed video data (without electrodes in the face and on the head) was desired in at least one of the vehicles, to ensure generalizability of video-based sleepiness detection algorithms developed based on the recorded data. The data collection was a collaborative effort which has implications for data sharing. This effectively means that face video data is only available to the MEDIATOR project for the XC90 group, and the eye tracking data in combination with the face videos can only be accessed by VTI and Autoliv AB.

Physiological data were recorded with two different recording systems. The V60 group used a Vitaport 3 (Temec Instruments BV, the Netherlands) to record ECG and respiration while the homogenous group (XC90) used an ego Sport system (ANT Neuro, Hengelo, Netherlands) to record EEG+EOG+ECG+respiration. The 64 EEG-electrodes are attached to the scalp via a cap (Figure 13), whereas the EOG/ECG electrodes are attached with adhesive tape. In addition, Garmin Fit (Forerunner 645) watches paired with a Polar H10 pulse sensor were used by all participants at both visits, from arrival in the afternoon until departure the next morning, see Figure 14.

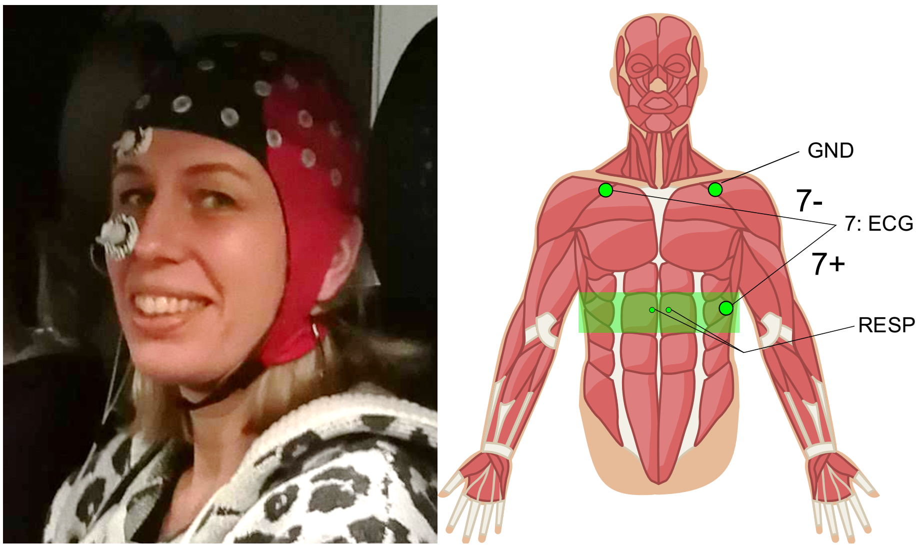


Figure 13. ANT Neuro (XC90) (left) and the position of electrodes using VitaPort (V60) (right)

Figure 14. Garmin fit watch used by all participants during the stay.

The drivers' eye movements were tracked with a 3-camera eye tracking system (TrackingSW SDK version 0.9.0, Smart Eye AB, Gothenburg, Sweden). The videos from the eye tracker were stored, allowing post-processing of the videos to improve eye tracking data if needed. Videos of the upper body of the driver and of the forward roadway were also recorded. In addition to gaze direction, also blink behaviour and pupil diameter were extracted from the face videos. The Karolinska sleepiness scale for self-reported sleepiness was used (Åkerstedt & Gillberg, 1990). The participants were asked by the experiment leader to rate their sleepiness every 5th minute during the drives.

1 = extremely alert
2 = very alert
3 = alert
4 = rather alert
5 = neither alert nor sleepy
6 = some signs of sleepiness
7 = sleepy, no effort to stay awake
8 = sleepy, some effort to stay awake
9 = very sleepy, great effort to keep awake, fighting sleep

The participants also performed a simple 10-minute reaction time test called the Psychomotor vigilance task (PVT) before and after each drive (Balkin et al., 2004; Loh, Lamond, Dorrian, Roach, & Dawson, 2004). The test randomly shows a digit on a screen and the participant must respond as fast as possible.

Data and statistical analysis: Of the 356 planned trials, 2 were cancelled due to bad weather, 1 was cancelled due to technical issues with the logging equipment, 4 were cancelled due to hazardous drivers, and 18 were cancelled due to scheduling and availability issues, leaving 333 trials for analysis. The data from the physiological measurements and eye tracking system were synchronized with the other data sources, e.g., driving data and the self-reported sleepiness. The analyses in this section are done based on average 5-minute values for the sleepiness indicators KSS, blink duration, Percentage Eye Closure (PERCLOS) and PVT mean reaction time and PVT lapses. A mixed model ANOVA was used with participant as a random factor and distance driven, gender and type of vehicle as fixed factors. In this report we also evaluate the driver's opinion about driving with and without L2 activation. Differences between drivers were evaluated using t-tests or Chi-square tests. Pre-processing was done in MATLAB R2020a (Mathworks Inc., Natick, US). All statistical analysis was conducted using IBM SPSS 26.0 statistical software (IBM Corp., Armonk, NY, USA). An alpha level of 0.05 was used to determine statistical significance.

3.1.4 Results

Questionnaire results

After each drive, the drivers were asked to fill in a questionnaire. The drivers were asked to rate, on a scale from 1 = not at all to 7 = very, how they experienced the drive, in terms of:

- how **demanding** it was to stay awake while driving.
- how **stressful** the drive was?
- how **anxious** they were during the drive.

There was no difference in how demanding it was to stay awake between the first and second visit to VTI. Neither was there a difference between manual or level 2 automated driving. As expected, it was much more demanding to stay awake in the night-time drive compared to the daytime drive (Figure 15). The drivers did not find the driving to be stressful at any time (Figure 16) and in general they did not feel anxious during the drives (Figure 17). However, they did report slightly higher anxiousness levels during night-time driving, regardless if they were driving with or without automation.

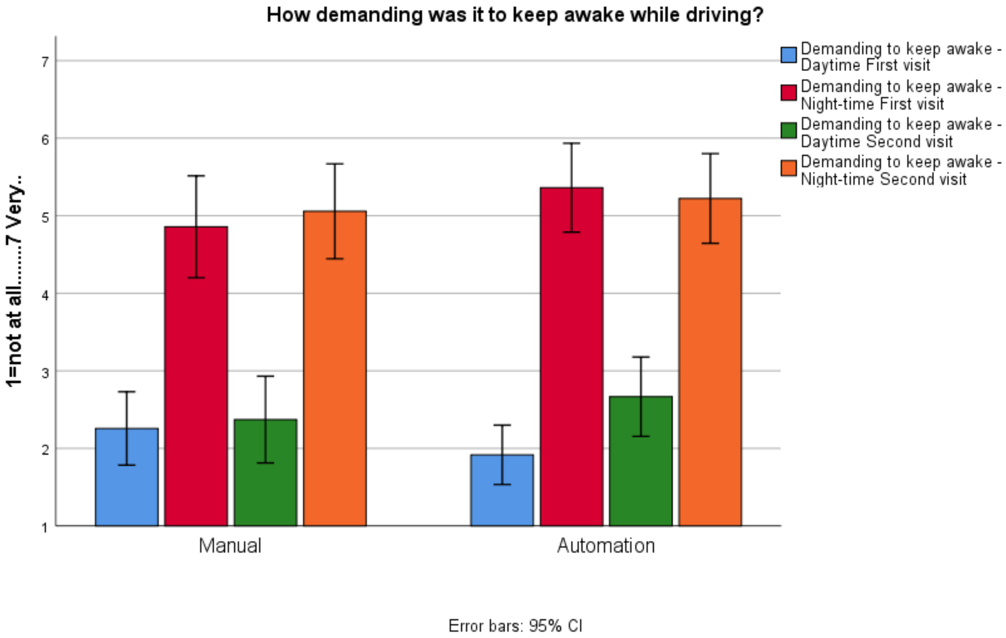


Figure 15. Did you find it demanding to keep awake while driving? (1=not at all....7= very demanding.).

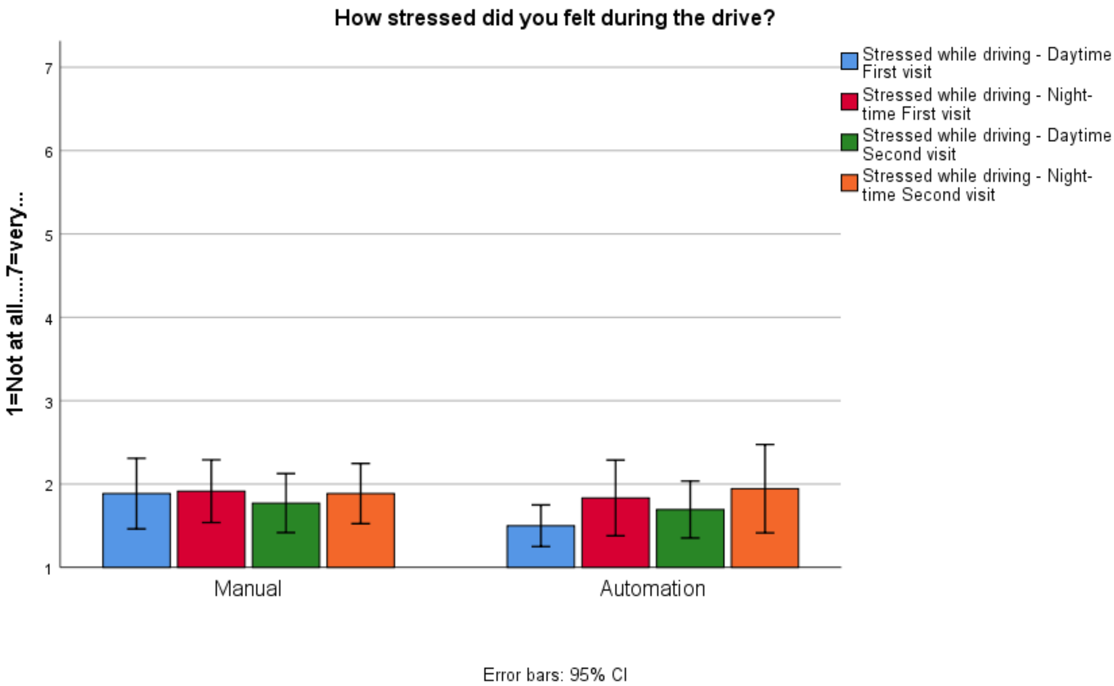


Figure 16 Did you felt stressed while driving? (1=not at all....7=very stressful).

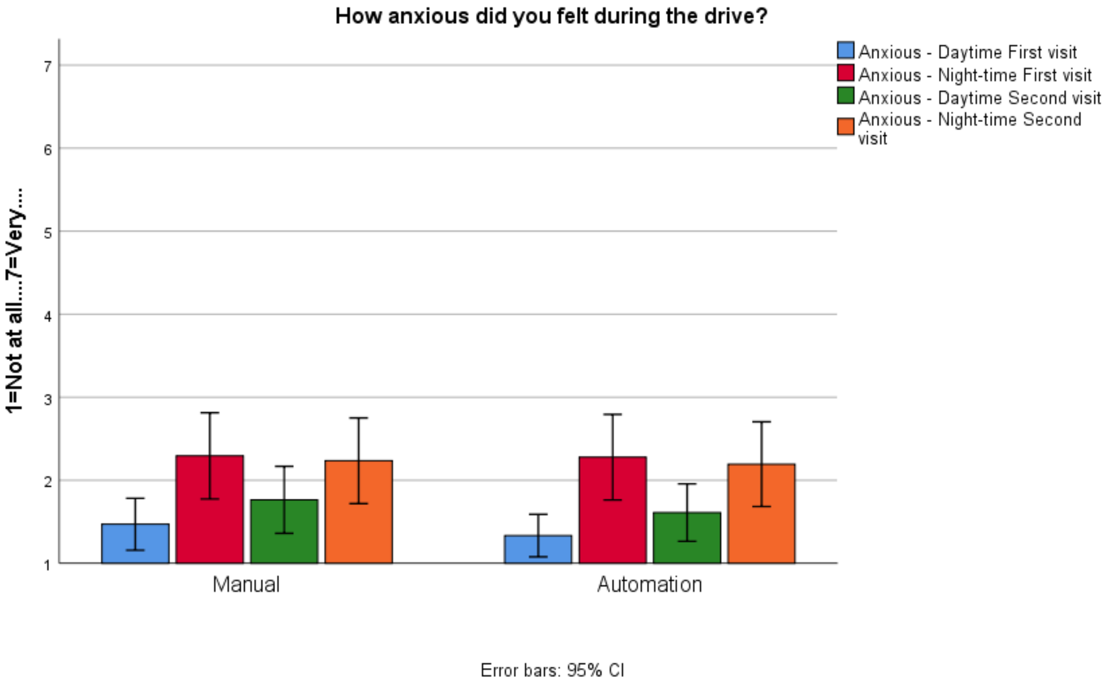


Figure 17. Did you felt anxious while driving? (1=not at all....7=very…).

Behavioural markers

On a group level, the subjective KSS ratings showed a difference in sleepiness during daytime and night-time driving ($F_{1,4361}$=231.7, p<0.001), Figure 18. There is a difference in sleepiness development (time on task) with and without automation ($F_{14,4361}$=71.0, p<0.001) for day and night-time driving ($F_{14,4361}$=24.1, p<0.001).

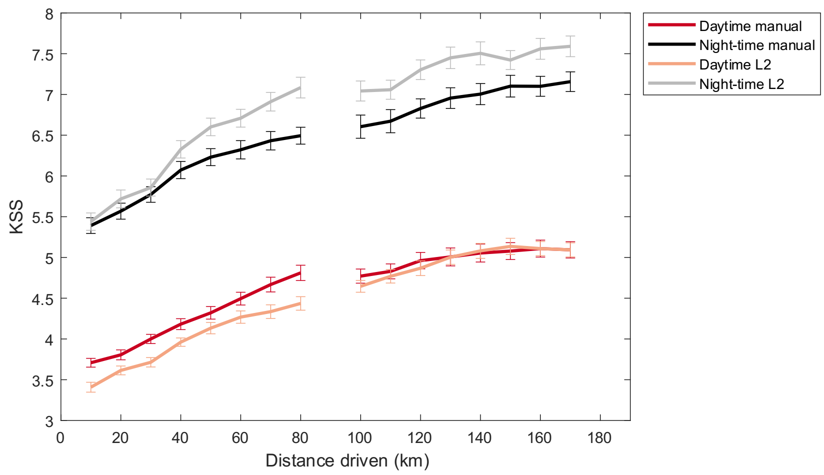


Figure 18. KSS development during driving at night and daytime, with and without pilot assist activated.

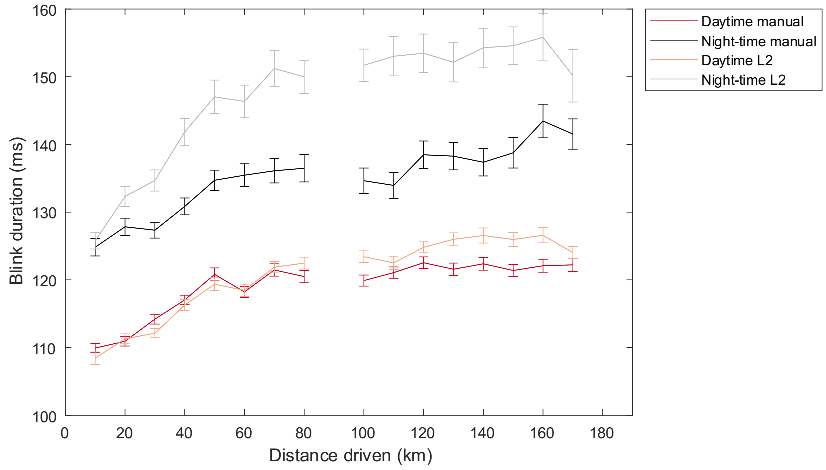


Figure 19. Blink duration development during daytime and night-time with and without pilot assist activated

Blink duration showed higher levels of sleepiness during night-time ($F_{1,4029}$=52.9, p<0.001) and when driving with automation activated ($F_{1,4029}$=41.7, p<0.001). As can be seen in **Fout! Verwijzingsbron niet gevonden.**, these higher levels of sleepiness during level 2 automated driving mainly originate from the night-time drives ($F_{1,4029}$=60.5, p<0.001). During daytime driving, there was no difference between automated and manual driving in the beginning of the drive, but after half of the drive, sleepiness increased more when level 2 automation was activated ($F_{14,4029}$=61.4, p<0.001).

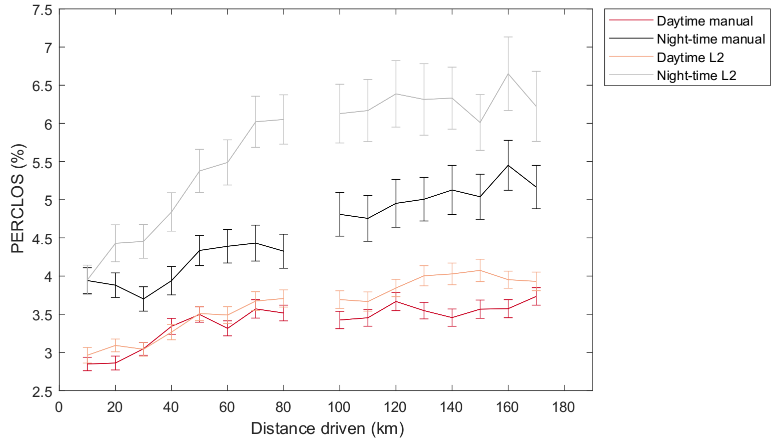


Figure 20 PERCLOS development during daytime and night-time with and without pilot assist activated.

The only significant main effect of PERCLOS was an increase due during night-time driving ($F_{1,3989}$=11.0, p<0.001), **Fout! Ongeldige bladwijzerverwijzing.**. The interaction effects showed increased PERCLOS levels during level 2 automated driving at night-time drives ($F_{1,3989}$=10.6, p<0.001). During daytime driving, there was no difference between automated and manual driving in the beginning of the drive, but after half of the drive, sleepiness increased more when level 2 automation was activated ($F_{14,3989}$=14.4, p<0.001).

Heart rate was slower (i.e., longer interbeat intervals) during the night-time drive ($F_{1,3651}$=24.0, p<0.001), with L2 activated ($F_{1,3651}$=28.6, p<0.001) and with time on task ($F_{15,3651}$=42.5, p<0.001), see Figure 21.

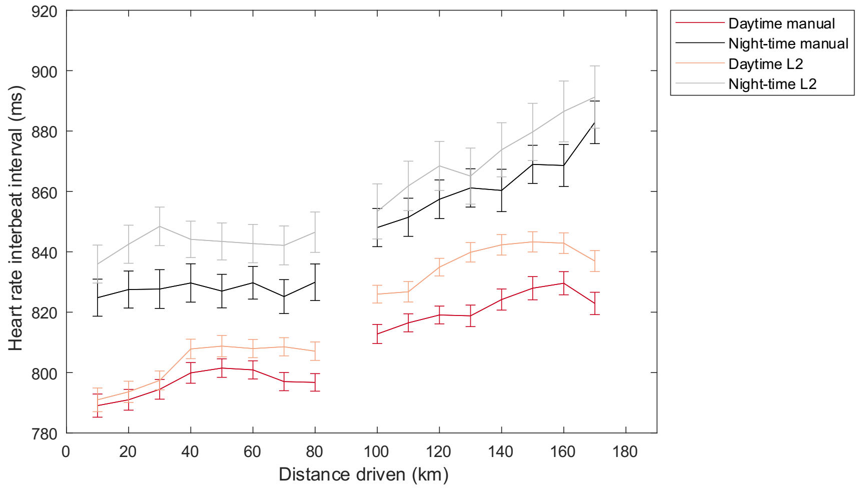


Figure 21. Heart rate interbeat interval development during daytime and night-time with and without pilot assist activated.

Psychomotor vigilance task

Higher mean reaction times (ms) were found in the PVT tests performed after the drive ($F_{1,557}$=80.5, p<0.001), with a tendency for shorter mean reaction times after driving with automation activated ($F_{1,557}$=9.3, p=0.002), see Figure 22. There were no differences in mean reaction time between manual versus level 2 automated driving in the test performed before the drives. Concerning PVT lapses, defined as reaction times longer than 500 ms, the tendencies were the same, with fewer lapses after driving with level 2 automation during night-time ($F_{1,557}$=1.4, p=0.23), see Figure 23.

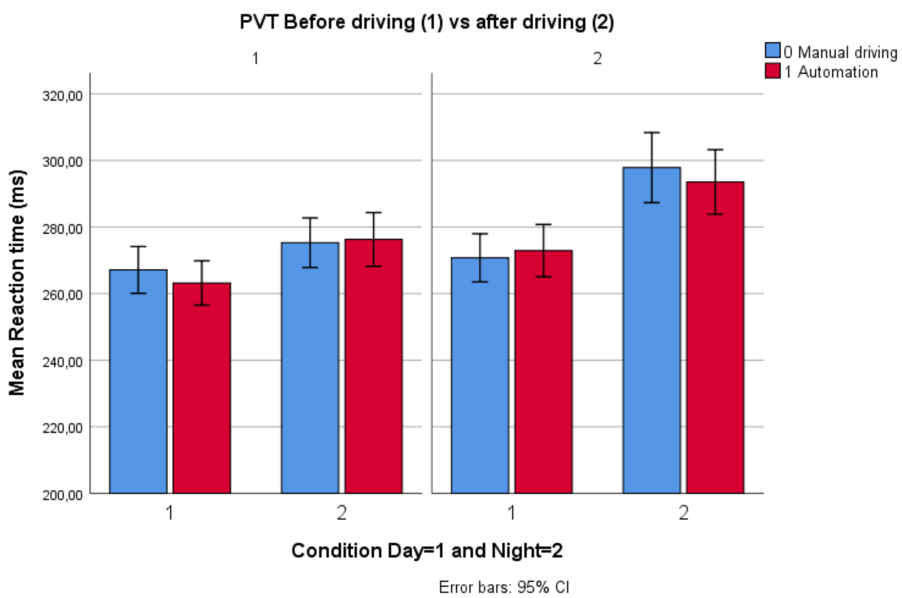


Figure 22. PVT expressed as mean reaction time (ms)

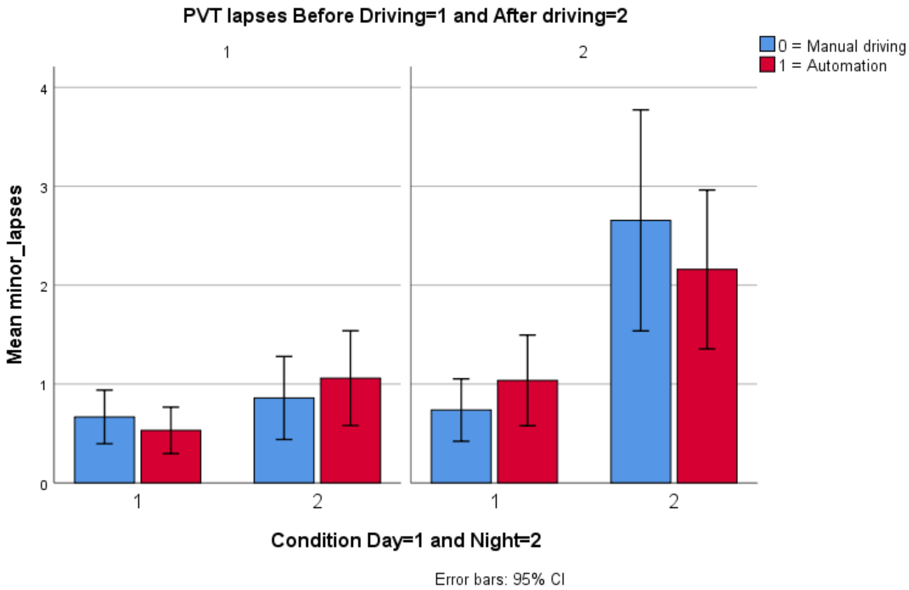


Figure 23. PVT expressed as minor lapses (reaction time > 500 ms).

Preliminary results based on real-time analysis of video data

The group-level results presented above represent highly averaged data where intra- as well as inter-individual differences and contextual confounding factors are smoothed out. By smoothing out the differences, there is a risk that a classifier based on these fatigue indicators may work well on a group level, but not for a certain individual in a certain situation at a certain time (cf. Golz, Sommer, Trutschel, Sirois, & Edward, 2010). It is therefore important to investigate how sleepiness indicators, or detection algorithms, generalise to new individuals, and importantly, also to data from the same person but on another day.

Cygnify, in collaboration with VTI, has previously developed a personalised sleepiness detection algorithm based on a two-stage model with (1) a generic deep feature extraction module combined with (2) a personalised sleepiness detection module, see Figure 24.

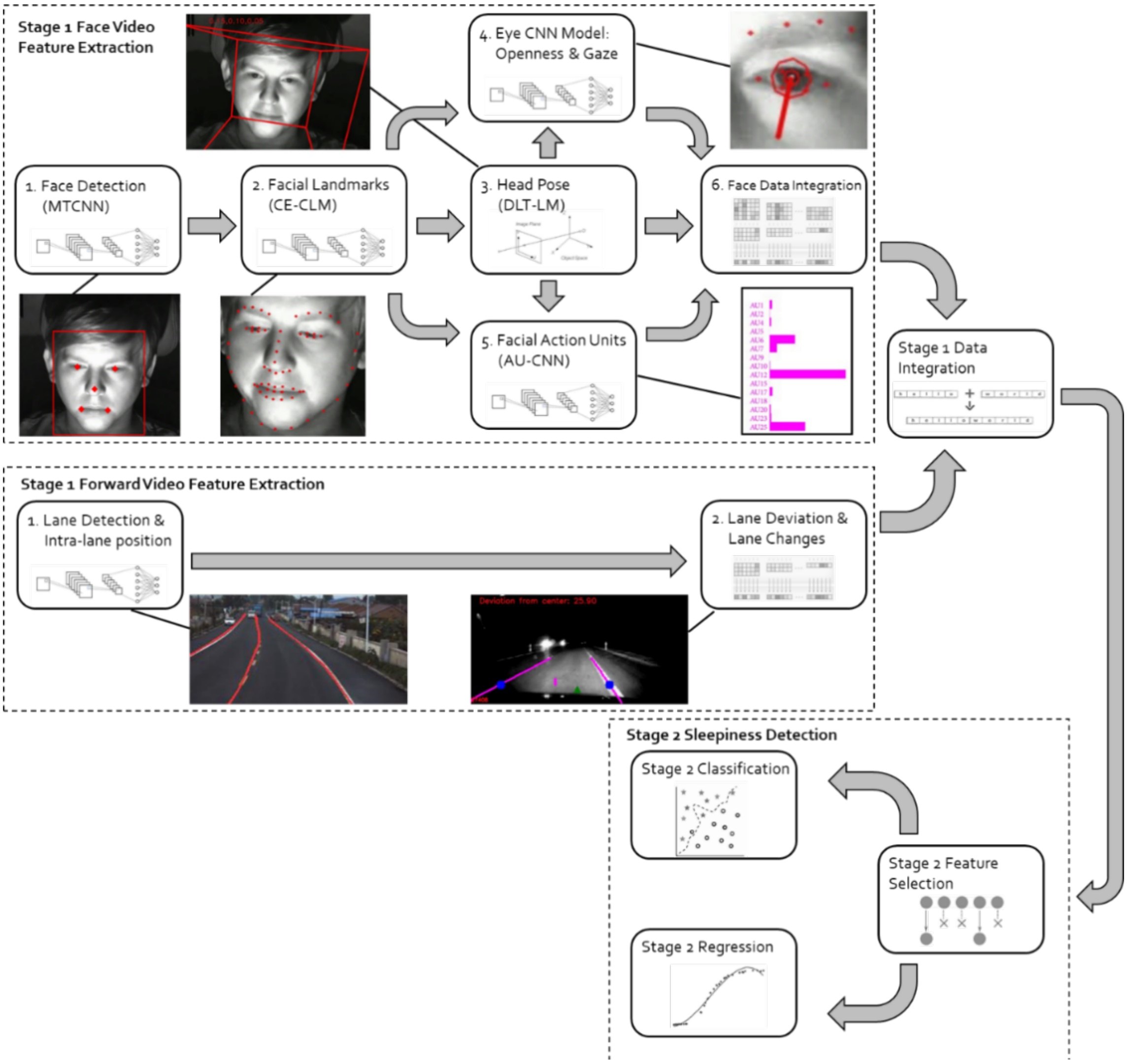


Figure 24 The complete architecture consisting of Stage 1 (feature extraction) and Stage 2 (sleepiness detection), each with multiple subcomponents.

As shown in Figure 24, the sleepiness detection method can operate in two modes: binary classification ("sleepy" vs. "alert"); and in continuous-output regression-like mode, estimating directly the KSS value for the driver. Both modes have already been attempted in our preliminary experiments. For binary classification we use the criterion that the "sleepy" class corresponds to KSS>=7, and "alert" corresponds to KSS<7.

In both modes, the detection system operates by considering the past 5 minutes of data, and at the end the 5 minutes, it produces the fatigue estimate in real-time. The stage 1 system extracts, from those 5 most recent minutes of data, a fixed size deep feature vector, combining the outputs of multiple internal deep networks. This feature vector captures information concerning the overall statistics as well as unusual events related to eyes opening and closing, blink frequencies and eye closure times, facial expressiveness, gaze patterns, face behaviours like yawning and head nodding, and—optionally—lane changes and lane departures and other information derived from forward video and/or vehicle sensors.

The feature vector is used as input to the second stage system, which performs the actual classification or non-linear regression. This second stage is trained using one part of the dataset, the train dataset, using the driver reported KSS scores as the ground truth ‘target’ datapoints (adapted for either binary classification, see above, or continuous regression-like estimation). This training process is personalised, leading to personalised models fine-tuned for individual drivers. At test time, another part of the dataset, the test set, which was held apart from the model training process, is then used to evaluate how well the system does on such previously unseen data.

In the current set of experiments, we focus on the head, face, and eye patterns extracted from the face videos. Thus, for now, we exclude the forward-looking video in the experiments, as well as vehicle data such as longitudinal and lateral acceleration, steering wheel movements, etc. The reason is that the use of such data is mainly for extracting vehicle behaviour and driver corrective behaviour data (lane change, lane centre variability, and lane departure data; corrective, harsh steering wheel movements and braking, etc.), which is not informative for fatigue estimation when L2 automation takes care of lateral and longitudinal vehicle control.

From the data of 43 drivers available to Cygnify (the XC90 group—see the description of what data was shared above), for 25 drivers there was complete (2 days and 2 nights) face video and KSS data. Out of those, 19 drivers had both low fatigue (“alert”, KSS<7) and high fatigue (“sleepy”, KSS>=7) KSS scores on both the first visit (day 1 and night 1) and the second visit (day 2 and night 2). In the preliminary analysis, we have focused on those 19 drivers, as they are the most ‘interesting’ ones in terms of both completeness of data and the challenge of accurately detecting both low fatigue and high fatigue.

For those 19 drivers, we considered two different train-test splits. First, we used a relatively ‘easy’ train-test split, where we randomly set 40% of the drivers’ 5-minutes episodes from both days and nights as test episodes. Using that split, we get 92.5% binary classification accuracy on the out-of-sample test data. This means that only 7.5% of the episodes are misclassified, i.e. are either “false alarms” (false positives, indicating fatigue when the driver says (s)he is not fatigued) or “misses” (false negatives, indicating no fatigue when the driver says (s)he is not fatigued). We get a Mean Absolute Error (MAE) for the continuous-output regression-like direct estimate of the KSS of 0.445. This means that the estimated KSS is, on average, less than half a unit away from the true reported KSS score; even though there are occasionally larger deviations.

A risk of this approach, using this random train-test split, is that performance is overestimated because the machine learning model may exploit too many similarities between ‘neighbouring’ episodes from the same drive. In other words, there are both train and test episodes within one drive, often directly adjacent to each other. Typically, these neighbouring episodes have similar KSS scores because fatigue normally does not change dramatically from one 5-minute episode to the next. This means that some of those KSS scores are ‘known’ to the training system. Thus, the machine learning model might learn to ‘simply’ match test data to the most (superficially?) similar episode in the training data, which may be from an immediately neighbouring (train) episode (even though the machine learning detection system does not explicitly know the drive ID or the episode number or order).

For this reason, we also experimented with another train-test split, the “visit 1/visit 2” train-test split, where data from the first day and first night is used as train data and data from the second day and the second night are used as test data. Binary classification accuracy with this “visit 1/visit 2” train-test split was 89.7%; and MAE was 0.734. The preliminary conclusions from this is that the random train-test split leads to some data leakage. Otherwise, the results would not be worse for the "visit 1/visit 2" train-test split. However, this effect is not very prominent.

Preliminary conclusions are that the approach we are working on is promising and will serve the goals of the MEDIATOR project sufficiently well. The approach is also a good basis for next steps in WP2. These future developments will include looking at generalising to completely new, unseen drivers; forecasting KSS development over time in a drive; and combining this facial video-based approach with the physiological data approach outlined below.

Preliminary results based on real-time analysis of physiological data
Autoliv has implemented a machine learning based analytics factory with the goal of facilitating data analysis and algorithm development. Analysis stages include systematic review, cleaning and pruning of data, machine learning design and training as well as development of runnable algorithms for deployment. The main goal is to develop, based on the current/past state of different variables, an algorithm for real-time prediction of driver sleepiness.

The state of the driver is characterized by the following data:

- Bio sensors: ECG, HRV, respiratory measures
- CAN and map data (future use)
- Historical data on sleep and activity (future use)
- Visual data (future use)

KSS is used as the target label in a supervised regression model based on machine learning algorithms such as random forest, XGBoost, MLP (Soon), Model stacking, etc.

Five minutes of data prior to each KSS rating is used to extract driver sleepiness features. In this way a sample is generated in (processed data) for each 5 min window and matched with the KSS for that window. In the present setup, we use fixed time windowing, but this will be extended to a sliding window approach in WP2. Extracted features include temporal, statistical and frequency domain features, as well as domain specific features such as various heart rate variability metrics, from both the chest ECG and the seatbelt respiration sensor. All in all, about 850 features are considered.

Out-of-fold validation (worst case scenarios where the trained model is doing inference on unseen drivers) shows MAE performance of around 1.0 with good bounds for confidence level, and a normalized MAE of 20%. In the case where the model has seen all the drivers during training, the MAE performance was around 0.5.

Next steps include testing additional machine learning algorithms and continued fine-tuning of hyper-parameters, including model pooling. Further work will also be devoted to improving the feature extraction algorithms and to merge the physiological data approach with the video-based approach described above.

3.1.5 Summary and conclusions

Driving with partial automation leads to higher levels of sleepiness, especially during night-time driving when the sleep pressure is high. During daytime, when the drivers in the study were alert, partially automated driving had little or no detrimental effects on driver fatigue (compared to manual driving).

Preliminary fatigue detection algorithms, based on video data and physiological data, respectively, have been developed based on the data set. The results are encouraging, showing a MAE of about 0.5 KSS units, or a binary classification accuracy of about 90%.

3.2 Evaluating the effects of task induced fatigue during L2 driving with and without a mitigating secondary task on hazard perception ability: A driving simulator study

3.2.1 Background

3.2.1.1 Fatigue in manual driving

Fatigue is a transitional state between awake and sleep and if continuous, it may lead to sleep. It is characterised by reduced efficiency, lowered alertness level, memory impairment, longer reaction times and a general unwillingness to work (Lal & Craig, 2001). Fatigue is classified into three types: sleep-related fatigue, active task-related fatigue and passive task-related fatigue. Active task-related fatigue occurs in situations of poor visibility, high traffic density or secondary task engagement. Sleep related is a result of sleep deficiency and related to the circadian rhythm. Passive fatigue can be seen as the opposite of active fatigue. Task underload, monotony and automated systems are seen as the reasons for vigilance decrement (May & Baldwin, 2009). Fatigue is becoming a leading cause for crashes. Worldwide, it is estimated that between 20% to 30% of road crashes are due to fatigue (Vogelpohl, Kühn, Hummel, & Vollrath, 2019). Already a long time ago it was established that manual driving for prolonged periods of time leads to fatigue. In a simulator study conducted by Dureman & Bodén (1972), participants had to manually keep a small car on the road and react to intermittent auditory stimuli. Results indicated an increase in reaction times and a deterioration in the tracking task. Besides performance decrement participants reported increased feelings of fatigue (Dureman & Bodén, 1972). Objective indicators of fatigue were also used for quantifying mental fatigue. Power spectrum of EEG signals indicated an increase in alpha band in a 15 minutes simulated drive which could be interpreted as a dip in attention due to fatigue onset (Balasubramanian, Adalarasu, & Gupta, 2011). Heart Rate Variability is also often used to evaluate levels of task induced fatigue. A recent literature overview of this measure suggests that HRV is often increased due to drowsiness and a lack of engagement in the driving task during automated mode (Lohani et al.,2019).

3.2.1.2 Fatigue in Level 2 automation

Unlike manual driving, where the driver is the sole responsible for both the driving and monitoring tasks, in L2 the automation is able to provide longitudinal and lateral control of the vehicle in restricted situations. However, the driver is expected to monitor the environment at all times and may be required to proactively take control under unannounced circumstances depending on the driving context.

Paradoxically, partial automation induces passive fatigue due to underload and monotony, rendering the driver with a loss in attentional resources and performance decrement (Körber, Cingel, Zimmermann, & Bengler, 2015). Schmidt and colleagues studied the effect of a 3-hour monotonous highway drive on driver's vigilance and drowsiness. Results showed an increase in KSS scores as a subjective indicator of fatigue, as well as an increase in reaction time and vigilance decrement (Schmidt et al., 2009). Not only does automation induce passive fatigue, due to lack of involvement, drivers under partial automation become fatigued faster than under manual driving (Schömig et al., 2015). Feldhütter et al. found indicators for the development of fatigue during an automated drive of only 20 minutes based on eye-tracking measurements compared to fatigue signs that appeared after 45–50 minutes of manual driving (Feldhütter et al., 2016).

3.2.1.3 Hazard Perception in Level 2 automation

Hazard perception is a skill with a high cognitive component. As such, general attention and vigilance factors are significant contributors to performance on hazard perception tests (Smith, Horswill, Wetton, & Chambers, 2009). It is common to measure HP visual drivers' eye movements variables (e.g., fixations duration on areas of interests such as mirrors) alongside with drivers' response time to actual or potential road hazards. Hazard perception ability in L2 driving is impaired two-fold. One by the automation itself and second by fatigue. The underload imposed on drivers may encourage them to engage in tasks that are unrelated to the driving task. In their work, Körber et al. showed that participants reported more engagement with their own thoughts in the last five minutes of the experimental drive compared to the first five minutes. This is in line with our expectation that because of the low task demand, participants' thoughts and attention drift away from performing the driving task to engage with their own thoughts (Körber et al., 2015). During fatigued-driving in vehicles with high levels of automation, knowledge-based and rule-based behaviours are seriously affected. When the driver is required to take control due to a new road situation, he must reach a high level of consciousness in order to properly handle and process the load of new information. Fatigue will cause a slower response time to the new situation and hinder the driver from applying the right rule and actions (Mccoy, 2009).

Saxby and colleagues conducted a driving simulator study and demonstrated situational awareness decrement under L2 partially automated driving conditions. One-hundred and seventy participants had to make a transition between an inattentive state during partially automated driving conditions to control reclaim triggered by a road hazard. Results suggested that the likelihood of colliding with a roadway hazard was greater after driving for 30-minutes under partially automated conditions than after 10 minutes of partially automated driving. These findings, according to the authors, indicate the loss of situation awareness during partially automated driving that may eventually compromise HP (Saxby, Matthews, Warm, Hitchcock, & Neubauer, 2013).

3.2.1.4 NDRT under partially automated driving (L2) conditions

Studies have shown that one of level 2 automation undesired effects is the proneness of the drivers to engage in non-driving related tasks (Solís-Marcos, Ahlström, & Kircher, 2018). However, little is known about the deliberate incorporation of a secondary task as a countermeasure of passive fatigue in L2. Though the primary countermeasure for fatigue is rest, the countermeasure for passive fatigue is monotony rupture, which is not necessarily best achieved through a rest break. If fatigue is due to underload, then activation, such as playing a trivia game or eating sunflowers seeds, rather than rest may help (Gershon, Ronen, Oron-Gilad, & Shinar, 2009). A different countermeasure for fatigue was presented in the paper of May and Baldwin who reviewed several existing fatigue countermeasure methods. One of these methods is an algorithm that calculates the proportion of time within 1 min blocks that the eyelid covers 80% of the pupil. If fatigue level reaches a predefined limit, the monitor gives an auditory alert for the driver and visually presents the PERCLOS duration and distance travelled during the PERCLOS. This system is currently used by commercial truckers (May & Baldwin, 2009). A study conducted by Jarosch et. al., using a motion-based driving simulator at the BMW Group laboratories was aimed to investigate the effects of two different NDRTs in conditional automated driving (CAD) on drivers' fatigue and takeover performance. NDRT was either a monotonous monitoring task or an activating quiz task. Results show that participants who had to engage in the monitoring task showed higher PERCLOS measures compared with the participants who had to engage in the activating quiz task (Jarosch, Bellem, & Bengler, 2019). Positive effects of a quiz task to counter- measure fatigue in automated driving were reported. In another related study, participants indicated significantly lower levels of fatigue when they had to deal with a quiz task for 15 minutes (Schömig et al., 2015).

3.2.2 Research hypotheses

Based on the literature overview the following hypotheses were elicited

- H1. Partially automated driving conditions will be more likely to induce fatigue than manual driving. This hypothesis will be confirmed if:
  - KSS scores under L2 driving conditions will be higher than under manual driving conditions
  - Heart Rate Variability under L2 driving conditions will be higher than under manual driving conditions
- H2. Engagement in a secondary task (i.e., Simon) during partially automated driving (level 2) will counteract the effects of fatigue levels. This hypothesis will be confirmed if:
  - Driving under L2 driving conditions in the presence of a secondary task (i.e., Simon) will lead to lower KSS scores than driving under L2 driving conditions in the absence of a secondary task.
  - Driving under L2 driving conditions in the presence of a secondary task (i.e., Simon) will lead to lower HRV values than driving under L2 driving conditions in the absence of a secondary task.
- H3. Hazard perception performance will deteriorate during L2 driving but this deterioration will be mitigated by engagement with a secondary task. This hypothesis will be confirmed if:
  - Drivers will be less likely to glance to hazard under L2 conditions than in manual conditions.
  - Drivers will be more likely to glance to hazard under L2 driving conditions when a secondary task is presented than when a secondary task is absent.

3.2.3 Method

3.2.3.1 Participants

Thirty-two participants, 13 females (mean age=25.8, SD=1.53 years) and 19 males (mean age = 25.8, SD=1.56) participated in the study and were randomly assigned into one of four experimental groups (n=8 for each group). All participants were students from Ben- Gurion University of the Negev (BGU) and participated as paid volunteers. They received 150 NIS only after completion of the experiment. All participants were naive to the study's hypotheses, right-handed, had normal or corrected to normal visual acuity (9/6 or better), normal contrast sensitivity, normal colour vision and no background of heart problems. The participants had a valid driver license for at least 5 years and reported to have no experience with a driving assistance system that provides both lateral and longitudinal support.

3.2.3.2 Materials and apparatus

The driving simulator, the automated virtual vehicle, the eye tracker and the Biopac devices were the same as the ones described in Section 2.1.4.2.

<u>Driving environment and hazard perception scenarios</u>
The simulated drive included a trip of approximately 40 min. The virtual drive included both highways environments and urban environments. The roads in the urban environment always included two lanes in each travel direction. Six unmaterialized hazards and one materialized hazard were located along the drive. The seven hazards were distributed along the drive with an average time interval of 4.5 minutes between two consecutive hazardous scenarios. The materialized hazard scenario always appeared as the seventh scenario at the end of the drive. Six different combinations of the six

hazardous scenarios including the seventh hazard at the end were built and were randomly assigned between participants.

Fout! Verwijzingsbron niet gevonden. describes in detail the seven hazardous scenarios that were designed for the study. The "cue" column describes the hazard related visual cues that preceded each hazardous situation and indicated to the participant that a hazard is likely to occur in the near future. For each of the seven hazards a time window was defined during which the participant should identify the hazard in order to avoid a potential crash. The start of the time window is called a launch zone, a predefined area along the route where the hazard becomes visible and the driver must begin glancing toward the target zone. The target zone is a visual area of interest (AOI) of a latent hazard (Krishnan et al., 2019). The time window ended when detection of the hazard would have been too late to avert a crash (Vlakveld et al., 2018). The duration of the seven time-windows ranged from 9 to 38 seconds.

<u>The secondary Task: Simon game</u>

The interface of the Simon game that was used as a secondary task consisted of four squares of different colours (*Figure 25*). For each trial of the game, the interface showed the participant a sequence of coloured squares where each coloured square was associated with a distinct tone. The participant had to repeat the sequence by touching the coloured squares in the correct order. In the first and easiest level, the sequence included one colour. Each time the participant repeated the sequence correctly (i.e. a successful trial), the next trial was presented. The following trial included an additional colour that was added to the sequence and was thus harder than the preceding trial. For example, the second level included a sequence of two-coloured squares, the third level a sequence of three coloured squares and so on. If the participant made an error or had a delayed response larger than 2 seconds, the game was restarted at the first level. Each successful trial entitled the participant with points corresponding to the sequence's length at that level (e.g. 3 points for a sequence of 3 colours). Points were accumulated during the whole experiment to give final score. Drivers' goal was to have as many successful trials as possible. Points were accumulated per trial independent of the number of game restarts. The opportunities to play the Simon game were triggered after 9, 24 and 37 minutes from the beginning of the experiment, for 2 minutes each time. These playing opportunities were identical for all participants under this condition. Participants were motivated to do so by indicating that the three drivers with the highest scores will win a monetary prize of an additional 50 NIS.

Figure 25 The in-vehicle interface including the Simon game

Questionnaires
Participants were asked to complete four different questionnaires: (1) a demographic questionnaire that included 9 questions about the participant (e.g., age, driving experience, gender). (2) Multidimensional Driving Style Inventory questionnaire (MDSI) (Taubman-Ben-Ari et al., 2004) that included 44 questions that are related to the driver's driving style. Each question contained a 6-points Likert scale where "1" represents no agreement and "6" represents full agreement. This questionnaire distinguishes between four driving styles: Risky and irresponsible, Anxious, Hostile and aggressive, Careful and tolerant. (3) User's adoption and trust aspects of autonomous vehicles questionnaire. It includes 12 questions with a 7-point Likert scale (Choi and Ji, 2015). (4) Karolinska sleepiness scale (KSS) (Shahid et al., 2011). KSS includes 9 phrases describing fatigue level. The participant had to choose the phrase that best describes his current level of fatigue. Upon completion of each experimental drive (2 drives per driver), participants were asked to complete the Karolinska sleepiness scale (KSS) questionnaire for the second time, to asses post-experiment fatigue. After each experimental drive, participants also completed the (5) NASA-TLX (Hart and Staveland, 1988) questionnaire to measure their perceived mental workload in each driving condition. All the questionnaires were displayed on a computer and were created by using Google Forms application. Analysis of the questionnaires was done by a simple average over the scores, except for the MDSI which was analysed by weighed scores.

3.2.3.3 Experimental design

The experimental design included two arrays of 2*2 mixed design. Participants were divided into 2 main groups of 16 participants each. The first group experienced two driving modes in two separate experimental drives: manual and Level 2 automation. None of these modes included a secondary task. The second group experienced Level 2 automation only, where a secondary task was included in one of the two experimental drives. Two independent variables were designed for group 1 and included: (1) driving mode (within subjects), (2) order of drives (between subjects). Similarly, two independent variables were designed for group 2 and included: (1) existence of a secondary task (within subjects), (2) order of drives. Each group experienced a 40-minute drive twice (there was a one-week break between the drives). The experimental design is summarized in Table 6.

Table 6 Summary of experimental conditions

Group number	Drive 1	Drive 2
1(a)	Manual	L2 automation without a secondary task
1(b)	L2 automation without a secondary task	Manual
2(a)	L2 automation without a secondary task	L2 automation with a secondary task
2(b)	L2 automation with a secondary task	L2 automation without a secondary task

The independent variables were the level of automation (L2, manual), secondary task existence (yes/no), drive number (drive 1, drive 2), and scenario's chronological order (between 1 and 7) as it appeared in each experimental drive. The dependent variables are summarized in Table 7.

Table 7 Description of the dependent variables that were analysed

Measure	Purpose	Transformation of measure	Name of variable
RMSSD	Measurement of heart rate variability	$\log\left(\frac{RecordedRMSSD}{RecordedRMSSDRestBaseline}\right)$	LogQuotientHRV
KSS score	Subjective measure of fatigue level	(KSS score at the end of the drive)-(KSS score at the beginning of the drive)	DeltaKSS
Number of glances (>100 msec) on AOI	HP quantification	No transformation	GlanceCount

3.2.3.4 Procedure

Participants were first requested to digitally sign an online informed consent form. Then, each participant completed online the MDSI, demographic and trust in automated vehicles technology questionnaires before they arrived at the lab. Upon their arrival at the HPEL, each participant underwent a visual acuity test (Snellen Chart) and a functional acuity contrast test (FACT; Ginsburg, 2002). Those who qualified received a written explanation about the study and the simulator capabilities in accordance with the driving condition they were assigned to. Participants were instructed to follow the Israeli traffic laws and drive as they would normally do in similar real-world situations. Prior to the experimental condition including a partially autonomous driving conditions, participants were given an explanation about the PAV and its capabilities (adaptive cruise control and lane keeping system). All participants were informed that they are fully responsible of the driving task and of the safety of other road users and themselves at all times even when the automation is active (i.e. they should keep monitoring the driving environment all the times). Deactivating the PAV driving mode and shifting to manual driving was possible at all times by pressing the brake pedal. Returning to partially autonomous driving mode was done by pressing a designated button inside the vehicle. Prior to the experimental condition that included a secondary task, participants were taught the rules of the game and could practice it without driving for several minutes until they felt comfortable with playing the game. Next, they answered a short quiz about the automation capabilities to see if they understood how it works. Then, the ECG electrodes were positioned on the participants' chest and they were given an IKEA magazine to browse while they were resting. Their heart rate and R-R interval baseline was measured for 7 minutes at rest.

Next, the participant drove a 4-minute training drive to become familiarized with the simulated driving environment. During the training session the participant's heart rate and R-R intervals data were recorded and served as the participant's heart rate at baseline. In this training session, participants experienced driving on straight roads, curves, intersections, and a transition from a two-lane road to a four-lane road. The environment characteristics of the training drive matched the environmental characteristics of the experimental driving conditions (manual driving conditions/ L2 with a secondary task/ L2 without a secondary task).

After training, the participant was asked to wear the eye-tracker glasses and his gaze position was calibrated. Then, the first experimental drive began depending on the experimental condition see Table 6. At the end of the experimental drive the participant was able to take a short break and then to complete the NASA-TLX and KSS questionnaires. A week later each participant came back to the lab approximately at the same hour as in the first session, for the second session where he or she underwent the second experimental drive. On their second session, participants underwent the same procedure as in session one except for the seven minutes rest period (the baseline at rest was only measured once in session 1). Each session lasted for an hour. At the end of the two sessions, the participant was thanked and received monetary compensation (150 NIS) for his/her participation.

3.2.4 Results

3.2.4.1 Statistical analysis

Prior to the statistical analyses all data were merged and synchronized. All statistical analyses were carried out at a significance level of 5 percent. For dependent variables that are normally or log-normally distributed we used a linear regression model within the Linear Mixed Models (LMM) framework. Participants and scenario number were included as random effects. Specifically, the LMM regression was applied on the variables DeltaKSS and logQuotientHRV. Next, for the hazard identification variable that is binary distributed we used a logistic regression model with a logit link function within the Generalized Linear Mixed Models (GLMM) framework. Finally, for the number of glances at the hazard variable that is Poisson distributed we used a Poisson regression with a log link function within the GLMM framework. The initial model in each analysis included the fixed effects of scenario's chronological order, automation level, drive number, task existence (yes/no), the second order interaction between automation level and scenario's chronological order, the second order interaction between drive number and task existence and the third order interaction between automation level, scenario's chronological order and drive number as fixed effects. The final model of each analysis was achieved via backwards elimination where all non-significant fixed effects interactions were removed from the model. For significant fixed effects with more than two levels, post hoc pair-wise contrasts comparisons analysis was applied and the Tukey HSD/ LSD procedure was used to correct for multiple comparisons (all statistical analysis computed in R software). Finally, the analysis is based on the data of all 32 participants. However, due to some technical issues the data for some scenarios were omitted from analysis. The following Table 8 summarizes missing data.

Table 8. Summary of cases excluded from the analyses and their associated reasons

Participant	Missing scenario data	Drive	Reason for exclusion
7, 24, 29, 31, 39, 43, 45	traffic	2	Eye-tracker came out of calibration
12	first drive missing	1	Simulator dynamics output missing
29	gas station	2	Gaze not detected
31	traffic	1	Eye-tracker came out of calibration
45	blinking car	2	Power shortage
28	crosswalk 2	1	Technical error

3.2.4.2 Levels of fatigue induction

There were two direct measures that were aimed at evaluating how manual driving and L2 driving (with or without engaging with a secondary task) contribute to the development of fatigue, namely KSS scores analysis and HRV analysis. The analysis of these two measures are presented next.

<u>KSS scores analysis</u>

In this analysis the dependent variable was the KSS score difference between the beginning and end of each drive per participant. The final model included 2 significant main effects of automation level (X_1^2=39.2802, p <0.05) and secondary task (X_2^2=18.4538, p <0.05), and one significant interaction between task existence and drive number (X_2^2=13.1254, p <0.05). First, with respect to the main effect of automation level, under the condition of manual driving, participants reported on average significantly lower KSS scores (Estimated Mean =1.9, SE=0.312) than under the condition of Level 2 driving (EM =2.73, SE=0.293). Second, with respect to the main effect of task existence, in the absence of a secondary task L2 driving conditions, participants reported on average significantly higher KSS scores (EM =2.55, SE=0.407) than under the L2 driving conditions when the secondary task was present (EM =1.99, SE=0.41). Post hoc analysis of the significant interaction did not yield significant differences and thus it is not discussed further.

Heart rate variability analysis

The dependent measure that was used to quantify heart rate variability analysis is RMSSD (see Table 7 for the computational expression). The final model included two significant main effects: task existence (X_2^2=10.061, p <0.05) and the event chronological order (X_6^2=18.658, p <0.05). With respect to the main effect of the chronological order of the scenarios, the only significant difference (p_{adj}=0.0069) was between the estimated RMSSD mean of event 1 (EM=-0.42, SD=0.153) and event 6 (EM=-0.005, SD=0.153). Driving under L2 driving conditions in the absence of a secondary task showed a significantly smaller estimated RMSSD (EM=-0.236, SE=0.185) than under L2 driving conditions in the presence of a secondary task (EM =0.0475, SE=0.187). Pearson correlation between RMSSD and the chronological order of the scenarios showed a weak but positive and significant correlation (ρ=0.122, p=0.011).

3.2.4.3 Evaluation of HP performance

This section presents the analyses that were carried out to evaluate the extent to which drivers hazard perception performance was affected when participants drove under partially automated driving conditions and to determine whether the inclusion of a fatigue mitigating secondary task can help in mitigating the assumed negative effects of L2 driving on hazard perception. We currently report the analysis of the number of glances on the hazards during each experimental condition. The other analyses are currently a work in progress and will be reported in WP2.

The dependent variable in this analysis is the number of glances that were allocated to the hazard. This variable is Poisson distributed where each glance is considered an event in time. The final GLMM Poisson Regression model included two significant main effects of automation level - F(1,421)=5.575, p <0.05) and event chronological order F(6,421)=2.134, p <0.05). However, the interaction automation level*event order was found not significant (shown in Figure 26). Participants under manual driving had a significantly larger number of glances at the hazard (EM =5.38, SE=0.55) than under the condition of Level 2 driving (EM =4.16, SE=0.4).

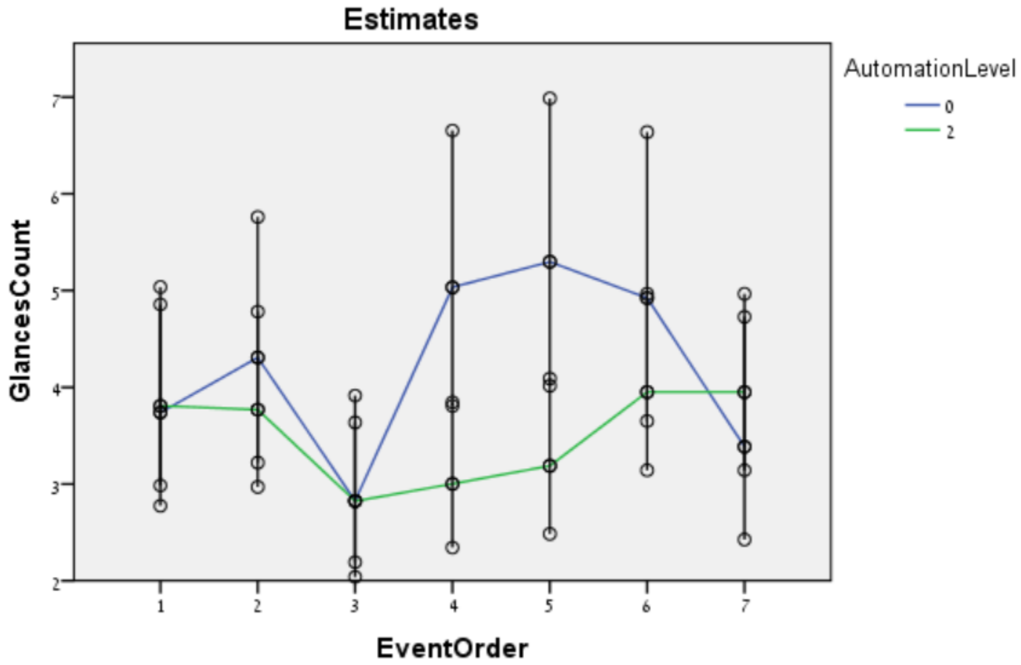


Figure 26. The interaction between scenario chronological order and automation level

3.2.5 Discussion

This study had two goals. The first goal was to investigate the effects of partially automated driving (i.e., L2), on fatigue progression and hazard perception performance with respect to manual driving. The second goal of the study was to examine whether engagement with a secondary task during L2 driving conditions can mitigate fatigue progression and mitigate the assumed decline in HP performance under L2 driving conditions due to driver's underload.

In this study, participants were asked to drive in a driving simulator two drives of 40 minutes each in one of two experimental conditions. The first condition examined fatigue progression and HP performance differences between manual driving and partially automated driving conditions without a secondary task. The second experimental condition examined fatigue progression and HP performance differences between L2 driving conditions in the absence of a secondary task and L2 driving conditions in the presence of a secondary task.

The discussion is organized around the three hypotheses that followed the literature overview and discusses the findings considering the literature.

In accordance with our first hypothesis H1(a) the results demonstrated that fatigue levels are higher when driving under partial automation compared to manual driving. This result complies with the findings of Körber et.al as their research showed that fatigue developed when there is no active engagement in the driving task (Körber et al., 2015). Specifically, the KSS questionnaire scores showed higher scores for L2 drives than for manual driving conditions. The chronological order of the scenarios showed a positive and significant Pearson correlation with the variable RMSSD, implying that as the experiment progressed the HRV measure also increased. This finding is in line with previous findings showing that an increase of HRV correlates with relaxation, time-on-task and fatigue (Gershon et al., 2009). However, no significant differences were found between L2 and manual driving.

Our results also supported H2(a). Secondary task engagement leads to lower KSS scores under L2 driving conditions than L2 driving conditions without task engagement. This hypothesis was based on the theory of fatigue mitigation using an activating task. A previous research concluded that during CAD (conditional automated driving), participants who had to engage in the monitoring of the driving task showed higher PERCLOS measures than participants who had to engage in the activating quiz task (Körber et al., 2015).

With respect to H2(b), driving under L2 with a secondary task present resulted in higher values of the RMSSD, which corresponds with higher values of HRV. Secondary task existence, as shown previously in this section, mitigates driver fatigue, thus we would expect that HRV will decrease when a secondary task exists. This contradiction might be explained by the short-term psychological stress that was induced by the competition generated by the Simon game. Participants were encouraged to engage and perform well in the game by granting a monetary prize for the best performers, therefore they were stressed to perform well. As suggested in Delaney et.al, short term psychological stress leads to a significant increase in the low frequency to high frequency ratio (LF/HF) causing an increase in HRV (Delaney and Brodie, 2000).

Finally, with respect to H3 (a and b), it was hypothesized that drivers under L2 driving conditions will glance at potential hazards for shorter durations than drivers under the manual driving condition. As shown in Figure 26, throughout the drive the number of glances on the hazard's AOI was smaller under L2 and larger under manual driving (there was no difference within the L2 conditions with respect to the secondary task). An exception to this was the final hazard that corresponds with the Traffic Cascade scenario for all the drives. The hazard in this scenario is materialized, thus it is reasonable to assume that participants glanced more at it even in the automated condition. This hypothesis was based on previous studies on the mindlessness theory, also called underload theory. Driving under L2 automation is a typically referred to as a vigilance task, monotonous and under-stimulating. According to mindlessness theory, as the mind disengages from the task, it wanders and becomes occupied with task irrelevant thoughts. During the task, the observers' ability to respond to infrequent critical signals separated by long intervals wanes over time. A possible explanation is that their supervisory attentional system loses its efficacy to focus on the vigilance task (Helton and Warm, 2008). Moreover, over-reliance on the automated agent translates into complacency and creates less vigilant drivers. When situations get complex and the driver is expected to take control or detect hazards, they might not be able to perform optimally (Onnasch et al., 2014).

3.3 References

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, *52*, 29 - 37.

Åkerstedt, T., Connor, J., Gray, A., & Kecklund, G. (2008). Predicting road crashes from a mathematical model of alertness regulation - The Sleep/Wake Predictor. *Accident Analysis and Prevention*, *40(4)*, 1480-1485.

Åkerstedt, T., Folkard, S., & Portin, C. (2004). Predictions from the Three-Process Model of Alertness. *Aviation, Space, and Environmental Medicine*, *75*, A75-A83.

Balasubramanian, V., Adalarasu, K., & Gupta, A. (2011). EEG based analysis of cognitive fatigue during simulated driving. *International Journal of Industrial and Systems Engineering, 7(2)*, 135–149. https://doi.org/10.1504/IJISE.2011.038563

Balkin, T., Bliese, P., Belenky, G., Sing, H., Thorne, D., Thomas, M., . . . Wesensten, N. (2004). Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *Journal of Sleep Research*, *13*, 219-227.

Bioulac, S., Franchi, M., Arnaud, M., Sagaspe, P., Moore, N., Salvo, F., & Philip, P. (2017). Risk of Motor Vehicle Accidents Related to Sleepiness at the Wheel: A Systematic Review and Meta-Analysis. *SLEEP*, *40(10)*. doi:http://dx.doi.org/10.1093/sleep/zsx134

Connor, J., Norton, R., Ameratunga, S., Robinson, E., Civil, I., Dunn, R., . . . Jackson, R. (2002). Driver sleepiness and risk of serious injury to car occupants: population based case control study. *BMJ*, 324 1125.

Dureman, E. I., & Bodén, C. H. (1972). Fatigue in simulated car driving. *Ergonomics, 15(3),* 299–308. https://doi.org/10.1080/00140137208924432

European Commission . (2019). *Staff Working Document: EU Road Safety Policy Framework 2021-2030 – Next steps towards "Vision Zero"*. Brussels: European Commission.

European Transport Safety Council. (2018). *Briefing: 5th EU Road Safety Action Programme 2020-2030*. Brussels: European Transport Safety Council (ETSC).

Feldhütter, A., Gold, C., Schneider, S., Bengler K. (2016). How the duration of automated driving influences take-over performance and Gaze behavior *Kongress der Gesellschaft für Arbeitswissenschaft,* p. 62

Fors, C., Ahlström, C., & Anund, A. (2018). A comparison of driver sleepiness in the simulator and on the real road. Journal of Transporrtation, *Safety and Security*, *10(2)*, 72-87.

Gershon, P., Ronen, A., Oron-gilad, T., & Shinar, D. (2009). The effects of an interactive cognitive task (ICT) in suppressing fatigue symptoms in driving. *Transportation Research Part F: Psychology and Behaviour, 12(1)*, 21–28. https://doi.org/10.1016/j.trf.2008.06.004

Golz, M., Sommer, D., Trutschel, U., Sirois, B., & Edward, D. (2010). Evaluation of fatigue monitoring technologies. Somnologie, 14(3), 187-199. doi:10.1007/s11818-010-0482-9

Hallvig, D., Anund, A., Fors, C., Kecklund, G., Karlsson, J. G., Wahde, M., & Åkerstedt, T. (2013). Sleepy driving on the real road and in the simulator--A comparison. *Accident Analysis and Prevention, 50*, 44-50. doi:10.1016/j.aap.2012.09.033

Jarosch, O., Bellem, H., & Bengler, K. (2019). Effects of Task-Induced Fatigue in Prolonged Conditional Automated Driving. *Human Factors*, *61(7)*, p. 1186-1199. https://doi.org/10.1177/0018720818816226

Körber, M., Cingel, A., Zimmermann, M., & Bengler, K. (2015). Vigilance Decrement and Passive Fatigue Caused by Monotony in Automated Driving. *Procedia Manufacturing, 3(Ahfe)*, 2403–2409. https://doi.org/10.1016/j.promfg.2015.07.499

Kyriakidis, M., de Winter, J. C., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., . . . Merat, N. (2019). A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science*, *20(3)*, 223-249.

Lal, S. K. L., & Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological Psychology, 55(3)*, 173–194. https://doi.org/10.1016/S0301-0511(00)00085-5

Loh, S., Lamond, N., Dorrian, J., Roach, G., & Dawson, D. (2004). The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior research methods, instrument & computers*, *36(2)*, 339-346.

Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving. *Frontiers in human neuroscience, 13*, 57.

May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F: Traffic Psychology and Behaviour, 12(3)*, 218–224. https://doi.org/10.1016/j.trf.2008.11.005

Mccoy, B. H. B. (2009). L I V I Ng.

Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., & Neubauer, C. (2013). Active and passive fatigue in simulated driving: Discriminating styles of workload regulation and their safety impacts. *Journal of Experimental Psychology: Applied, 19(4),* 287–300. https://doi.org/10.1037/a0034386

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 and Prevention, 41(5)*, 1087–1093. https://doi.org/10.1016/j.aap.2009.06.007

Schömig, N., Hargutt, V., Neukum, A., Petermann-Stock, I., & Othersen, I. (2015). The Interaction Between Highly Automated Driving and the Development of Drowsiness. *Procedia Manufacturing, 3(Ahfe)*, 6652–6659. https://doi.org/10.1016/j.promfg.2015.11.005

Smith, S., Horswill, M., Wetton, M., & Chambers, B. (2009). Sleepiness and Hazard Perception While Driving. *Road Safety Grant Report 2009-001*. (April), 27.

Solís-Marcos, I., Ahlström, C., & Kircher, K. (2018). Performance of an Additional Task During Level 2 Automated Driving: An On-Road Study Comparing Drivers With and Without Experience With Partial Automation. *Human Factors, 60(6)*, 778–792. https://doi.org/10.1177/0018720818773636

Vogelpohl, T., Kühn, M., Hummel, T., & Vollrath, M. (2019). Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving. *Accident Analysis and Prevention, 126*, 70–84. https://doi.org/10.1016/j.aap.2018.03.013

Williamson, A., Lombardi, D. A., Folkard, S., Stutts, J., Courtney, T. K., & Connor, J. L. (2011). The link between fatigue and safety. *Accident Analysis and Prevention*, *43(2)*, 498-515. doi:10.1016/j.aap.2009.11.011

4 Driver competence and comfort

Research on driver competence and driver comfort was split into two parts presenting two different approaches: (1) Online / real-time comfort assessment presents empirical results from 2 driving simulator studies including 81 participants. The studies aimed at identifying behavioural indicators of discomfort in automated driving using face tracking and facial expression analysis. (2) Offline / long-term comfort and competence research aimed at collecting and identifying situations from previous studies, in which drivers would prefer automation instead of driving manually.

Both approaches are aimed at detecting driving situations that cause a decrease in drivers' comfort, to intervene as soon as possible (e.g. in advance before driving comfort is impaired or as soon as a decrease in comfort is detected), to avoid future uncomfortable driving situations (e.g., due to an inappropriate automated driving style or by offering automated driving in a traffic jam), and to adapt the Mediator system to drivers' individual preferences for automated vs. manual driving in future driving situations.

4.1 Online/Real-time comfort assessment

4.1.1 Relevance of real-time comfort assessment in automated driving

Automated driving is supposed to bring several benefits such as an increase of traffic efficiency and safety, reduced vehicle emissions and energy consumption, economic competitiveness and social inclusion by offering new options for all mobility users (Smith et al., 2018). However, to exploit these expected benefits, societal and user acceptance of automated vehicles is considered crucial (ERTRAC, 2019). Due to the changing role of the active driver towards passenger/user, satisfaction with automated vehicle operations becomes an important novel topic in automotive human factors research (Elbanhawi et al., 2015; Hartwich et al., 2018). Vehicle operations hereby intend the full range of vehicles' behaviour on the road, including deceleration, acceleration, car following, gap acceptance, lane keeping, lane changing etc. (Smith et al, 2018). Ensuring a comfortable and positive driving experience is considered a fundamental prerequisite for the acceptance and usage of automated vehicles, connected to the realization of promises concerning new opportunities during driving, such as relaxation, work and entertainment (Bellem et al., 2018). Thus, monitoring and enhancing drivers' comfort and satisfaction with automated vehicle operations is an important aim to maintain well-being, associated with the absence of discomfort. Next to traditional comfort aspects such as noise, vibrations or sitting comfort, additional factors are discussed in automated driving such as apparent safety, motion sickness, trust in the system, controllability, familiarity of vehicle operations as well as information about system states and actions (Beggiato, 2015; Elbanhawi et al., 2015; Domeyer et al., 2019). As these novel comfort issues are primarily related to specific and dynamic driving situations and differ between individuals, continuous evaluation is indicated (Domeyer et al., 2019). Thus, continuous driver comfort monitoring could provide useful input to adapt and improve automated vehicle operations, based on the idea of a vehicle-driver team that knows each other's limitations, strengths and current states and is able to act accordingly (Klein et al., 2004).

The Mediator system is intended to intelligently assess the weaknesses and strengths of both the driver and the automation in the current driving context and mediate between them. Based on driving fitness, but also comfort and preferences in a certain context, the system could actively propose automation features as a service to the driver. Next to comfort and safety aspects, the active proposal of vehicle automation features could tackle a well-known problem of current Advanced Driver Assistance Systems (ADAS). A great proportion of drivers are not even aware of having a potentially helpful system in their own vehicle, report disuse/misuse or demonstrate misperceptions about what the system can and cannot do (McDonald et al., 2018). Therefore, potential safety effects of ADAS are diminished. In addition, comfort aspects during automated driving are not only relevant for a pleasant driving experience and acceptance of automated technology, but they can also have safety impacts. Unnecessary interventions by the driver due to uncomfortable or unexpected vehicle operations (e.g., if apparent safety is perceived as compromised) could lead to safety-critical and unnecessary takeover situations (Hergeth et al., 2016; Techer et al., 2019). Thus, detection of comfort issues could allow for adapting automation features such as driving style aspects (e.g. distance to vehicle ahead, speed, lateral distance) and/or information presentation with the overall aim to prevent disengagement of automation or dangerous and unnecessary takeover situations.

4.1.2 Facial expressions as behavioural indicator for comfort and discomfort

Traditional comfort measurements usually make use of questionnaires (a collection of comfort questionnaires can be found at http://www.icc.tudelft.nl/Questionnaires/Questionnaires.htm). However, there is a continuous research effort to get more objective human comfort measures that can be assessed by sensors in real-time. Current approaches use for example heart rate, heart rate variability, electrodermal response, pupillometry, oxygen saturation, electromyography, electroencephalography, pressure distribution by seat mats and posture analysis (Beggiato et al., 2018; Ikeda et al., 2018; Tan et al., 2008). In addition, facial expressions could be a promising source of information about users' current perception of automated vehicle operations. Changes in the activation of the muscles in the face are a fundamental channel for communicating and understanding emotional states in social interactions (Ekman & Friesen, 2003). Thus, incorporating the capability to read facial expressions in technical systems could potentially increase the understanding of current user states, as envisioned in the vehicle-driver team metaphor (Domeyer et al., 2019; Ihme et al., 2018b, Neubauer et al., 2020). The Facial Action Coding System (FACS, Ekman et al., 2002) is a system to encode the movements of specific individual facial muscles or muscle groups called Action Units (AUs). AUs reflect distinct momentary changes in facial appearance and do not interpret the meaning of expressions. However, combinations of specific AU changes can be used to infer particular emotional states (Ekman et al., 2002). Traditional FACS analyses are based on human annotation of AU changes, which is time-consuming and requires intensive specialized training. However, technological developments in automated face tracking based on video processing and machine learning offers promising results in automated detection of AU changes (Bryant & Howard, 2019; Ko, 2018).

These recent developments make video-based facial expression analysis attractive as in-vehicle driver state sensor for ADAS and automated vehicles. First, video cameras are relatively inexpensive, compared to some other sensors built into the steering wheel or driver seat and/or based on physiological measures. Second, cameras allow for unobtrusive measurement in contrast to sensors attached to the face, head, or skin. Third, one or just a few cameras can measure a variety of driver state properties, such as eye openings and closures, eye gaze direction, head movements, hand and body pose information as well engagement in secondary tasks (Christoph et al., 2019). As a consequence, facial expression analysis has already been investigated in the automotive context to detect fatigue and distraction (Sigari et al., 2014), stress (Gao et al., 2014), frustration (Ihme et al., 2018a) as well as emotional reactions to specific automated driving manoeuvres (Domeyer et al., 2019). However, general issues when using video cameras in vehicles are related to privacy concerns, problems under varying lighting conditions, glare, occlusions and processing of rather large video data. In addition, reliable, stable and valid identification of distinct driver states is still challenging. On the one hand, environmental conditions can heavily influence face tracking quality, such as illumination (glare, sunlight, contrasts, directional lighting), face angle, occlusions of face parts due to reflecting (sun)glasses, beards, jewellery, piercings, caps, hats, hands, arms or other objects such as the steering wheel in vehicles. On the other hand, individual differences in the quality and quantity of facial expressions complicate the establishment of a direct relationship between certain patterns of AU changes and distinct states. Thus, recent works aim at identifying groups/clusters of facial expressivity and relate these clusters to personal characteristics in order to gain more knowledge about individual specifics in facial expressions (Domeyer et al., 2019; Neubauer et al., 2020).

4.1.3 Empirical study aims

The present empirical study aims at exploring the potential of automated facial expression analysis for discovering AU changes related to uncomfortable automated driving manoeuvres. The data are combined from two driving simulator studies including 81 participants, all experiencing the same automated close approach manoeuvre to a truck driving ahead three times. This highly standardized situation was selected due to multiple reasons. First, keeping short distances is one of the most mentioned causes for discomfort when driving as a co-driver (Beggiato et al., 2019). Second, maintaining a comfortable distance to vehicles driving ahead is already relevant for existing ADAS systems and partial automation. Finally, the high standardization and therefore experimental control of this particular automated vehicle operation allows for gaining larger amounts of data to analyse AU changes in the same situation by different individuals. To ensure that all analyses relate to individually perceived discomfort, a handset controller was integrated into the driving simulator. By means of this device, participants could report discomfort gradually and continuously during the whole trip (Figure 27, right). In addition to identifying relevant AU changes during perceived discomfort, the influence of situational characteristics on face tracking quality (such as camera type and perspective) was in the focus of investigation. A clustering approach finally aimed at separating two groups, showing either high or low situation-related effects in relevant AUs. These two clusters were subsequently characterized by situational and personal characteristics (such as personality questionnaires and demographics) to obtain information about individual differences in facial expressivity.

4.1.4 Study materials and methods

Driving Simulator Studies: The data for all analyses originate from two driving simulator studies with different participants, but the exactly identical automated trip. Data collection was funded by the Federal Ministry of Education and Research under grant No. 16SV7690K (Project KomfoPilot) and took place between 2017 and 2019. Both studies were conducted in a fixed-base driving simulator (Software SILAB 5.1) with a fully equipped interior, two side mirrors, a rear-view mirror, and a 180° horizontal field of view (Figure 27 left). A three-minute drive was pre-recorded by the investigators and replayed while the participants sat in the driver’s seat, pretending that all aspects of the driving task were executed by an automated vehicle. The steering wheel and the pedals were inoperative. Each trip consisted of three consecutive potentially uncomfortable close approach situations to a truck driving ahead. Each time the truck drove at a constant speed of 80 km/h, whereas the own car approached in automated mode with a constant speed of 100 km/h. Automated braking started at a rather short distance of 9 m, reaching a minimum distance of 4.2 m and a minimum time to contact of 1.1 s. Participants were not informed beforehand about the situation and were instructed to press the lever of a handset control (Figure 27 right) according to the extent of perceived discomfort. To capture the driver's face, two video cameras were used in Study 1 and four video cameras in Study 2, placed at various positions (extensive details and screenshots can be found in the results section).

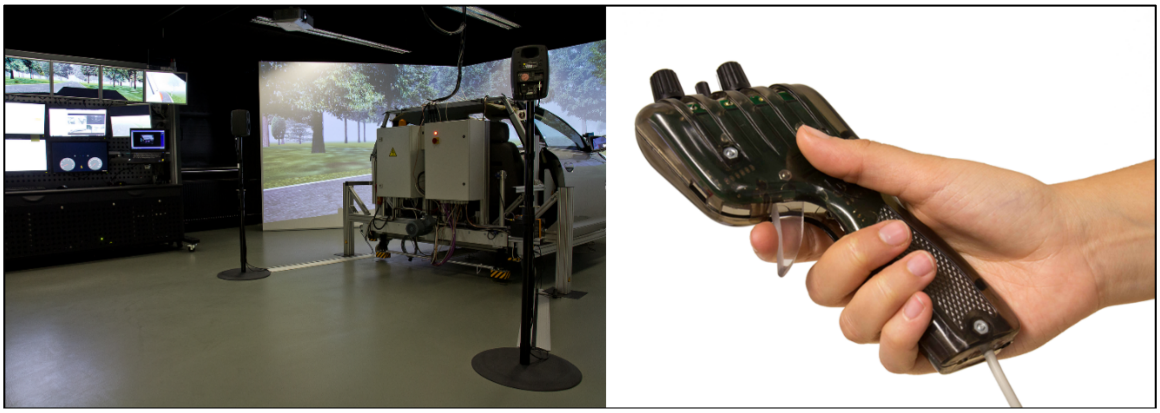


Figure 27. Fixed-base driving simulator (left) and handset control for gradual and continuous reporting of discomfort during each automated trip (right).

Participants: A total of 81 persons took part in the studies; 40 in Study 1 and 41 in Study 2. The sample consisted of 32 females and 49 males in two distinct age groups, a younger group under 40 years (M = 29 years, SD = 4.2, range = 24 to 29 years, N = 40) and an older group over 60 years (M = 70 years, SD = 5.6, range = 61 to 84 years, N = 41). All participants were required to hold a valid driver's license. The studies were carried out in line with the recommendations, regulations and consent templates of the TU Chemnitz ethics commission. All subjects gave written informed consent in accordance with the Declaration of Helsinki. Participants were compensated with 20 euros for participation.

Personality questionnaires: In addition to demographic data such as age, gender and annual mileage, each participant compiled several questionnaires on personality traits that were expected to potentially influence the perception of automated driving. The personality trait sensation seeking indicates the need for varied, novel and complex sensations and experiences and was assessed by the brief sensation seeking questionnaire (Hoyle et al., 2002). A closely related personality trait is uncertainty tolerance, expressing the overall tendency to perceive new and uncertain situations as challenging, exciting, intense and fruitful (Dalbert, 1999). Desirability of control indicates the general level of motivation to control the events in one's life (Burger & Cooper, 1979). The emotional relation to driving (Schmidt & Piringer, 2012) expresses the emotional value of driving a car, an exemplary question is “Driving a car is one of best things in life”. Finally, a questionnaire on the Big Five Personality model was used to assess the personality dimensions extraversion, agreeableness, conscientiousness, neuroticism and openness to experience (NEO-FFI-30; Körner et al., 2008). Extraverted persons are characterized as excitable, outgoing, energetic and talkative with high amounts of emotional expressiveness. Agreeableness intends altruism, affection, kindness and other prosocial behaviours. Conscientious persons plan ahead, are organized, mindful of details and think about how their behaviour affects others. Higher scores on neuroticism mean higher anxiety, moodiness, irritability, sadness and emotional instability. Openness to experience characterizes persons with a broad range of interests, imagination and creativeness.

Discomfort sequences determination: Whenever a participant pressed the handset control lever during a truck approach situation, this time period was marked as a discomfort interval (independent of magnitude). Having 81 participants and 3 approach situations, a theoretical maximum of 243 intervals could be present. However, the handset control was not pressed in all situations, thus, 192 discomfort intervals could be extracted (94 in Study 1 and 98 in Study 2). The mean duration of these 192 intervals was 7.72 s (SD = 5.45). To assess discomfort-related AU changes, an additional 10 s time interval before and after each discomfort interval was added to obtain a discomfort sequence (10 s + discomfort interval + 10 s). Due to the fact that discomfort intervals varied in duration, a common percent scale from 0% to 300% was created to integrate the AU trends of all sequences in a common chart (Figure 288 and Figure 29). Time periods before and after each discomfort intervals were always 10 s long, thus, 1% corresponds to 0.1 s. Each discomfort interval was divided into percent slices, and the mean of AU values was calculated for the time interval of the respective percent slice. For further details on this method see Beggiato et al. (2018).

Face tracking software and pre-processing of raw AU data: To extract the raw AU values, all video recordings from the six video cameras (details in Table) were processed by the Visage facial feature detection and face analysis SDK (Version 8.4 for Windows, visagetechnologies.com). The SDK was integrated into an in-house developed logging application, providing detailed log files. A total of 23 AUs were tracked by the software, each one as arbitrary decimal number for each video frame. In addition, the software estimated the tracking quality (TQ) for each video frame as decimal number from 0 to 1, with 0 as tracking lost completely and 1 as optimal quality (statistics for each video camera are reported in Table 9). To avoid distortions by low TQ, only AU values with a TQ equal to or greater than 0.3 entered the analyses; all lower values were treated as missing data. The cut-off score of 0.3 was chosen as trade-off between sufficient data quality and minimal data loss (loss < 1%). A cut-off score of 0.4 would have led to additional missing up to 3%. As raw AU values consistently showed high frequency fluctuations, a moving average was calculated over ± 2 s for each raw AU value. By this smoothing procedure, signal noise was reduced to obtain more stable patterns of changes. The resulting values were still arbitrary decimal numbers; therefore z-transformation was applied for all smoothed AU values of each video camera over each discomfort sequence. Z-transformation expresses raw values as distance to the mean in units of standard deviations, with a total mean of zero and a standard deviation of one. The transformed values represent relative changes of each AU within the discomfort sequence in a unified scale of measurement. This procedure allowed for data fusion of all six video cameras. Having two video cameras x 94 sequences in Study 1 and four video cameras x 98 sequences in Study 2, a total number of 580 standardized AU sequence values could be integrated. Missing AU values due to lost tracking in specific video cameras were excluded from calculations, resulting in 428 total sequences with valid AU values. These 428 sequences formed the basis for the overall AU trend charts displayed in Figure 28. The z-scores were averaged over the 428 sequences for each percent slice of the discomfort sequence (bold blue line in Figure 28) and the 95% confidence interval was calculated pointwise and plotted as a light red area around the means. If the confidence band does not overlap between two particular points in time or between two groups (Figure 29), these two means differ in a statistically significant manner. To identify situation-related AU changes, all resulting plots were checked for significant rises or falls around the discomfort interval with subsequent recovery, i.e., u-shaped or n-shaped profiles. The resulting 10 relevant AUs showing these effect patterns are displayed in Figure 28.

Clustering of high and low AU effect groups: Overall, u- or n-shaped AU trends represent general patterns of change during discomfort sequences. To obtain more information about differences in these patterns, a separation into two equally sized groups of sequences showing high and low AU effects was carried out. Clustering was performed based on the strength of discomfort-specific AU effects in all the 10 relevant AUs, either rise (n-shaped) or fall (u-shaped). Thus, for each of the 428 AU sequences, the effect strength was calculated by averaging the z-values during the discomfort interval, the 10 s beforehand (pre) and afterwards (post) and summing up the absolute differences between these means. The corresponding formula is $|(\bar{z}_{discomfort} - \bar{z}_{pre}) + (\bar{z}_{discomfort} - \bar{z}_{post})|$. The higher the value, the greater the change during the discomfort interval, either rise or fall. The resulting values were averaged across the 10 relevant AUs, resulting in a combined indicator of effect strength. A subsequent median split of this indicator divided the 428 sequences into 214 high and 214 low effect sequences, which are presented in Figure 29.

4.1.5 Results

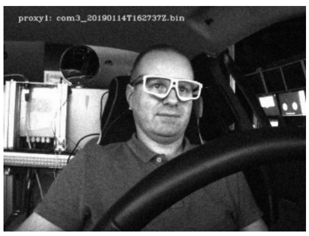
4.1.5.1 Video camera perspectives and face tracking quality

One aim of the two studies was to investigate effects of different video camera types and perspectives on face tracking quality and results. Three different video camera models were used, placed at various positions to capture the driver's face. In Study 1, a GoPro Hero 5 camera was mounted in the centre below the instrument cluster behind the steering wheel (Table 9, S1_G) and an Intel RealSense SR300 central over the steering wheel (Table , S1_SR300). These central positions allowed for capturing the face in an optimal frontal perspective, however, the position of the S1_SR300 would partially interfere with the field of view on the street ahead. Due to the optimal angle, face tracking quality was high with only 6% of video frames without tracking for S1_G and 12% for S1_SR300 during all discomfort sequences (Table). As both camera tracking values were combined for the AU analysis, an overall tracking rate of 97% could be reached with an overall mean TQ of 0.79.

In Study 2, four video cameras were used, capturing the driver's face from different perspectives. The main aim was to compare the impact of different camera angles on the face tracking results. Two GoPro Hero 5 cameras were placed at the right (Table , S2_GR) and left side from the driver's perspective (Table , S2_GL) below the instrument cluster behind the steering wheel. The Intel Realsense SR300 camera was again placed centrally over the steering wheel (Table , S2_SR300), however, this time slightly lower than in Study 1 to avoid potential obstructions of the view on the street. An additional AVT Mako G-234B grayscale video camera was placed at the right side from the driver's perspective next to the steering wheel (Table , S2_AVT). All face tracking quality values diminished significantly in Study 2 with percentages of not tracked video frames ranging from 26% to 43% (Table). For the AU analyses, all four camera tracking values were again combined, resulting in a total tracking rate of 80% during the discomfort sequences, with an average TQ of 0.71.

Overall, face tracking quality was primarily influenced by the camera angle toward the driver's face. Lower tracking rates for S2_SR300 compared to S1_SR300 mainly resulted from obstructed face parts by the upper part of the steering wheel. The seating position of some, mainly smaller, participants was so low that the mouth region was obstructed by the steering wheel, resulting in lost tracking or worse TQ. In addition, analyses of reduced TQ for specific participants revealed that TQ in both studies was lowered when persons wore reflecting eyeglasses or beards.

Table 9. Video camera specifications and face tracking quality statistics.

Study (1/2)	Camera type, specifications and placement	Face tracking percentage and tracking quality (TQ from 0..1) over all sequences	Example screenshots from original video
1	[S1_G] GoPro Hero 5, 1280 x 720 pixel, colour, 50 fps, central behind the steering wheel, wide angle	Tracking lost: 6.3% TQ < 0.3 = 0.3% TQ ≥ 0.3 = 93.4% *M* (*SD*) TQ ≥ 0.3 = 0.78 (0.13)	
1	[S1_SR300] Intel RealSense SR300, 1280 x 720 pixel, colour, 30 fps, central over the steering wheel	Tracking lost: 11.8% TQ < 0.3 = 0.1% TQ ≥ 0.3 = 88.1% *M* (*SD*) TQ ≥ 0.3 = 0.77 (0.14)	
2	[S2_GR] GoPro Hero 5, 1920 x 1080 pixel, colour, 50 fps, right side behind the steering wheel	Tracking lost: 25.7% TQ < 0.3 = 0.4% TQ ≥ 0.3 = 73.9% *M* (*SD*) TQ ≥ 0.3 = 0.67 (0.14)	
2	[S2_GL] GoPro Hero 5, 1920 x 1080 pixel, colour, 50 fps, left side behind the steering wheel	Tracking lost: 34.9% TQ < 0.3 = 0.1% TQ ≥ 0.3 = 65.0% *M* (*SD*) TQ ≥ 0.3 = 0.65 (0.11)	
2	[S2_SR300] Intel RealSense SR300, 1280 x 720 pixel, colour, 30 fps, central over the steering wheel (lower than in Study 1)	Tracking lost: 43.3% TQ < 0.3 = 0.3% TQ ≥ 0.3 = 56.4% *M* (*SD*) TQ ≥ 0.3 = 0.67 (0.13)	
2	[S2_AVT] AVT Mako G-234B, 640 x 480 pixel, grayscale, right side next to steering wheel	Tracking lost: 42.6% TQ < 0.3 = 0.5% TQ ≥ 0.3 = 56.9% *M* (*SD*) TQ ≥ 0.3 = 0.62 (0.14)	

4.1.5.2 Discomfort-related AU effects

To identify situation-related AU changes, all resulting AU plots for the 428 discomfort sequences were checked for significant rises or falls around the discomfort interval with subsequent recovery, i.e. u-shaped or n-shaped profiles. Figure 288 displays the averaged z-standardized values of all 10 AUs which showed such a discomfort-specific change.

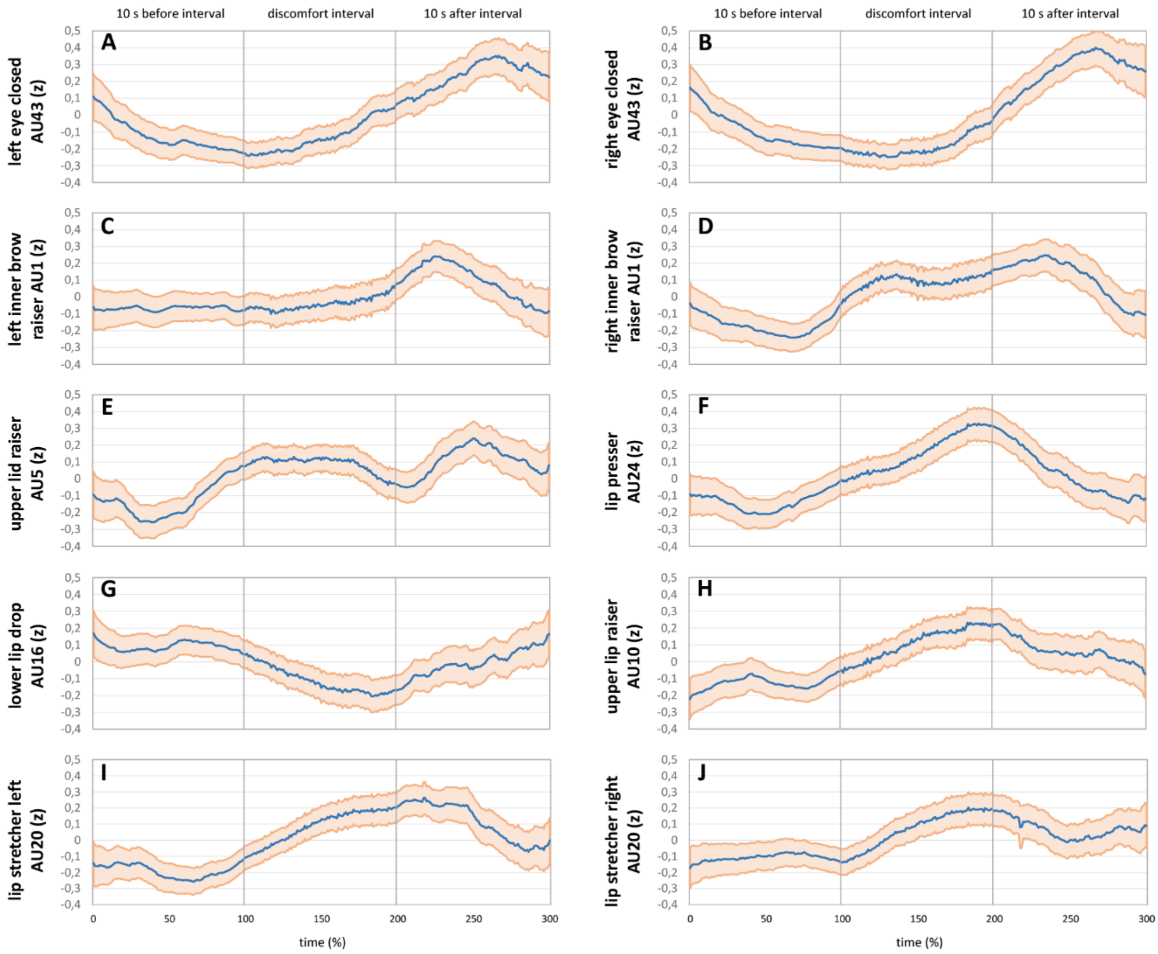


Figure 28. Mean z-scores of relevant AUs before, during and after discomfort intervals showing situation-specific effects as u- or n-shaped trend. The bold blue line shows the mean z-values over all 428 sequences, the light red area represents the 95% pointwise confidence interval.

In line with previous findings based on eye tracking (Beggiato et al., 2018), a reduction in eye closure during the truck approach could be observed for the right and left eye with a subsequent increase after the discomfort interval (AU43, Figure 288 A and B). This means that participants reduced eye blinks during the approach situation and kept their eyes open. An increasing trend could be observed for the inner brow raiser (AU1, Figure 28 C and D), more pronounced for the right inner side (Figure 288 D) and with a "delay" for the left inner brow (Figure 288 C, see clustering results for potential explanations). An increasing trend could as well be observed for the upper lid raiser (AU5, Figure 28 E). All these eye-related AU trends including rises of upper lids and inner brows as well as reduced eye blinks point towards a reaction of surprise and visual attention. Situation-specific changes in AUs around the mouth showed a considerable increase in pressing the lips (AU24, Figure 288 F). At the same time, lower lip drop diminished (AU16, Figure 288 G) and upper lip raiser increased (AU10, Figure 288 H), resulting in an upward oriented lip pressing motion. In addition, lips were stretched during the approach situation, both for the left (AU20, Figure 288 I) as well as for the right side (AU20, Figure 288 J). The combination of pressing and stretching the lips during the discomfort interval points towards a reaction that could be interpreted as tension.

4.1.5.3 High and low AU effect clusters

Even though the overall AU effects showed general patterns related to discomfort, the confidence interval bands suggest that there is still variability in these trends. Thus, a separation into two equally sized groups of sequences showing high and low AU effects was carried out, based on the combined strength of situation-specific changes in the relevant AUs (details on the clustering procedure are reported in the method section). Figure 29 shows the direct comparison between the high (red/bold line) and low effect group (green/dotted line), averaged over 214 sequences for each group. Regarding eye closure (AU43, Figure 29 A and B), the low effect group did not show situation-specific changes during the discomfort interval, whereas the high effect group reduced eye closure significantly. The raise of the inner brows during the discomfort interval was more pronounced for the high effect group (AU1, Figure 29 C and D). The low effect group in turn showed an opposite trend with a reduction of inner brow raising during the discomfort interval and a subsequent increase (in particular for the left inner brow). These opposite trends cancel each other out during the discomfort interval and the low effect group seems to react later with raising the inner brows.

A similar effect could be observed for the upper lid raiser (AU5, Figure 29 E), showing a delayed increase in raising the upper lid for the low effect group. The overall picture for the eye-related AU trends in the two groups suggests lower visual attention by unchanged eye closure as well as a potentially delayed surprise reaction in the low effect group. Pressing the lips (AU24, Figure 29 F) was markedly more pronounced in the high effect group, whereas the low effect cluster showed none to slightly decreasing trends. Distinct differences between the high and low effect group could also be observed for lower lip drop (AU16, Figure 29 G) and upper lip raiser (AU10, Figure 29H). Whereas the low effect group did not show any changes during the discomfort interval, the high effect group showed the situation-related decrease of lower lip drop and increase of upper lip raiser. With regard to lip stretcher at the left and right side (AU20, Figure 29 I and J), the means of both groups did not differ considerably, however, the discomfort-specific increase is more pronounced in the high effect group. Overall, lip-related AUs showed distinct situation-specific signs for tension in the high effect group, in contrast to almost no such signs in the low effect group.

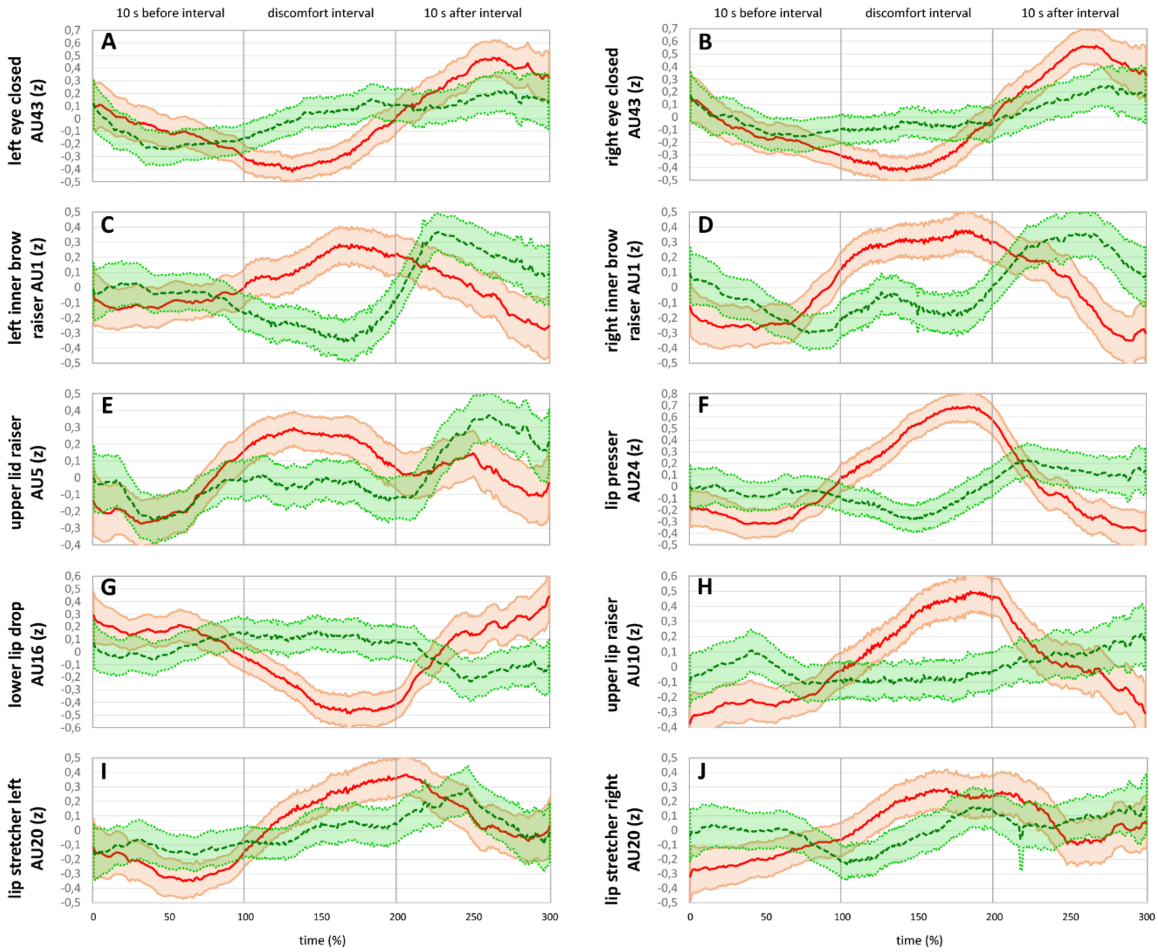


Figure 29. Mean z-scores of relevant AUs clustered in two equally sized groups of high and low discomfort-specific effects. The bold red line shows the mean z-values over all 214 high-effect-sequences, the green dotted line shows the mean z-values over all 214 low-effect-sequences. The light areas around the means represent the 95% pointwise confidence interval.

4.1.5.4 Situational and personal characteristics of the high and low effect clusters

After clustering high and low effect groups, an additional research question aimed at characterizing these two groups by situational and personal characteristics. Table 10 summarizes the situational characteristics of both groups, including a statistical comparison (significant differences are marked in bold). The proportion of sequences from Study 1 and Study 2 did not differ significantly between the groups, i.e. no study artefacts could be observed. The same applies for the six video cameras/positions used in the two studies. The proportion of sequences in the high and low effect group originating from the different cameras did not differ significantly, even if overall face tracking quality varied between the video cameras (see Table). This means that, even if tracking quality differed and higher loss of tracking of course leads to reduced potential of discovering AU changes, the effects itself were not artefacts of the camera model or position. The duration of reported discomfort intervals was similar in the two groups as well as the proportion of sequences with persons wearing glasses (including eye tracking glasses as shown in Table for S2_SR300 and S2_AVT), which could potentially have influenced the AU effects. A statistically significant difference was observed for the order of the three truck approach situations. The high effect group contains more sequences originating from the first and second situation, whereas the proportion of sequences from the last approach is higher in the low effect group. This means that AU effects in the high effect group were stronger at initial encounter of such situations and diminished over time.

Table 10. Comparison of situational characteristics between the high and low effect clusters (percentage values indicate the relative number of sequences within each group; statistically significant differences are marked in bold).

Situational characteristic	High effect group (*N* = 214 sequences)	Low effect group (*N* = 214 sequences)	Statistical significance between group comparison
	M (*SD*) or %	*M* (*SD*) or %	
Study 1/2	Study 1 = 38%, Study 2 = 62%	Study 1 = 43%, Study 2 = 57%	χ^2 (1, 428) = 0.968, *p* = .325
Video camera	S1_G = 21% S1_SR300 = 18% S2_GR = 18% S2_GL = 15% S2_SR300 = 16% S2_AVT = 13%	S1_G = 22% S1_SR300 = 21% S2_GR = 20% S2_GL = 13% S2_SR300 = 10% S2_AVT = 15%	χ^2 (5, 428) = 4.283, *p* = .509
Duration discomfort interval (s)	7.59 (4.98)	7.91 (5.33)	*t*(426) = 0.624, *p* = .533, *d* = 0.060
Wearing glasses (yes/no)	glasses = 69%	glasses = 61%	χ^2 (1, 428) = 2.957, *p* = .085
Truck approach situation (1/2/3)	**1st approach = 35%** **2nd approach = 36%** **3rd approach = 29%**	**1st approach = 28%** **2nd approach = 32%** **3rd approach = 40%**	**χ^2 (2, 428) = 6.040, *p* = .049**

Table 9 summarizes personal characteristics of the two groups; statistically significant differences are again marked in bold. It should be noted that all statistics were calculated over the 2 x 214 sequences and therefore data from one person could enter multiple times in the calculations. A first group difference appeared for age; the mean age in the high effect group was significantly lower than in the low effect group. The extraversion score was higher in the high effect group, meaning that this group includes more persons characterized by excitability, sociability, talkativeness, assertiveness, and emotional expressiveness. Persons in the high effect group showed higher scores in uncertainty tolerance as well as sensation seeking. Both constructs refer to a similar concept (and indeed showed a high correlation of r = .448), i.e. a tendency to search and tolerate experiences that are varied, novel, complex and intense. The last statistically significant difference showed higher scores for desirability of control in the high effect group, characterizing this group by higher motivation to control events in one's environment. No differences between the high and low effect clusters could be assessed for gender, driving experience (mileage per year), emotional relation to driving and the remaining Big-Five dimensions agreeableness, conscientiousness, neuroticism and openness. Surprisingly, no group differences could be noted for facial expressivity, calculated as mean standard deviation of the relevant AUs before performing the z-transformation. To summarize personal characteristics, the group showing higher AU effects is characterized by younger, extraverted persons with higher tendencies to search for varied, novel and intense situations that, however, are under their own control.

Table 9. Comparison of personal characteristics between the high and low effect clusters (statistically significant differences are marked in bold).

Personal characteristic	High effect group (N = 214 sequences)	Low effect group (N = 214 sequences)	Statistical significance between group comparison
	M (SD) or %	M (SD) or %	
Age (years)	**40.73 (19.42)**	**46.26 (21.17)**	**t(422.9) = 2.815, p = .005, d = 0.274**
Extraversion	**3.51 (0.58)**	**3.31 (0.66)**	**t(426) = -3.220, p = .001, d = 0.312**
Uncertainty tolerance	**3.53 (0.83)**	**3.20 (0.77)**	**t(426) = -4.120, p < .000, d = 0.399**
Sensation seeking	**3.12 (0.65)**	**2.99 (0.67)**	**t(426) = -1.979, p = .048, d = 0.192**
Desirability of control	**5.08 (0.59)**	**4.95 (0.65)**	**t(426) = -2.238, p = .026, d = 0.217**
Gender	female = 52%	female = 50%	□^{2} (1, 428) = 0.150, p = .699
Mileage (km/year)	10,779 (13,204))	11,652 (13,379)	t(426) = 0.680, p = .497, d = 0.066
Emotional relation to driving	3.33 (1.02)	3.33 (1.12)	t(426) = 0.001, p = .999, d = 0.000
Agreeableness	4.03 (0.54)	3.98 (0.56)	t(426) = -0.923, p = .357, d = 0.089
Conscientiousness	4.11 (0.54)	4.00 (0.66)	t(409.4) = -1.865, p = .063, d = 0.184
Neuroticism	2.25 (0.69)	2.28 (0.79)	t(426) = 0.398, p = .691, d = 0.039
Openness	3.67 (0.61)	3.58 (0.64)	t(426) = -1.484, p = .138, d = 0.144
Facial expressivity (SD relevant AU)	0.0926 (0.0411)	0.0916 (0.0447)	t(426) = -0.522, p = .602, d = 0.051

4.1.6 Discussion

Advantages of video-based facial expression analysis: Technological developments make video-based facial expression analysis attractive as in-vehicle driver state sensor for ADAS and automated vehicles. Cameras are inexpensive, unobtrusive and allow for an economic assessment of multiple driver state indicators. Monitoring drivers' satisfaction/comfort by facial expression analysis could provide useful input to adapt and improve automated vehicle operations, based on the idea of a vehicle-driver team that knows each other's limitations, strengths and current states. The present empirical study aimed at exploring the potential of automated facial expression analysis for discovering AU changes related to uncomfortable automated close approach situations. A first aim was to identify the influence of different video camera types and perspectives on face tracking quality. Second, general situation-related AU changes were identified as u- or n-shaped trends. To obtain information about differences in the overall effects, a clustering procedure based on the strength of the AU effects was applied and the resulting groups were characterized by situational and personal characteristics.

Video camera placement: With regard to optimal video camera placement for face tracking, deviations from a central frontal perspective on the driver's face resulted in significantly reduced tracking quality. Thus, a central camera placement behind the steering wheel appeared as best option, without limiting the field of view on the road ahead and avoiding obstructions due to differing seating positions. Next to the camera perspective on the face, tracking problems mainly occurred due to obstruction of face parts, reflecting eye glasses and beards. Even though tracking quality differed markedly between the video cameras, analyses of situational characteristics did not show influences of camera type/position on the AU changes. However, the better the tracking quality, the higher the chance of identifying AU effects just with few/one video camera. If face tracking should also be performed in situations with expected higher head rotation (e.g. secondary tasks in the centre console), using several video cameras covering different angles could still make sense.

Behavioural marker for discomfort / KSI: Changes in AUs related to the close approach situation showed that eyes were kept open (AU43) and eye blinks were reduced, indicating attentive visual monitoring of the situation. These findings are in line with previous eye tracking results from Beggiato et al. (2018, 2019). Observed raises of upper lids (AU5) and inner brows (AU1) are considered as essential component in all prototypes and major variants of surprise (Ekman et al., 2002). In addition, the lips were pressed (AU24) and stretched (AU20) during the discomfort interval. Similar AU patterns in the mouth region could be identified by Ihme et al. (2018a) for frustrated drivers. Simultaneously, lower lip drop diminished (AU16) and upper lip raiser increased (AU10), resulting in an upward oriented lip pressing motion. This combination of lip movements could be interpreted as sign for tension. Thus, the combination of all AU changes points towards a reaction of surprise, tension and visual attention during the close approach situation.

Personalization / subgroups: Even though these overall AU effects showed general patterns related to discomfort, the confidence interval bands suggest that there is still variability in these trends. Thus, a separation into two equally sized groups of sequences showing high and low AU effects was carried out, based on the combined strength of situation-specific changes in the relevant AUs. As intended, the high effect group showed clearly stronger effects in all AUs, especially in lip pressing, lip upward movement as well as eye closure. Less pronounced group differences resulted for lip stretching. Although raising the inner brows and upper lids differed between the groups, the low effect group showed a rather delayed than missing reaction (i.e. effects after the discomfort interval). To gather knowledge about characteristics of high and low effect clusters, both groups were compared with regard to situational and personal parameters. Situational characteristics were partly analysed as check for systematic bias, i.e. if certain study circumstances influenced the clustering results. The proportion of sequences from both studies, all camera models/positions, duration of discomfort intervals and percentage of wearing glasses did not differ significantly between the two groups. Thus, no evidence could be found for a systematic bias of these situational characteristics on the AU effects. In contrast, the order of driving situations showed a significantly different distribution in the groups, with higher proportions of earlier encounters in the high effect group. It is plausible that facial expressions are strongest at initial experience of such situations and a habituation effect can be expected over time (especially due to the fact that the three situations occurred within relatively short time periods). Personal characteristics of the two groups showed a higher extend of extraversion, sensation seeking, uncertainty tolerance and desirability of control in the high effect group, as well as lower age. Thus, the high effect group could be characterized by younger, extraverted persons with higher tendencies to search for varied, novel and intense situations that, however, are under their own control. These results are mainly in line with the core definition of these concepts, e.g. extraversion including higher emotional expressiveness (Körner et al., 2008; Neubauer et al., 2020) and sensation seeking/uncertainty tolerance as tendency to get involved in novel and intense emotional situations (Dalbert, 1999; Hoyle et al., 2002). However, the calculation of facial expressivity out of the AU data itself (as average standard deviation of the relevant AU during the sequences before z-transformation) did not show differences between the groups. However, it could be related to shortcomings of the indicator itself (average SD) and further research could try to find better/different overall indicators of variability in AU. No differences between the high and low effect clusters could be identified for gender, driving experience in mileage per year, emotional relation to driving and the remaining Big-Five dimensions agreeableness, conscientiousness, neuroticism and openness.

4.1.7 Conclusions

Overall, automated video-based facial expression analysis revealed situation-specific AU effects on aggregated level over participants and situations. Thus, there is general potential in face tracking for contributing useful information about users' satisfaction with the current operations of the automated system. The findings provide valuable indications of potentially relevant AUs indicating surprise and tension in this situation (including direction, timing and magnitude) as well as situational and personal influences on these effects. However, these results were obtained by aggregating data from several video cameras, participants and situations. Even though the situations were exactly the same, confidence bands indicate that there is still a high amount of variability in these trends, even if personalized z-transformation was applied and clustering disclosed parts of the overall variance. Thus, stable and reliable detection of discomfort based on AUs on individual level and in real time remains a challenging task. First, analyses of personal characteristics already revealed differences in effects according to personality. Therefore, facial expression analysis will not perform equally well for every person, even when personalizing AU changes at individual level. Second, the discovered effects could be shifted in time (as probably for raising the inner brows and upper lids), which already impacts aggregated data analysis and is even more challenging at individual level. Third, the applied clustering method aimed at separating effects in quantitative manner (high/low), however, it is not able to discover qualitative differences. It could be the case that individuals do not only show different strength of the (same) effects, but completely different AU patterns. Identifying these qualitatively different patterns would require different clustering methods and a higher amount of data, especially when aiming at the individual level. Machine learning algorithms could potentially tackle several of the mentioned problems, however, much higher amounts of data would be required to develop and train such algorithms. Even though, it is planned to test the potential of such algorithms within the MEDIATOR project. Video cameras will be present in MEDIATOR's driver state monitoring system anyway for detecting distraction and fatigue. Using machine learning-based facial analysis algorithms that overlap, to a large extent, with the distraction and fatigue detection algorithms, AU information can be extracted from these face videos. Fourth, the close approach situation is just one of several potential automated vehicle operations that could cause discomfort. Although distance regulation is considered an important comfort factor (Beggiato et al., 2019), further research is required to investigate if facial expressions are comparable in other potentially uncomfortable driving situations.

4.2 Offline / long-term comfort and competence

4.2.1 Relevance of long-term prediction of comfort in automated driving

Real time comfort assessment based on physiological data (e.g., facial expressions) is a promising approach to detect a decrease in drivers' comfort as soon as it occurs (e.g., close approach manoeuvre to a vehicle driving ahead during automated driving; see Chapter 4.1.). Once the decrease in drivers' comfort is detected, appropriate reactions can be implemented by the automated system (e.g., initiate a braking manoeuvre to keep a greater distance to the vehicle ahead). Such strategies are aimed to increase the drivers' comfort level again or at least to maintain the current comfort level (i.e., avoid a further decrease in drivers' comfort). Additionally, the automated driving style can be adapted accordingly (e.g., initiate the braking manoeuvre earlier and maintain a greater distance to the vehicle ahead in future approach manoeuvres). Hence, real time comfort assessment can help to detect driving situations that cause a decrease in drivers' comfort, to intervene immediately and to avoid future uncomfortable situations due to an inappropriate automated driving style. Hence, it is planned to test the potential of real time comfort assessment algorithms within the MEDIATOR project. Nevertheless, stable and reliable real time detection of discomfort on individual level based on physiological data (e.g., facial expressions) remains a challenging task. Further, to ensure a comfortable and positive driving experience it is not only important to react to decreasing drivers' comfort but also to avoid a potential decrease in comfort beforehand if possible.

An additional approach was chosen enabling the long-term prediction of a possible decrease in drivers' comfort based on previous knowledge. Thus, a literature research was conducted to collect and identify potentially uncomfortable driving situations from previous studies as well as the probability for a decrease in drivers' comfort in the respective situations. The information will be implemented in the Mediator system. If such a potential uncomfortable situation occurs, the automated system (e.g., the Mediator system) can suggest a take-over from manual to automated driving (or vice versa) in advance to maintain a high level of drivers' comfort. Based on the drivers' decision to accept or reject the take-over request, the automated Mediator system can be adapted to drivers' preferences by increasing or decreasing the frequency of future take-over requests before approaching a potentially uncomfortable driving situation. Hence, long-term prediction of comfort can help to detect potentially uncomfortable driving situations in advance, to prevent a decrease in drivers' comfort and to consider drivers' preferences for automated vs. manual driving in future driving situations.

4.2.2 Identifying potentially uncomfortable driving situations

A literature research was conducted to identify situations, in which drivers would prefer automation instead of driving manually or vice versa. Various queries were used in scientific databases such as ScienceDirect and Google Scholar. The research covers studies focusing mainly on comfort/discomfort, trust, perceived safety, satisfaction, acceptance, preferences, attitudes, intention to use, user needs, stress and identified use cases in the context of automated driving. Further, previous studies dealing with discomfort, stress and difficulties in manual driving were considered. More than 3.000 articles were scanned for potentially uncomfortable driving situations.
However, previous research focused mostly on:

- the identification of situations that can be fairly reliably covered by current automated driving systems, for instance, driving on a highway (e.g., Noh & An, 2017),
- specific driving manoeuvres that need to be optimized regarding drivers' comfort, for instance, automated lane change manoeuvres or automated parking manoeuvres (e.g., Suh et al., 2018; Furgale et al., 2013),
- drivers' preferences regarding automated driving styles (e.g., Beggiato et al., 2019; del Campo et al., 2018)
- drivers' opinion regarding specific automated functions, for instance, lane keeping assistance (Abraham et al, 2016),
- drivers' opinion regarding automated driving in general (e.g., Becker & Axhausen, 2017; Bazilinskyy et al., 2015).

Research on drivers' preferences for automated driving in specific driving situations is quite sparse. Therefore, research focusing on drivers' attitudes regarding different manual driving scenarios were also considered. It is assumed that drivers would prefer automated driving in situations they do not like to deal with themselves (e.g., driving in a line of cars; Engelbrecht, 2013). Moreover, previous literature could show that driving safety has the highest priority for drivers (e.g., Frison et al., 2018). Hence, it is further assumed that drivers would prefer automated driving in situations they do not feel competent enough to deal with themselves (e.g., driving in bad weather conditions or at night; Moták et al., 2014). The difference in drivers' competence (e.g., inexperienced vs. experienced drivers, elderly drivers) was considered in the process of identifying potentially uncomfortable driving situations. Especially drivers' age is assumed to play an important role due to age-related changes in cognitive, sensory and psychomotor functions possibly effecting driving performance (e.g., Bayam, Liebowitz, and Agresti 2005; Eby et al. 1998; Kaiser and Oswald 2000).

In a next step, the expected probability for a decrease in drivers' comfort was derived for the different identified driving situations based on the results found in the literature. It is suggested to use the derived probability values (see Table 12) as starting values for the implementation in the Mediator system. For studies showing that a certain percentage of participants prefer automated driving or do not like to drive manually in a specific situation, the expected probability for a decrease in drivers' comfort is equivalent to this value. For instance, if 60% of the participants stated that they do not like driving in a line of cars, the probability for a decrease in drivers' comfort when approaching a traffic jam in manual driving mode was assumed to be 60%. For studies in which participants were asked to rate certain driving situations regarding, for instance, the experienced stress level on a Likert scale another approach was chosen. The probability for a decrease in drivers' comfort was derived in relation to the values indicated on the Likert scale. For instance, when the Likert scale ranged from 0 (not stressful) to 4 (very stressful), then the probability for a decrease in drivers' comfort was set to 0% for the scale point 0 and to 100% for the scale point 4. If drivers rated a certain situation with a mean value of 1.6 on this Likert scale, the expected probability for a decrease in drivers' comfort was 40%. If diverse results were found in the literature, the mean value was calculated to derive the expected probability of a decrease in drivers' comfort. Studies reporting drivers' intention to completely avoid certain driving situations were also considered to identify comfort-relevant situations and differences between younger and older drivers. However, the avoidance scores were not used to derive the probability for a decrease in drivers' comfort, because avoidance represents the highest level of drivers' discomfort. For the situations for which the literature provided not enough information to calculate the probability of an expected decrease in drivers' comfort, it is proposed to start with a probability of 50% (i.e., equal probability for change vs. no change in drivers' comfort). Alternatively, the drivers can be asked in advance for their individual preferences for automated driving in predefined situations and the initial values can be set accordingly.

The automated system (e.g., Mediator system) should provide take-over requests from manual to automated driving or vice versa based on the expected probability of a decrease in drivers' comfort (based on literature, set to random level or based on drivers' stated preferences). For instance, if the probability for a decrease in drivers' comfort was set initially to 60%, the take-over request should be offered at least in every second traffic jam. However, the driver should have the possibility to initiate a take-over action by him- or herself at any time given that automated driving is available or he / she is fit to drive manually. Based on the decision of the driver to accept or reject the take-over request or to initiate a take-over action him- or herself, the initial probability value for a decrease in drivers' comfort need to be adjusted. In case the driver accepts or initiates the take-over request, the probability needs to be increased leading to a higher frequency of take-over requests if approaching a similar situation. Accordingly, the probability needs to be decreased leading to a lower frequency of take-over requests in case the driver rejects the take-over request. It is suggested to implement higher changes in the probability values around 50% (i.e., random level) and smaller changes in the periphery (i.e., near 0 or 100%). With this approach, the frequency of take-over requests can be adapted to the individual preferences of the drivers for specific situations. Moreover, intra-individual differences can be considered. It might be possible that the same driver will have different preferences for automated vs. manual driving on different days depending on his / her current mood. For instance, if a driver typically accepts take-over requests on long highway drives, the frequency of take-over requests is high (e.g., 95%). In case the driver wants to drive him- or herself because he / she wants to enjoy the manual drive, he / she will reject the take-over request once. The frequency of take-over requests will be adapted accordingly, but just slightly due to the smaller changes in the periphery. Hence, for the next trip on a highway the driver will still receive a take-over request. Just in case the driver's preference has changed completely (i.e., the driver does not want to be driven automatically on highways anymore), he / she will reject the take-over request more often leading to a noticeable decrease of take-over request frequency. It is suggested to provide the possibility to change the frequency of system-initiated take-over requests manually in the system settings enabling the drivers to adapt the system to their changed needs and preferences.

In the special case that the driver approaches a situation with several indicators for a decrease in drivers' comfort (e.g., morning commute combined with heavy rain), it is suggested to provide take-over requests with the frequency associated with the most relevant situation for the individual driver (i.e., the situation with the highest probability for a decrease in drivers' comfort). Based on the decision of the driver, the system should adapt the frequency of take-over requests for the specific combination of situational characteristics. Further, if previous research revealed that drivers' competence (e.g., age) has an influence on the expected probability of a decrease in drivers' comfort (e.g., elderly drivers tend to avoid driving at night more often compared to younger drivers; Baldock et al., 2016), the initial probability value should be adapted accordingly. It is suggested to increase / decrease the probability by one step (i.e., pretend the driver has accepted / rejected a take-over request once).

Additionally, the probability (high vs. low) to detect potentially uncomfortable situations in advance was discussed for each situation (e.g., predictability of a traffic jam vs. incoming messages). Further, it was discussed how much in advance (long-term vs. short-term) the situations can be detected (e.g., driving at night vs. interaction with passengers). For the discussions, it was assumed that the vehicle is equipped with all necessary technology (e.g., sensors, cameras) to gather the relevant information.

Further, trust was identified as a prerequisite for acceptance and use of automated driving (e.g., Lee & Moray, 1992). Hence, drivers' trust in the automated driving system will influence the probability of take-over request acceptance or rejection. Therefore, the system (e.g., the HMI) need to be designed to maximize and maintain drivers' trust, for instance, by providing relevant information in an appropriate way (e.g., Diels & Thompson, 2017). For instance, the reason for a take-over request should be given together with a certainty rating (e.g., heavy rain is forecasted) to prevent a reduction in drivers' trust in case he / she accepted the takeover request but the respective situation will not occur (e.g., there will be no rain on the trip).

Moreover, drivers require that the automated driving system will be at least as safe as manual driving is (e.g., Lee et al., 2018). It needs to be stated that the situations were collected based on the expected change in drivers' comfort (i.e., situations in which the driver is expected to prefer automated driving). It was not considered if current automated driving systems are able to handle those situations better than the human driver (e.g., complex traffic scenarios). This topic will be covered in Deliverable 1.3 "Degraded automation performance".

4.2.3 Results regarding potentially uncomfortable driving situations

Two clusters of situations will be presented. The first and most important cluster "potentially uncomfortable situations in manual driving" consists of situations for which a take-over from manual to automated driving should be offered by the automated system to maintain drivers' comfort (e.g., car following scenarios, situations with poor visibility). The second cluster "potentially uncomfortable situations in automated driving" contains of situations for which a hand-over from automated to manual driving should be offered (e.g., road conditions that increase the risk of motion sickness).

One potential use case for fully automated driving is to increase mobility for people who are not able to drive themselves (e.g., people without a driving license, children, elderly people who do not want to drive anymore, people with mental or physically disabilities, drivers under the influence of alcohol). Although literature could show that people prefer automated driving to be able to deal with these use cases (e.g., Cunningham et al., 2019; Payre et al., 2014), they will not be considered in this chapter. Within the MEDIATOR project, several automation levels as well as transitions between the different levels will be considered. Hence, this chapter focuses on specific situations (not on a whole trip) in which either automated or manual driving is preferred. Use cases that require fully automated driving without the possibility for a take-over action by the driver are not in the scope of this chapter. Further, this chapter focuses on individual transport and preferences of private car users. Professional drivers (e.g., truck drivers, bus drivers) or car-sharing approaches are not in the scope of this chapter.

An overview of the identified potentially uncomfortable driving situations, the derived expected probability for a decrease in drivers' comfort based on the results found in the literature, the probability to detect the situations as well as the possible time span for detection in advance can be found in Table 12.

Table 12. Overview of the identified potentially uncomfortable driving situations, the expected probability for a decrease in comfort, the probability to detect the situations and the possible time span for detection in advance.

		Situation	Probability of a decrease in drivers' comfort: Starting value	Probability of a decrease in drivers' comfort: Suggested adaptions due to drivers' competence	Probability to detect situations in advance	Time span for the premature detection of the situations
Potentially uncomfortable situations in manual driving	Car following scenarios	Traffic jams on a highway	55%	increase for inexperienced drivers	high	several minutes
		Following a slow vehicle on a country road	55%	/	medium	several seconds or minutes
		Stop and go traffic during rush hour in cities	55%	increase for inexperienced drivers	high	several minutes
	Situations impairing drivers' vision	Driving at night	54%	increase for elderly drivers and inexperienced drivers	high	several minutes
		Driving in rain	40%	increase for elderly drivers and inexperienced drivers	medium	several seconds or minutes
		Driving in snow	50%	increase for elderly drivers and inexperienced drivers	medium	several seconds or minutes
		Driving in fog	33%	increase for elderly drivers and inexperienced	medium	several seconds or minutes
		Driving with low sun	50%	/	medium to high	several seconds
	Situations with high complexity	High traffic density (e.g., rush hour in cities, complex intersections)	40%	increase for elderly drivers and inexperienced drivers	high	several seconds or minutes
	Challenging driving manoeuvres	Overtaking vehicles	15%	increase for elderly drivers and inexperienced drivers	medium to high	several seconds
		Passing obstacles	25%	increase for elderly drivers	medium to high	several seconds

		Narrow single-lane roads or narrow lanes in construction zones	50%	/	high	several seconds or minutes
		Parking	25%	increase for elderly drivers	high	several seconds or minutes
	Situations causing high levels of uncertainty	Driving in unknown environments (e.g., unknown cities or routes, foreign countries with differing traffic rules)	50%	increase for elderly drivers	high	several minutes

		Situation	Probability of a decrease in drivers' comfort		Probability to detect situations in advance	Time span for the premature detection of the situations
			Starting value	Suggested adaptions due to drivers' competence		
Potentially uncomfortable situations in manual driving	Demanding or highly prioritized NDRTs	Watching the scenery	50%	/	medium to high	several minutes
		Interaction with passengers	45%	/	medium	no time in advance
		Eating / drinking	40%	increase for younger drivers	medium to high	several seconds
		Incoming phone calls or text messages	40%	increase for younger drivers	high	several seconds or minutes
		Resting / sleeping	33%	increase for younger drivers	medium	several seconds or minutes
		Reading / watching movies	27%	increase for younger drivers	low to medium	several seconds or minutes
		Working	25%	increase for younger drivers	medium	several seconds or minutes
	Impaired driver state	Fatigue	70%	increase for elderly drivers	high	several seconds or minutes
		Frustration	50%	/	high	several seconds or minutes
		Stress	50%	/	high	several seconds or minutes
		Distraction	50%	/	high	no time in advance
	Longer and monotonous trips	Long duration car driving	20%	/	high	several seconds or minutes
		Regular commutes	60%	/	high	several seconds or minutes

		Situation	Probability of a decrease in drivers' comfort: Starting value	Probability of a decrease in drivers' comfort: Suggested adaptions due to drivers' competence	Probability to detect situations in advance	Time span for the premature detection of the situations
		Longer trips (not work-related)	35%	/	high	several seconds or minutes
		Monotonous driving	65%	/	high	several seconds or minutes
	Other	Optimize fuel or energy efficiency	80%	/	high	several seconds or minutes
Potentially uncomfortable situations in automated driving	Situations with higher risks of motion sickness	While engaging in NDRTs	75%	/	high	several seconds or minutes
		When not engaged in NDRTs	50%	/	medium	no time in advance
	Situations with high demands on communication with other road users		50%	/	high	several seconds or minutes
	Driving under time pressure		50%	/	medium	several seconds or minutes
	Purpose of the trip (e.g., driving for pleasure)		50%	/	high	several seconds or minutes
	Situations that cannot be handled by the automated system		100%	/	high	several seconds or minutes

4.2.3.1 Potentially uncomfortable situations in manual driving

Car following scenarios: Driving in a line of cars is one of the most stressful and negatively perceived driving scenarios in manual driving. In an online survey conducted with 1.800 car drivers in Germany, it could be revealed that nearly 60% of the respondents disliked driving in a line of cars like a traffic jam (Engelbrecht, 2013). According to Engelbrecht (2013) the probability for a decrease in drivers' comfort should be set to 60%. Another study revealed that 52% of the respondents expressed the intention to use automated vehicles in traffic jams (Sommer, 2013 as cited in Kyriakidis et al., 2015), corresponding to a probability for a decrease in drivers' comfort of 52%. In the study of Taylor & Paki (2008), respondents even stated, that they try to avoid traffic jams completely representing the highest possible level of discomfort. On a 5-point Likert scale ranging from 0 – never avoid to 4 – always avoid, respondents rated traffic jams with 1. Hence, car following scenarios like traffic jams on highways, following a slow vehicle (e.g., agricultural vehicles) on a country road or stop and go traffic during the rush hour in larger cities are prototypical situations with the potential to cause a decrease in drivers' comfort while driving manually. Hence, the chance that drivers will accept a take-over from manual to automated driving is quite high right from the beginning. It is suggested to set the probability for a decrease in drivers' comfort to 55% according to Engelbrecht (2013) and Sommer (2013 as cited in Kyriakidis et al., 2015).

Further, it could be revealed that inexperienced drivers disliked starting a driving manoeuvre more (17.5%) compared to more experienced drivers (4%; Engelbrecht, 2013). Hence, it could be expected that stop and go driving during car following scenarios can lead to an even higher decrease in drivers' comfort for inexperienced drivers. Hence, the frequency for system-initiated take-over requests should be slightly higher for inexperienced drivers in case stop and go driving is expected. Apart from this, driver competence is not expected to have an influence on drivers' comfort during car following scenarios.

The probability to detect situations like traffic jams on highways or rush hour traffic in larger cities is rated as high. By using information from traffic news as well as prior knowledge of the usual times with high traffic density (i.e., rush hour) and other disadvantageous factors (e.g., construction zones, bad weather) the situations are well predictable. The information is also available several minutes in advance. Hence, the take-over request can be offered with an adequate preparation time for the driver. Situations that occurred shot-termed (e.g., emerging traffic jam due to a sudden car accident, approaching to a slow vehicle on a single-lane country road) are harder to predict. Vehicle-to-Vehicle communication might be one solution to increase the predictability of such events. Vehicles approaching a traffic jam can share this information with the rear traffic and, hence, the take-over request can be offered in advance for the approaching vehicles.

Situations with poor visibility impairing drivers' vision: impairing drivers' vision: Driving in situations with poor visibility is quite challenging in manual driving (e.g., Cahour, 2008). Driving at night is more demanding due to the poor visibility conditions compared to daylight, especially for people with impaired night vision (e.g., Kilpeläinen & Summala, 2007). Further, some drivers have problems with oncoming traffic and feel glared by the headlights (e.g., Anderson & Holliday, 1995). Additionally, the probability to get sleepy increases due to the circadian rhythm (e.g., Harris, 1977), which makes it harder to concentrate on the driving task. Hence, driving at night-time is a potentially uncomfortable situation and take-over requests should be offered by the automated system. In an online survey conducted with over 6.000 respondents it could be revealed that 33% somewhat agreed and 21% strongly agreed to the question if they want to use a fully automated car at night or in bad weather (Cunningham et al., 2019). Another study could reveal that younger drivers rated driving at night on average with M = 21 on a Likert scale ranging from 0 – I do not avoid at all to 100 – I absolutely avoid it (Moták et al., 2014). For driving at night in the rain the value even increases to M = 35. Hence, the probability of a decrease in drivers' comfort when driving at night should be set to 54% according to Cunningham et al. (2019).

Especially elderly drivers report problems with night driving and tend to avoid such situations (e.g., Cahour, 2008; Collia et al., 2003; Keall & Frith, 2004). Baldock et al. (2006) could show that 51% of the elderly drivers feel reasonably, not very or not at all confident driving at night. The percentage even increased to 70% for driving at night in the rain. 20% of the respondents reported to avoid driving at night sometimes, often or even always (Baldock et al., 2006). Another study revealed that on average 45% of the elderly drivers reported stress when driving at night, 35% even reported to avoid such situations (Hakamies-Blomqvist & Wahlström, 1998). Meng & Siren (2012) could show that on average 65% of the elderly drivers feel a little unpleasant, unpleasant or very unpleasant when driving at night. 40% of the respondents indicated to avoid driving at night if possible (Meng & Siren, 2012). Consequently, the probability of a decrease in drivers' comfort should be set higher for elderly drivers compared to younger drivers.

Bad weather conditions like heavy rain or snow as well as dense fog will also reduce visibility while driving (e.g., Andrey, 2010; Meyers et al., 2011). Furthermore, it is riskier to drive on wet and slippery or even icy roads (e.g., Hakamies-Blomqvist & Wahlström). Drivers seem to feel more unsafe while driving in bad weather conditions and tend to prefer automated driving. According to Cunningham et al. (2019) 54% of the respondents would prefer automated driving in bad weather and at night. Unfortunately, the respondents rated both situations combined in this study. Hence, it is not possible to say if the preferences differ for driving in bad weather and driving at night. Another study could show that participants rated driving when it is raining and while there is fog on the road as moderately stressful (Argandar et al., 2016). On a Likert scale ranging from 0 – not stressful to 4 – very stressful, rain was rated on average with 1.6 (corresponding to a probability of a decrease in drivers' comfort of 40%) and fog with 1.35 (corresponding to a probability of 34%). Another study revealed that respondents were even slightly anxious regarding driving in fog (Taylor & Pali, 2008). On a 5-point Likert scale ranging from 0 – no anxiety to 4 – extreme anxiety, driving in fog was rated on average 1.27 (corresponding to a probability of a decrease in drivers' comfort of 32%). Further findings could show that younger drivers rated driving in the rain on average with M = 21, driving in fog with M = 39 and driving in the snow with M = 42 on a Likert scale ranging from 0 – I do not avoid at all to 100 – I absolutely avoid it (Moták et al., 2014). Hence, take-over requests from manual to automated driving should be offered by the automated system. Regarding the probability of a decrease in drivers' comfort, we propose a starting value of 40% for rain and 33% for fog according to the findings of Argandar et al. (2016) and Taylor & Paki (2008). No literature was found regarding the probability of a decrease in drivers' comfort when it snows. Hence, it is suggested to set the probability for a decrease in drivers' comfort to 50%. This corresponds with the findings of Cunningham et al. (2019) revealing that 54% of the respondents would prefer automated driving in bad weather conditions or at night. It should be considered that participants were asked to rate driving in bad weather in general. It can be assumed that drivers' preference for automated driving increases if there is heavy rain, heavy snow or dense fog that impair the visibility strongly (e.g., Andrey, 2010). Hence, a take-over request should be offered with a slightly higher frequency when heavy rain, heavy snow or dense fog is detected.

It is expected that drivers who have more experience with driving in bad weather conditions (e.g., due to a higher driven mileage or a road-safety training) feel more competent in dealing with, for instance, icy roads or aquaplaning. . On the other hand, research could show that inexperienced drivers tend to show incomplete, slower or even missing behavioural adaptation in response to a changing situation (e.g., Mueller & Trick, 2012). Hence, it can be assumed that less experienced drivers will be more stressed in the respective situations leading to an impaired driving comfort. Therefore, it is suggested to set the probability for a decrease in drivers' comfort slightly higher for inexperienced drivers compared to experienced drivers. Nevertheless, it is not assumed that more experience (i.e., driver competence) has an influence on drivers' comfort (e.g., experienced stress) in such situations (e.g., Argandar, 2016). However, evidence could be found in the literature that elderly drivers feel more unsafe driving in bad weather compared to younger drivers (e.g., Moták et al., 2014). Baldock et al. (2006) could show that 56% of the elderly drivers feel reasonably, not very or not at all confident driving in the rain. 14% of the respondents reported to avoid driving in the rain sometimes, often or even always (Baldock et al., 2006). Another study revealed that on average 55% of the elderly drivers reported stress when driving on slippery roads, 45% even reported to avoid such situations (Hakamies-Blomqvist & Wahlström, 1998). Further, 20% of the elderly drivers expressed that they try to avoid driving in winter (Hakamies-Blomqvist & Wahlström, 1998). Meyers et al. (2011) could show that 31% of the respondents stated they do not drive on days with bad weather (either rain or snow) and 30% do not drive when the road conditions are poor (e.g., snow covered, icy, wet). Meng & Siren (2012) could show that on average 80% of the elderly drivers feel a little unpleasant, unpleasant or very unpleasant driving when slippery. 65% of the respondents indicated to avoid driving at night if possible (Meng & Siren, 2012). Consequently, the probability of a decrease in drivers' comfort should be set higher for elderly drivers compared to younger drivers.

Another situation that can be assumed to be relevant in terms of poor visibility is driving when the sun is very low on the horizon, for instance, during winter months (Pink et al., 2015). No literature was found regarding the probability of a decrease in drivers' comfort in such situations. Hence, it is suggested to set the probability of a decrease of drivers' comfort to 50%. Driver competence is not expected to have an influence on drivers' comfort in driving situations with low sun.

The probability to detect driving situations during nighttime is rated as very high based on current time of day, expected time for travel and existing knowledge regarding sunset times. The information is available several minutes (or even hours) in advance allowing for take-over requests that can be offered with an adequate preparation time for the driver. Detecting bad weather situations in advance is more challenging. Information from weather forecasts can be used to predict future weather conditions but sometimes they are quite inaccurate. Alternatively, current weather conditions can be used to predict future weather conditions (e.g., light fog increases the probability for heavy fog when driving through a dip) and the take-over request can be offered in advance. However, weather conditions can change rapidly (e.g., heavy rain starts immediately) and, hence, cannot be predicted in advance. Vehicle-to-Vehicle communication can help to increase the predictability of bad weather conditions. Vehicles driving in heavy rain, snow or dense fog can share this information with the rear traffic and, hence, the take-over request can be offered in advance for the approaching vehicles.

Situations with high complexity: Complex driving situations like situations with high traffic density (e.g., in large cities especially during the rush hour), complex intersections, roundabouts or multiple driving lanes with different destinations are potential uncomfortable driving situations (e.g., Guy et al., 2020; Frison et al., 2019; Payre et al., 2014; Cahour, 2008; Cottrell & Barton, 2011). Liljamo et al. (2018) could show that people living in densely populated urban areas had the highest positive attitudes towards automated driving. In their online survey, Cunningham et al. (2019) revealed that 33% somewhat agreed and 22% strongly agreed to the question if they want to use a fully automated car when traffic is congested leading to a probability for a decrease in drivers' comfort of 55%. Another study revealed that respondents were even slightly anxious regarding driving in heavy traffic (Taylor & Pali, 2008). On a 5-point Likert scale ranging from 0 – no anxiety to 4 – extreme anxiety, driving in fog was rated on average 1.24, corresponding to a probability of a decrease in drivers' comfort of 31%. Further findings could show that younger drivers rated driving during the rush hour on average with M = 46 on a Likert scale ranging from 0 – I do not avoid at all to 100 – I absolutely avoid it (Moták et al., 2014). According to Cunningham et al. (2019) and Taylor & Pali (2008), it is suggested to set the probability for a decrease in drivers' comfort to 40%.

It is expected that drivers who have more driving experience feel more competent in dealing with complex traffic situations compared to inexperienced drivers. Literature could reveal that inexperienced drivers tend to show inadequate visual scanning behaviour for potential obstacles or hazards (e.g., Underwood et al., 2002), a poorer hazard awareness (e.g., Borowsky, & Oron-Gilad, 2013) or prediction (e.g., Crundall, 2016), longer detection and reaction times to hazards (e.g., Whelan et al., 2004) as well as less anticipatory driving (e.g., Lehtonen et al., 2014) which can impair driving in complex traffic situations. Hence, it can be assumed that less experienced drivers will be more stressed in the respective situations leading to an impaired driving comfort. Therefore, it is suggested to set the probability for a decrease in drivers' comfort slightly higher for inexperienced drivers compared to experienced drivers. Especially elderly drivers reported difficulties in dealing with complex traffic situations. Baldock et al. (2006) could show that 35% of the elderly drivers feel reasonably, not very or not at all confident driving on high traffic roads. 44% of the respondents feel less confident driving during peak hours. 11% of the respondents and 21% respectively reported to avoid such situations sometimes, often or even always (Baldock et al., 2006). Another study revealed that on average 40% of the elderly drivers reported stress when driving during rush hour, 35% even reported to avoid such situations (Hakamies-Blomqvist & Wahlström, 1998). Meng & Siren (2012) could show that on average 55% of the elderly drivers feel a little unpleasant, unpleasant or very unpleasant when driving in dense traffic. 35% of the respondents indicated to avoid driving in dense traffic if possible (Meng & Siren, 2012). Hence, a take-over request should be offered in such situations with a higher frequency for elderly drivers.

The probability to detect complex driving situations is high when prior knowledge of the usual times with high traffic density (i.e., rush hour, all-day strongly frequented intersections in large cities) is used. The drivers can be informed one or two minutes before, for instance, reaching a complex intersection and, hence, the take-over request can be offered with enough preparation time for the driver.

Challenging driving manoeuvres: Certain driving manoeuvres like overtaking vehicles (e.g., Engelbrecht, 2013, Cahour, 2008), passing obstacles on the street (e.g., Engelbrecht, 2013), using narrow single-lane roads or narrow lanes in construction zones (e.g. de Vos et al.,

1998) and parking (e.g., Engelbrecht, 2013, Cahour, 2008) are quite challenging for some drivers. Engelbrecht (2013), for instance, could show that 15% of the drivers indicated that they do not like overtaking other vehicles, 25% stated that they do not like to pass obstacles on the street and 25% reported that they do not like to park their car. Hence, a potential decrease in drivers' comfort can be expected and the expected probability for a decrease in drivers' comfort should be set accordingly: 15% for overtaking manoeuvres, 25% for passing obstacles on the street and 25% for parking manoeuvres. No literature was found regarding the probability of a decrease in drivers' comfort when driving on narrow lanes. Hence, it is suggested to set the probability for a decrease in drivers' comfort to 50%.

Especially inexperienced drivers report that they do not like these kinds of driving manoeuvres (Engelbrecht, 2013). Nearly twice as much inexperienced drivers indicated negative experiences with such driving manoeuvres compared to more experienced drivers. Hence, a take-over request should be offered in such situations with a higher frequency for inexperienced drivers. Further, research could show that elderly drivers might accept take-over requests in such situations with a higher probability. Baldock et al. (2006) could show that 70% of the elderly drivers feel reasonably, not very or not at all confident parking their vehicle. 37% of the respondents reported to avoid such situations sometimes, often or even always (Baldock et al., 2006). Another study revealed that on average 18% of the elderly drivers reported stress when parking and 20% reported stress when overtaking other vehicles (Hakamies-Blomqvist & Wahlström, 1998). Meng & Siren (2012) could show that on average 37% of the elderly drivers feel a little unpleasant, unpleasant or very unpleasant when overtaking other vehicles. 27% of the respondents indicated to avoid overtaking other vehicles if possible (Meng & Siren, 2012). Consequently, the probability of a decrease in drivers' comfort should be set higher for elderly drivers compared to younger drivers.

The probability to detect such situations in advance is rated as medium to high. Prior knowledge regarding narrow streets or construction zones on the road ahead can be used to offer a take-over request several minutes in advance. The parking manoeuvre can also be predicted several minutes or even hours in advance by using information regarding the destination of the trip. The take-over request can be combined with possibilities for parking lot reservations (e.g., Wang & He, 2011). The predictability of manoeuvres like overtaking vehicles or passing obstacles on the street is high as soon as the vehicle's sensors detect the slow driving vehicle or obstacle ahead. The moment of detection in relation to the time to reach the vehicle or obstacle determines how much in advance a take-over request can be offered.

Situations causing high levels of uncertainty: Some drivers reported to have difficulties in situations, which cause a higher level of uncertainty like driving in unknown environments (e.g., in foreign cities, unfamiliar routes or even countries with differing traffic rules). In a study of McKenna et al. (1991) nearly 100 car drivers rated their own drivers' skill in different driving scenarios on 11-point Likert scale ranging from 0 – very poor to 10 – very good. Navigating and driving in unfamiliar areas was rated worse compared to other driving manoeuvres like overtaking, parking or busy town driving. No literature was found regarding the probability of a decrease in drivers' comfort when driving in unfamiliar areas. Hence, it is suggested to set the probability of a decrease of drivers' comfort to 50%.

Especially elderly people prefer to drive common routes and avoid driving unknown routes (e.g., Meyers et al., 2011). Research could show that on average 55% of the elderly drivers feel a little unpleasant, unpleasant or very unpleasant when driving in unknown places or on unknown routes (Meng & Siren, 2012). 25% of the respondents indicated to avoid such situations if possible. Another study revealed that on average 35% of the elderly drivers reported stress when driving in unfamiliar surroundings (Hakamies-Blomqvist & Wahlström, 1998). Consequently, the probability of a decrease in drivers' comfort should be set higher for elderly drivers compared to younger drivers.

The probability to detect unknown routes is rated as very high based on the information from the navigation system (e.g., planned route, destination of the trip). The information and, hence, the take-over request can be provided several minutes or even hours in advance. Redirections on familiar routes that force the driver to take a potentially unknown alternative route can also be predicted several minutes in advance based on traffic news or information from vehicle-to-vehicle communication.

Demanding or highly prioritized non-driving related tasks: When automated driving becomes available, the execution of non-driving related tasks (NDTRs) becomes more and more attractive for drivers (e.g., Kyriakidis et al., 2015). Further, drivers rated situations with the need of distributed attention (e.g., restless children, conversations with passengers) as uncomfortable (e.g., Cahour, 2008). When asked if automated driving will allow to spend time on other activities, 33% of the respondents somewhat agreed and 14% strongly agreed (Cunningham et al., 2019). The most likely activities the respondents would like to be engaged in (rated with somewhat likely or extremely likely) are observing the scenery (70%), interaction with passengers (70%), eating / drinking (60%), using personal devices like mobile phones (55%), resting (45%) or even sleeping (20%), reading (40%) or working 30%). In an online survey 1.000 participants were asked for perceived advantages of automated driving (Cyganski et al., 2014). They rated different activities on a 6-point Likert scale ranging from 1 – not at all true to 6 – absolutely true regarding the question if the possibility to perform these activities is a particular advantage of automated vehicles. Results could show that participants rated the possibility to enjoy the trip and the scenery (70%), interact with passengers (60%), use their smartphone and surf on the internet (40%), watch movies (30%), relax or even sleep (40%) and work (30%) as particular advantages of automated vehicles. In the study from Wadud & Huda (2019) participants were asked how they would spend their time when driving an automated vehicle. The indicated values for different activities were smaller compared to the findings from Cunningham et al. (2019) and Cyganski et al. (2014). Participants stated they would like to observe the scenery (32%), interact with passengers (28%), eat / drink (16%), use personal devices like mobile phones (28%), sleep (20%), read (21%) or work (20%). Further, the results from Wadud & Huda (2019) indicate that the purpose of the trip plays an important role. 27.6% of the commuters and 45% of the business travellers reported to spend some time working during the outbound trip, but had fewer intentions to work during return trips. Hence, it can be expected that in situations with high NDRT demands (e.g., incoming phone calls, voice messages or mails; interaction with passengers in the car especially with children) or when the priority of the NDTR is rated higher compared to the manual driving task (e.g., watching the nice road scenery, eating, working, watching movies, listening to podcasts, reading a book, sleeping), drivers would prefer automated driving. The probability for an expected decrease in drivers' comfort according to the findings of Cunningham et al. (2019), Cyganski et al. (2014) and Wadud & Huda (2019) should be set to:

- 50% for driving in a nice scenery (side note: hence, the scenery can also be watched peripherally while driving manually, a decrease in comfort is not necessarily expected but an increased preference for automated driving)
- 45% for interaction with passengers,
- 40% for eating / drinking,

- 40% for incoming phone calls or text messages,
- 33% on a trip home from work, trips at night or very early in the morning when the driver might want to rest,
- 27% for leisure trips when the driver might, for instance, want to read or watch a movie,
- 25% for work-related trips when the driver might want to work.

Especially younger drivers stated that they want to execute NDRTs during automated driving (e.g., Frison et al., 2018; Wadud & Huda, 2019). Hence, a take-over request should be offered in such situations with a higher frequency for younger drivers.

The probability to detect the relevant situations in advance depends on the NDRT in question. Incoming phone calls are impossible to predict in advance. In case the driver receives a phone call, the automated system can provide a take-over from manual to automated driving immediately but with no preparation time for the driver. Incoming voice or text messages (e.g., emails) are also impossible to predict. Nevertheless, the automated system can hold back the messages, inform the driver about the messages and provide a take-over request. The driver can choose if and when he / she wants to drive in automated mode and deal with the messages. The probability to detect passenger interaction is rated as quite high. If there is no passenger in the car, the probability of passenger interaction is zero, hence no take-over request is needed. Accordingly, the presence of a passenger in the car increases the possibility of passenger interaction considerably. Nevertheless, it could not be predicted when exactly a conversation with the passengers is initiated or when interactions with children in the car becomes necessary. One possibility is to use voice or camera recordings to identify situations that require a take-over request (e.g., when a continuous conversation is detected or when the child becomes restless). The take-over request will be offered immediately in the relevant situation but not in advance. This is the same for situations in which the driver wants to eat or drink something. The system can offer the take-over request as soon as the driver reaches for food or drinks. If the system detects that the driver brings food or drinks with him / her and stores it in reachable distance to the driver seat, the take-over request can be offered in advance. However, the system does not know if the driver plans to eat or drink immediately or maybe stored the food for later usage. Whereas the identification of a nice road scenery might be possible in advance based on geographical information, the identification of drivers' preference for NDRT execution is challenging. One indication that can be used is the purpose of the trip. During working trips, the preference for automated driving enabling the driver to work might be higher compared to leisure trips. On the other hand, during leisure trips the preference for automated driving might also increase in favour of watching a movie or reading a book. The probability, however, might be reduced when other passengers are present or the travel time is too short. Hence, both variables – presence of passengers and travel time – need to be taken into account for the decision to offer automated driving. It is assumed that the preference for the execution of NDRTs during automated driving is highest if there are no passengers present and the travel time is at least 20 minutes (e.g., minimum time to watch one episode of a TV series, AT&T, 2017; preferred length of a podcast episode in Germany, G+J, 2019). Another indicator might be the daytime. In the evening hours or during the night hours the driver might prefer automated driving enabling him / her to rest or even sleep for some hours. Hence, a take-over request should be offered especially when the automated system detects that the driver is getting fatigued.

Impaired driver state: A very important use case for automated driving is to compensate for performance losses due to an impaired driver state (e.g., fatigue, frustration, stress, distraction). As drivers stated that driving safety has the highest priority, it is assumed that they would prefer automated driving in case an impaired driving state is detected in favour of driving safety. The most relevant use case seems to be the driver state fatigue. Cunningham et al. (2019) could show that 36% of the respondents would somewhat agree and 33% would strongly agree to use an automated car when they are fatigued. Hence, it is suggested to set the probability of a decrease of drivers' comfort due to fatigue to 70%. For the other driving states, it is suggested to start with a probability of 50% for a decrease of drivers' comfort.

Especially elderly people report discomfort when driving fatigued. Research could show that on average 70% of the elderly drivers feel a little unpleasant, unpleasant or very unpleasant when driving fatigued (Meng & Siren, 2012). 75% of the respondents indicated to avoid such situations if possible. Another study revealed that on average 70% of the elderly drivers reported stress when driving fatigued (Hakamies-Blomqvist & Wahlström, 1998). Consequently, the probability of a decrease in drivers' comfort should be set slightly higher for elderly drivers compared to younger drivers.

The probability to detect impaired driver states is rated as high. For more information see Chapter 2 (distraction) and Chapter 3 (fatigue). Data from respective driver state monitoring systems can be used to detect changes in the driver state (e.g., the driver gets fatigued, stressed or distracted) and the automated system can react immediately by offering a take-over from manual to automated driving. Although, context factors (e.g., driving at night) can be used to estimate the probability of a change in driver state (e.g., the driver might get fatigued), detecting a change in drivers' state in advance (e.g., before the driver gets fatigued) is hardly possible. However, the driver state monitoring system can detect first signs of a change in driver state and the automated system can react before the level of impairment will affect the driving performance. Hence, automated driving can be offered several seconds or even minutes in advance, before the driver state reaches a critical level (e.g., the driver gets too fatigued to drive safely).

Longer and monotonous trips: Long duration car driving without stops is expected to reduce drivers' comfort due to, for instance, the long time sitting in the driver seat and the vibrations related to the characteristics of the road surface and the vehicle (e.g., Cucuz, 1994). It could be revealed that drivers' discomfort increased over a 150 min drive from no discomfort to noticeable discomfort (e.g., el Falou et al., 2003). By the end of the trial participants rated their discomfort with 2.23 on a 11-point Liktert scale ranging from 0 – no discomfort to 10 – high discomfort. Hence, the probability for a decrease in comfort should be set to 20% after a 2.5 h time period spent driving manually. An influence of drivers' competence is not expected. Long duration car driving can be well detected and a take-over request from manual to automated driving can be offered several minutes in advance.

(Regular) longer trips (e.g., commutes) are also rated as unpleasant (e.g., Kahnemann & Krueger, 2006), because of the loss of potentially productive time (e.g., Cyganski et al., 2014). Kahnemann & Krueger (2006) could show that participants rated evening commutes on average with 2.09 and morning commutes even with 2.77 on a 6-point Likert scale ranging from 0 – not at all enjoyable to 6 – very much enjoyable. Hence, it is suggested to set the probability for a decrease in driving comfort to 65% for morning commutes and to 54% for evening commutes (on average 60% for commutes in general). Further findings could show that younger drivers rated driving long distances on average with *M* = 18 on a Likert scale ranging from 0 – I do not avoid at all to 100 – I absolutely avoid it (Moták et al., 2014). Longer trips of elderly drivers are more likely not work-related. Research could show that on average

35% of the elderly drivers feel a little unpleasant, unpleasant or very unpleasant when driving long distances (Meng & Siren, 2012). 35% of the respondents indicated to avoid such situations if possible. Another study revealed that on average 35% of the elderly drivers reported stress when driving long distances (Hakamies-Blomqvist & Wahlström, 1998). Hence, it is suggested to set the probability for a decrease in driving comfort to 35% to longer trips which are not work-related (i.e., morning or evening commutes). The probability to detect work-related situations is rated as high using the information from the navigation system (e.g., working place as starting point or destination) combined with daytime (e.g., morning vs. evening). Longer trips in general are also rated as well predictable using the information from the navigation system. The take-over request can be offered several minutes in advance (e.g., as soon as the driver enters the car).

Monotonous driving (e.g., prolonged highway driving with little surrounding traffic) is expected to decrease driving comfort (Cunningham et al., 2019; Payre et al., 2014). 37% of the respondents somewhat agreed and 25% strongly agreed that they want to use an automated car when driving is boring and monotonous. Another study revealed that 67% of the respondents expressed the intention to use automated vehicles on long freeway journeys (Sommer, 2013 as cited in Kyriakidis et al., 2015). Hence, the probability for a decrease in driver comfort should be set to 65%. An influence of drivers' competence is not expected. The probability for detecting situations characterized by monotonous driving (e.g., prolonged highway driving with little surrounding traffic) is rated as high and the take-over request can be given several minutes in advance (e.g., before entering a barely used highway).

Other situations: Automated driving is expected to increase fuel or energy efficiency (e.g., Cunningham et al., 2019). Themann et al. (2015) could show that an optimization of fuel efficiency is possible without influencing the travel times. Another study revealed that 29% of the respondents somewhat agreed and 10% strongly agreed to the statement that automated vehicles would consume less fuel (Golbabaei et al., 2020). Especially people who are concerned regarding the environment (e.g., concerns about the impact of pollutant emissions on global warming) might want to use automated driving in situations in which the automated driving style can increase fuel efficiency. Shabanpour et al. (2018) could show that 80% of the drivers rated better fuel efficiency as important. Hence, it is suggested to set the probability for a decrease in drivers' comfort to 80% for situations in which the automated system can optimize fuel or energy efficiency without influencing the travel time. An influence of drivers' competence is not expected. The probability for detecting situations in which fuel or energy efficiency can be optimized by the system is rated as high using traffic news and prior knowledge regarding the road ahead. The take-over request can be given some seconds in advance.

4.2.3.2 Potentially uncomfortable situations in automated driving

Situations with higher risks of motion sickness: Different on-road studies revealed that motion or car sickness might become a problem in automated driving (e.g.,Diels et al., 2016; Krause et al., 2016). Especially when engaged in NDRTs such as working, reading, watching movies or playing video games the probability of motion sickness increases (e.g., Wadud & Huda, 2019; Mühlbacher et al., 2020). Road characteristics (e.g., driving through curves) can further increase the risk for motion sickness (e.g., Mühlbacher et al., 2020). Based on the results of several on-road studies, Diels et al. (2017) suggested that motions sickness while driving in automated mode may be experienced by 50-75% of the drivers. Hence, it is suggested to set the probability for a decrease of drivers' comfort due to motion sickness to 50% if approaching a situation potentially provoking simulation sickness (e.g., curvy roads) and to 75% in case the driver is engaged in NDRTs like working or reading when approaching such situations. A take-over request from automated to manual driving should be offered in such situations. An influence of drivers' competence is not expected.

The probability for detecting potential situations in which the risk for motion sickness increases is rated as high using prior knowledge regarding the road ahead and information regarding NDRT-involvement of the drivers. The take-over request can be given some seconds or even minutes in advance. It needs to be considered that motion sickness can also occur without specific road characteristics provoking it or without engagement in NDRTs. In such cases it is important to intervene as soon as first signs of motion sickness are detected.

Situations with high demands on communication with other road users: Previous research could show that most (54%) of drivers' reported uncertainties during automated driving were related to awareness, behaviour or intentions of other road users (Grahn et al., 2020). Techer et al. (2019) revealed that participants were concerned that the automated vehicle will not interact with other road users in an appropriate manner. Hence, the participants stated that they would prefer to handle such situations themselves (Techer et al., 2019). Therefore, situations like pedestrians crossing the street, cyclists on the road or other road users who want to turn into the street are good examples for situations that might cause a decrease in drivers' comfort when approaching in automated driving mode. No literature was found regarding the probability of a decrease in drivers' comfort in such situations. Hence, it is suggested to set the probability of a decrease of drivers' comfort to 50%. An influence of drivers' competence is not expected.

The probability for detecting situations demanding high levels of communication with other road users is rated as quite high. As most of these situations might occur rather in cities (e.g., compared to highways), take-over from automated to manual driving should be offered when entering a village, small town or big city in automated driving mode. The take-over request can be given some seconds or even minutes in advance.

Driving under time pressure: In the study of Techer et al. (2019) it could be revealed that the number of driver-initiated take-over actions are higher when experiencing time pressure. According to the authors, time pressure enhanced drivers' need to take-over the driving task and handle the situation in manual driving mode. No literature was found regarding the probability of a decrease in drivers' comfort in such situations. Hence, it is suggested to set the probability of a decrease of drivers' comfort to 50%. An influence of drivers' competence is not expected.

The probability for detecting situations in which drivers are driving under time pressure might be possible by using information from the drivers' appointment calendar and the navigation system (e.g., planned route, destination of the trip, estimated driving time). In this case, the take-over request can be given some seconds or even minutes in advance.

Purpose of the trip: Wadud & Huda (2019) could show that the purpose of the trip plays an important role for the drivers' decision to engage in NDRTs. Nevertheless, some participants (16%) stated that they won't do anything during automated driving possibly leading to boredom and fatigue (e.g., Saxby et al., 2013; Coughlin et al., 2011). To prevent a decrease in drivers' comfort due to boredom and to prevent the driver from getting fatigued, a take-over request from automated to manual driving can be offered if the system detects that the driver is not engaged in any task and not asleep for a longer time period. No literature was found regarding the probability of a decrease in drivers' comfort in such situations. Hence, it is suggested to set the probability of a decrease of drivers' comfort to 50%. An influence of drivers' competence is not expected.

The probability for detecting situations in which the driver becomes bored or fatigued due to automated driving is rated as high using the respective driver state detection systems. The take-over request can be given some seconds or even minutes in advance.

More challenging are situations in which the driver wants to drive manually because of pleasure and joy associated with manual driving and, hence, wants a take-over from automated to manual driving. Such situations are hard to predict in advance. It is suggested to learn from driver-initiated take-overs during a trip (e.g., the driver initiates a take-over from automated to manual driving on a specific section of the route) and to offer a take-over request when approaching the same section in the future.

Further, it needs to be taken into account that drivers' ability to successfully manage difficult driving situations (e.g., handling complex intersections) might be impaired over time (e.g., driver loses or never develops essential driving skills) when the automated driving system always handles the respective situations. Therefore, solutions need to be developed to maintain drivers' driving skills (e.g., regular training) as long as system- or driver-initiated take-over manoeuvres from automated to manual driving can occur,

Side note: Situations that cannot be handled by the automated system: Situations in which the automated system cannot reliably handle the driving situation (e.g., due to missing lane markings without a lead vehicle, bad weather conditions, high traffic density and complexity) might impair driving safety and, hence, are expected to cause a decrease in drivers' comfort (e.g., Frison et al., 2018). In such cases a take-over from automated to manual driving is crucial in terms of safety but also in terms of drivers' comfort. Hence, it is suggested to set the probability of a decrease of drivers' comfort to 100%. An influence of drivers' competence is not expected. The probability for detecting situations in which the automated system is no longer capable to handle the driving situation is rated as high. The chances and limitations of current automated driving systems will be covered in Deliverable 1.3 "Degraded automation performance". Therefore, this topic will not further be discussed in this chapter.

4.2.4 Conclusions and implications

Overall, offline / long-term prediction of drivers' comfort based on prior knowledge seems to be a promising approach to detect potentially uncomfortable driving situations in advance, to prevent a decrease in drivers' comfort and to consider drivers' preferences for automated vs. manual driving in future driving situations. Therefore, it might be a useful addition to real time comfort assessment enabling the detection of a decrease in drivers' comfort as soon as it occurs. Nevertheless, there is much more research needed to enable stable and reliable real time detection of discomfort on individual level based on physiological data (e.g., facial expressions). Although, the development and testing of respective algorithms is in the scope of the MEDIATOR project, the offline prediction approach will also be considered. Based on prior knowledge, a variety of potentially uncomfortable driving situations could be collected (e.g., driving in traffic jams, driving at night, incoming text messages). Based on the results of existing studies regarding, for instance, drivers' experienced stress, uncertainty in competence, anxiety and preferences, a probability for a decrease in drivers' comfort could be derived. The information is summarized in Table 12 and can be implemented in the Mediator system. It is suggested to focus especially on the situations with the highest probability of a decrease in drivers' comfort (e.g., car following scenarios like traffic jams, driving at night, driving when fatigued, regular commutes or monotonous driving, situations with higher risks of motion sickness especially while engaged in NDRTs). By using information from other sources such as in-vehicle sensors, traffic news, weather forecasts or vehicle-to-vehicle-communication potentially uncomfortable situations can be detected in advance and take-over request can be offered. Further, the frequency of take-over requests can be adapted to individual preferences of the drivers helping to personalize the Mediator system. Nevertheless, it needs to be stated that the offline comfort prediction approach has its limits.

First, only well-defined situations that are known beforehand will be implemented in the system (e.g., traffic jams, complex intersections, incoming messaged while driving, driving at night). Other situations which provoke a decrease in comfort for specific divers will not be covered (e.g., if a driver feels uncomfortable driving on a certain road because he had an accident on this road in the past). For such situations, real-time comfort assessment is needed to intervene as soon as a decrease in drivers' comfort is detected.

Second, unexpected changes in drivers' comfort (e.g., due to a sudden headache, behaviour of other road users, bad news on the radio) cannot be covered by the offline prediction approach. Again, real-time comfort assessment is needed to deal with such situations.

Third, the initially set frequency of take-over requests (i.e., a-priory default state) is based on averaged results of prior studies. Hence, it is likely that the a-priory defined frequency of take-over requests will not meet all individual preferences at the beginning. Therefore, it is important to offer the flexibility for drivers to initiate take-over actions in case the system does not react to a situation, which leads to a decrease in drivers' comfort. It might be necessary for the driver to reject or initiate take-over actions more often when he / she starts to interact with the Mediator system until the system can learn the driver's preferences. It will be crucial to help the driver to overcome this initial phase without impairing his / her trust in the system or annoy him / her by too many take-over requests.

Fourth, with the offline detection approach, detecting situations in which the driver does not wish to get a take-over request is quite challenging. In case the driver is sleeping or engaged with a NDRT, no take-over requests should be sent to the driver even if approaching a section of the route, which the driver typically enjoys to drive manually (except for situations where the automated system is not capable to drive anymore). However, situations in which the driver does not want to receive a take-over request from manual to automated driving because he / she is enjoying the drive is hardly predictable. For such circumstances, it might be reasonable to implement the probability to switch off comfort-relevant take-over requests completely for a certain trip. Nevertheless, offline prediction of drivers' comfort is a feasible and promising approach. In a next step, the information regarding drivers' preferences need to be combined with knowledge regarding the limits of currently available automated driving system to derive situations in which a take-over by the automated system is possible. A promising use case in this regard might be the traffic jam on highways.

4.3 References

Abraham, H., Lee, C., Brady, S., Fitzgerald, C., Mehler, B., Reimer, B., & Coughlin, J. F. (2016). Autonomous vehicles, trust, and driving alternatives: A survey of consumer preferences. *Massachusetts Inst. Technol, AgeLab, Cambridge*, *1*, 16.

Anderson, S. J., & Holliday, I. E. (1995). Night driving: effects of glare from vehicle headlights on motion perception. *Ophthalmic and Physiological Optics*, *15*(6), 545-551. doi: 10.1016/0275-5408(95)00070-T

Andrey, J. (2010). Long-term trends in weather-related crash risks. *Journal of Transport Geography*, *18*(2), 247-258. doi: 10.1016/j.jtrangeo.2009.05.002

Argandar, G. D., Gil, F. T., & Berlanga, J. F. (2016). Measuring situations that stress Mexicans while driving. *Transportation research part F: traffic psychology and behaviour*, *37*, 154-161. doi: 10.1016/j.trf.2015.12.014

AT&T. (31. Dezember, 2017). Durchschnittliche Episodenlänge ausgewählter Serien im Jahr 2017 (in Minuten) [Graph]. In Statista. Retreaved from https://de.statista.com/statistik/daten/studie/990558/umfrage/episodenlaenge-ausgewaehlter-serien/

Baldock, M. R. J., Mathias, J. L., McLean, A. J., & Berndt, A. (2006). Self-regulation of driving and its relationship to driving ability among older adults. *Accident Analysis & Prevention*, *38*(5), 1038-1045. doi: 10.1016/j.aap.2006.04.016

Bayam, E., Liebowitz, J., & Agresti, W. (2005). Older drivers and accidents: A meta analysis and data mining application on traffic accident data. *Expert Systems with Applications, 29(3)*, 598-629. doi: 10.1016/j.eswa.2005.04.025.

Bazilinskyy, P., Kyriakidis, M., & de Winter, J. (2015). An international crowdsourcing study into people's statements on fully automated driving. *Procedia Manufacturing*, *3*, 2534-2542. doi: 10.1016/j.promfg.2015.07.540

Becker, F., & Axhausen, K. W. (2017). Literature review on surveys investigating the acceptance of automated vehicles. *Transportation*, *44*(6), 1293-1306. doi: 10.1007/s11116-017-9808-9

Beggiato, M. (2015). Changes in motivational and higher-level cognitive processes when inter-acting with in-vehicle automation (Doctoral Dissertation). University of Technology Chemnitz. Retrieved from http://nbnresolving.de/urn:nbn:de:bsz:ch1-qucosa-167333.

Beggiato, M., Hartwich, F., & Krems, J. (2018). Using Smartbands, Pupillometry and Body Motion to Detect Discomfort in Automated Driving. *Frontiers in Human Neuroscience, 12*, 3138. doi:10.3389/fnhum.2018.00338.

Beggiato, M., Hartwich, F., & Krems, J. (2019). Physiological correlates of discomfort in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *66*, 445-458. doi: 10.1016/j.trf.2019.09.018

Bellem, H., Thiel, B., Schrauf, M., & Krems, J. F. (2018). Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits. *Transportation Research Part F: Traffic Psychology and Behaviour*, 55, 90–100. doi:10.1016/j.trf.2018.02.036.

Borowsky A., & Oron-Gilad, T.(2013). Exploring the Effects of Driving Experience on Hazard Awareness and Risk Perception via Real-Time Hazard Identification, Hazard Classification, and Rating Tasks. *Accident Analysis, & Prevention, 59*, 548–565.

Bryant, D., & Howard, A. (2019). A Comparative Analysis of Emotion-Detecting AI Systems with Respect to Algorithm Performance and Dataset Diversity. In: *AIES '19 Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 377–382. ACM New York, NY.

Burger, J. M., & Cooper, H. M. (1979). The desirability of control. *Motivation and Emotion, 3(4)*, 381–393. doi:10.1007/BF00994052

Cahour, B. (2008, January). Discomfort, affects and coping strategies in driving activity. In *Proceedings of the 15th European conference on Cognitive ergonomics: the ergonomics of cool interaction* (pp. 1-7). doi: 10.1145/1473018.1473046

Christoph, M., Cleij, D., Ahlström, C., Bakker, B., Beggiato, M., Borowsky, A, … Van Nes, C.N. (2019). *Mediating between human driver and automation: state-of-the art and knowledge gaps*. Deliverable D1.1. Retrieved from https://MEDIATORproject.eu/deliverables

Collia, D. V., Sharp, J., & Giesbrecht, L. (2003). The 2001 national household travel survey: A look into the travel patterns of older Americans. *Journal of safety research*, *34*(4), 461-470. doi: 10.1016/j.jsr.2003.10.001

Cottrell, N. D., & Barton, B. K. (2013). The role of automation in reducing stress and negative affect while driving. *Theoretical Issues in Ergonomics Science*, *14*(1), 53-68. doi: 10.1080/1464536X.2011.573011

Coughlin, J. F., Reimer, B., & Mehler, B. (2011). Monitoring, managing, and motivating driver safety and well-being. *IEEE Pervasive Computing*, *10*(3), 14-21. doi: 10.1109/MPRV.2011.54

Crundall, D. (2016). Hazard prediction discriminates between novice and experienced drivers. *Accident Analysis & Prevention, 86,* 47-58.

Cucuz, S. (1994). Evaluation of ride comfort. *International Journal of Vehicle Design*, *15*(3-5), 318-325. doi: 10.1504/IJVD.1994.061864

Cunningham, M. L., Regan, M. A., Ledger, S. A., & Bennett, J. M. (2019). To buy or not to buy? Predicting willingness to pay for automated vehicles based on public opinion. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 418-438. doi: 10.1016/j.trf.2019.08.012

Cyganski, R., Fraedrich, E., & Lenz, B. (2015). Travel-time valuation for automated driving: A use-case-driven study. In *Proceedings of the 94th Annual Meeting of the TRB*.

Dalbert, C. (1999). Die Ungewißheitstoleranzskala: Skaleneigenschaften und Validierungsbefunde. *Hallesche Berichte zur pädagogischen Psychologie*: Vol. 1. Halle (Saale): Martin-Luther-Univ. Halle-Wittenberg.

de Vos, A. P., Hoekstra, W., & Pieterse, M. T. J. (1998). *Automativ vehicle guidance (AVG): Effects of lane width and physical separation on driver comfort*. Soesterberg (NL), TNO Human Factors Research Institute.

del Campo, I., Asua, E., Martínez, V., Mata-Carballeira, Ó., & Echanobe, J. (2018). Driving style recognition based on ride comfort using a hybrid machine learning algorithm. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 3251-3258). IEEE. doi: 10.1109/ITSC.2018.8569722

Diels, C., & Thompson, S. (2017). Information expectations in highly and fully automated vehicles. In *International conference on applied human factors and ergonomics* (pp. 742-748). Springer, Cham. doi: 10.1007/978-3-319-60441-1_71

Diels, C., Bos, J.E. Hottelart, K., Reilhac, P. (2016). *Motion Sickness in Automated Vehicles: The Elephant in the Room*. In: Meyer, G., Beiker, S. (eds) Road Vehicle Automation 3. Springer Verlag, Berlin. doi: 10.1007/978-3-319-40503-2_10

Diels, C., Erol, T., Kukova, M., Wasser, J., Cieslak, M., Payre, W., ... & Bos, J. (2017). Designing for comfort in shared and automated vehicles (SAV): a conceptual framework. In *1st International Comfort Congress*, 7-8 June, Salerno, Italy.

Domeyer, J., Alsaid, A., Liu, S.-Y., & Lee, J. D. (2019). Passenger Emotional Response Type and Timing during Automated Vehicle Intersection Negotiation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63*(1), 2061–2065. doi:10.1177/1071181319631226

Eby, D. W., Trombley, D. A., Molnar, L. J., & Shope, J. T. (1998). *The Assessment of Older Driver's Capabilities: A Review of the Literature*. UMTRI-98-24. Ann Arbor, MI: University of Michigan Transportation Research Institute.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. Cambridge (Mas.): Malor Books.

Ekman, P., Hager, J. C., & Friesen, W. V. (2002). *Facial action coding system: The manual*. Salt Lake City: Research Nexus.

El Falou, W., Duchêne, J., Grabisch, M., Hewson, D., Langeron, Y., & Lino, F. (2003). Evaluation of driver discomfort during long-duration car driving. *Applied Ergonomics*, *34*(3), 249-255. doi: 10.1016/S0003-6870(03)00011-5

Elbanhawi, M., Simic, M., & Jazar, R. (2015). In the passenger seat: Investigating ride comfort measures in autonomous cars. *IEEE Intelligent Transportation Systems Magazine, 7*(3), 4–17. doi:10.1109/MITS.2015.2405571.

Engelbrecht, A. (2013). *Fahrkomfort und Fahrspaß bei Einsatz von Fahrerassistenzsystemen*. disserta Verlag.

ERTRAC. (2019). *Connected Automated Driving Roadmap*. European Road Transport Research Advisory Council. Retrieved from https://www.ertrac.org/uploads/documentsearch/id57/ERTRAC-CAD-Roadmap-2019.pdf.

Frison, A. K., Aigner, L., Wintersberger, P., & Riener, A. (2018, September). Who is generation A? Investigating the experience of automated driving for different age groups. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 94-104). doi: 10.1145/3239060.3239087

Frison, A. K., Wintersberger, P., Liu, T., & Riener, A. (2019, March). Why do you like to drive automated? a context-dependent analysis of highly automated driving to elaborate requirements for intelligent user interfaces. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (pp. 528-537). doi: 10.1145/3301275.3302331

Furgale, P., Schwesinger, U., Rufli, M., Derendarz, W., Grimmett, H., Mühlfellner, P., ... & Schmidt, B. (2013, June). Toward automated driving in cities using close-to-market sensors: An overview of the v-charge project. In *2013 IEEE Intelligent Vehicles Symposium (IV)* (pp. 809-816). IEEE. doi: 10.1109/IVS.2013.6629566

G+J. (25. Juli, 2019). *Was ist für Sie die optimale Länge eines Podcasts*? [Graph]. In Statista. Retrieved from https://de.statista.com/statistik/daten/studie/1032178/umfrage/optimale-laenge-von-podcasts-in-deutschland/

Gabaude, C., Marquié, J. C., & Obriot-Claudel, F. (2010). Self-regulatory driving behaviour in the elderly: relationships with aberrant driving behaviours and perceived abilities. *Le travail humain*, *73*(1), 31-52. doi: 10.3917/th.731.0031

Gao, H., Yuce, A., & Thiran, J.-P. (2014). Detecting emotional stress from facial expressions for driving safety. In: *IEEE International Conference on Image Processing (ICIP)*, 5961–5965. IEEE. doi: 10.1109/ICIP.2014.7026203.

Golbabaei, F., Yigitcanlar, T., Paz, A., & Bunker, J. (2020). Individual Predictors of Autonomous Vehicle Public Acceptance and Intention to Use: A Systematic Review of the Literature. *Journal of Open Innovation: Technology, Market, and Complexity*, *6*(4), 106. doi: 10.3390/joitmc6040106

Grahn, H., Kujala, T., Silvennoinen, J., Leppänen, A., & Saariluoma, P. (2020). Expert Drivers' Prospective Thinking-Aloud to Enhance Automated Driving Technologies–Investigating Uncertainty and Anticipation in Traffic. *Accident Analysis & Prevention*, *146*, 105717. doi: 10.1016/j.aap.2020.105717

Guy, C. L., Ariel, T., Asaf, D., & Tal, O. G. (2020, September). Identifying Sources of Discomfort in Various Road Events While Riding Automated Vehicles. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)* (pp. 1-5). IEEE. doi: 10.1109/ICHMS49158.2020.9209461

Hakamies-Blomqvist, L., & Wahlström, B. (1998). Why do older drivers give up driving? *Accident Analysis & Prevention*, *30*(3), 305-312. doi: 10.1016/S0001-4575(97)00106-1

Harris, W. (1977). Fatigue, circadian rhythm, and truck accidents. In *Vigilance,* 133-146. Springer, Boston, MA. doi: 10.1007/978-1-4684-2529-1_8

Hartwich, F., Beggiato, M., & Krems, J. F. (2018). Driving comfort, enjoyment and acceptance of automated driving–effects of drivers' age and driving style familiarity. *Ergonomics*, *61*(8), 1017-1032. doi: 10.1080/00140139.2018.1441448

Hergeth, S., Lorenz, L., & Krems, J. F. (2016). Prior familiarization with takeover requests affects drivers' takeover performance and automation trust. *Human Factors, 59*(3), 457–470. doi:10.1177/0018720816678714

Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., & Donohew, R. L. (2002). Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences, 32*(3), 401–414. doi:10.1016/S0191-8869(01)00032-0

Ihme, K., Dömeland, C., Freese, M., & Jipp, M. (2018a). Frustration in the face of the driver. Interaction Studies, 19(3), 487–498. doi:10.1075/is.17005.ihm

Ihme, K., Unni, A., Zhang, M., Rieger, J. W., & Jipp, M. (2018b). Recognizing Frustration of Drivers From Face Video Recordings and Brain Activation Measurements With Functional Near-Infrared Spectroscopy. *Frontiers in Human Neuroscience, 12*, 669. doi:10.3389/fnhum.2018.00327

Ikeda, K., Endo, A., Minowa, R., Narita, T., & Kato, H. (2018). Ride comfort control system considering physiological and psychological characteristics: effect of masking on vertical vibration on passengers. *Actuators, 7*(3), 42. doi:10.3390/act7030042.

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of Economic Perspectives*, *20*(1), 3-24. doi: 10.1257/089533006776526030

Kaiser, H.J., & Oswald, W.D. (2000). Autofahren Im Alter – Eine Literaturanalyse [Driving in Old Age - a Literature Analysis]'. *Zeitschrift Für Gerontopsychologie & -Psychiatrie, 13*(3–4), 131–70. doi: 10.1024//1011-6877.13.34.131

Keall, M. D., & Frith, W. J. (2004). Older driver crash rates in relation to type and quantity of travel. *Traffic Injury Prevention*, *5*(1), 26-36. doi: 10.1080/15389580490269146

Kilpeläinen, M., & Summala, H. (2007). Effects of weather and weather forecasts on driver behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*, *10*(4), 288-299. doi: 10.1016/j.trf.2006.11.002

Klein, G., Woods, D., Bradshaw, J., Hoffman, R., & Feltovich, P. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems, 19*(06), 91–95. doi:10.1109/MIS.2004.74.

Ko, B. (2018). A Brief Review of Facial Emotion Recognition Based on Visual Information. *Sensors, 18(2), 401*. doi:10.3390/s18020401

Körner, A., Geyer, M., Roth, M., Drapeau, M., Schmutzer, G., Albani, C., … (2008). Personlichkeitsdiagnostik mit dem NEO-Funf-Faktoren-Inventar: Die 30-Item-Kurzversion (NEO-FFI-30). *Psychotherapie, Psychosomatik, Medizinische Psychologie, 58*(6), 238–245. doi:10.1055/s-2007-986199.

Krause, M., Henel, J., & Bengler, K. (2016). Performance and Behavior of a Codriver When Using a Mobile Device. In *62. Kongress der Gesellschaft für Arbeitswissenschaft, Germany, Aachen*.

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

Lee, C., Seppelt, B., Abraham, H., Reimer, B., Fitzgerald, C., Mehler, B., & Coughlin, J. F. (2018). Consumer Comfort with In-Vehicle Automation: Technology of Today Drives Acceptance of a Self-Driving Future. *White Paper 2018-2–MIT AgeLab*.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243-1270. doi: 10.1080/00140139208967392

Lehtonen, E., Lappi, O., Koirikivi, I., & Summala, H. (2014). Effect of driving experience on anticipatory look-ahead fixations in real curve driving. *Accident Analysis, & Prevention, 70*, 195-208.

Liljamo, T., Liimatainen, H., & Pöllänen, M. (2018). Attitudes and concerns on automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *59*, 24-44. doi: 10.1016/j.trf.2018.08.010

McDonald, A., Carney, C., & McGehee, D. V. (2018). *Vehicle Owners' Experiences with and Reactions to Advanced Driver Assistance Systems*. AAA Foundation for Traffic Safety. Retrieved from https://aaafoundation.org/wp-content/uploads/2018/09/VehicleOwnersExperiencesWithADAS_TechnicalReport.pdf

McKenna, F. P., Stanier, R. A., & Lewis, C. (1991). Factors underlying illusory self-assessment of driving skill in males and females. *Accident Analysis & Prevention*, *23*(1), 45-52. doi: 10.1016/0001-4575(91)90034-3

Meng, A., & Siren, A. (2012). Cognitive problems, self-rated changes in driving skills, driving-related discomfort and self-regulation of driving in old drivers. *Accident Analysis & Prevention*, *49*, 322-329. doi: 10.1016/j.aap.2012.01.023

Moták, L., Gabaude, C., Bougeant, J. C., & Huet, N. (2014). Comparison of driving avoidance and self-regulatory patterns in younger and older drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *26*, 18-27. doi: 10.1016/j.trf.2014.06.007

Mueller, A.S., & Trick, L.M. (2012). Driving in fog: The effects of driving experience and visibility on speed compensation and hazard avoidance. *Accident Analysis, & Prevention, 48*, 472-479.

Mühlbacher, D., Tomzig, M., Reinmüller, K., & Rittger, L. (2020). Methodological Considerations Concerning Motion Sickness Investigations during Automated Driving. *Information*, *11*(5), 265. doi: 10.3390/info11050265

Myers, A. M., Trang, A., & Crizzle, A. M. (2011). Naturalistic study of winter driving practices by older men and women: Examination of weather, road conditions, trip purposes, and comfort. *Canadian Journal on Aging/La Revue canadienne du vieillissement*, *30*(4), 577-589. doi: 10.1017/S0714980811000481

Neubauer, C., Gremillion, G., Perelman, B. S., La Fleur, C., Metcalfe, J. S., & Schaefer, K. E. (2020). Analysis of Facial Expressions Explain Affective State and Trust-Based Decisions During Interaction with Autonomy. In: T. Ahram, W. Karwowski, A. Vergnano, F. Leali, & R. Taiar (Eds.), *Intelligent Human Systems Integration 2020* (pp. 999–1006). Cham: Springer International Publishing. doi:10.1007/978-3-030-39512-4_152.

Noh, S., & An, K. (2017). Decision-making framework for automated driving in highway environments. *IEEE Transactions on Intelligent Transportation Systems*, *19*(1), 58-71. doi: 10.1109/TITS.2017.2691346

Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 252-263. doi: 10.1016/j.trf.2014.04.009

Pink, O., Becker, J., & Kammel, S. (2015). Automated driving on public roads: Experiences in real traffic. *it-Information Technology*, *57*(4), 223-230. doi: 10.1515/itit-2015-0010

Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., & Neubauer, C. (2013). Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts. *Journal of experimental psychology: applied*, *19*(4), 287. doi: 10.1037/a0034386

Schmidt, L., & Piringer, A. (2012). VIP Verkehrsspezifischer Itempool. Manual (Version 22, 2012). Wien: Kuratorium für Verkehrssicherheit.

Shabanpour, R., Golshani, N., Shamshiripour, A., & Mohammadian, A. K. (2018). Eliciting preferences for adoption of fully automated vehicles using best-worst analysis. *Transportation Research Part C: Emerging Technologies*, *93*, 463-478. doi: 10.1016/j.trc.2018.06.014

Sigari, M.-H., Pourshahabi, M.-R., Soryani, M., & Fathy, M. (2014). A Review on Driver Face Monitoring Systems for Fatigue and Distraction Detection. *International Journal of Advanced Science and Technology*, *64*, 73–100. doi:10.14257/ijast.2014.64.07

Smith, S., Koopmann, J., Rakoff, H., Peirce, S., Noel, G., Eilbert, A., & Yanagisawa, M. (2018). *Benefits Estimation Model for Automated Vehicle Operations: Phase 2 Final Report*. U.S. Department of Transportation. Retrieved from https://rosap.ntl.bts.gov/view/dot/34458/dot_34458_DS1.pdf?.

Suh, J., Chae, H., & Yi, K. (2018). Stochastic model-predictive control for lane change decision of automated driving vehicles. *IEEE Transactions on Vehicular Technology*, *67*(6), 4771-4782. doi: 10.1109/TVT.2018.2804891

Tan, C.F., Delbressine, F.L., Chen, W., & Rauterberg, M. (2008). Subjective and objective measurements for comfortable truck driver's seat. In: *Proceedings of the 9th International*

Symposium on Advanced Vehicle Control (AVEC 2008), p. 851–856. Retrieved from https://www.researchgate.net/publication/254758570.

Taylor, J. E., & Paki, D. (2008). Wanna drive? Driving anxiety and fear in a New Zealand Community Sample. *New Zealand Journal of Psychology*, 37(2), 31 – 37.

Techer, F., Ojeda, L., Barat, D., Marteau, J. Y., Rampillon, F., Feron, S., & Dogan, E. (2019). Anger and highly automated driving in urban areas: The role of time pressure. *Transportation Research Part F: Traffic Psychology and Behaviour*, *64*, 353-360. doi: 10.1016/j.trf.2019.05.016

Themann, P., Bock, J., & Eckstein, L. (2014). Optimisation of energy efficiency based on average driving behaviour and driver's preferences for automated driving. *IET intelligent transport systems*, *9*(1), 50-58. doi: 10.1049/iet-its.2013.0121

Underwood, G., Chapman, P., Bowden, K., & Crundall, D. (2002). Visual search while driving: Skill and awareness during inspection of the scene. *Transport Research Part F: Traffic Psychology and Behaviour, 5,* 87-97.

Wadud, Z., & Huda, F. Y. (2019, March). Fully automated vehicles: the use of travel time and its association with intention to use. In *Proceedings of the Institution of Civil Engineers-Transport* (pp. 1-15). Thomas Telford Ltd. doi: 10.1680/jtran.18.00134

Wang, H., & He, W. (2011, April). A reservation-based smart parking system. In *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 690-695). IEEE. doi: 10.1109/INFCOMW.2011.5928901

Whelan, M., Senserrick, T., Groeger, J., Triggs, T., & Hosking, S. (2004). Learner driver experience project, *Report No. 221.* Monash University Accident Research Center, Clayton, Victoria, Australia.

5 Key Performance Indicators

Key Performance Indicators are indicators for assessing partial or full system effectiveness, and as such will be fully defined and assessed in WP3 and WP4. For the purpose of defining data collection parameters, some initial thoughts about KPI(S) will be included in this chapter.

5.1 Road Traffic Safety

Traffic Conflict Techniques (TCT) have been used in traffic safety since its inception in the 1960s as a tool for safety evaluation where insufficient accident data is available (Kraay J. H., 1983; Grayson, Hydén, Kraay, Muhlrad, & Oppe, 1984; Kraay, van der Horst, & Oppe, 2013; Hydén, 2016; Hydén & Linderholm, 1984). The Swedish Traffic Conflict Technique (STCT) introduced *Time to Accident* (TA) (Hydén & Linderholm, 1984), later, the Dutch Objective Conflict Technique for Operation and Research (DOCTORS) (Kraay, van der Horst, & Oppe, 2013) combined *Time to Collision* (TTC) (Hayward, 1972) with the *Post Encroachment Time* (PET) (Amundsen & Hydén, 1979) introducing the concept of *Probability of Conflict*. *Conflict Speed* (CS) at the initiation of evasive manoeuvre is another indicator used in TCT (Hydén, 2016).

The Swedish Transport Administration, Trafikverket, considers *Compliance with Speed limits*, *Sobriety,* and *Use of seat belt* as some of the key road safety performance indicators. An emphasis on *Safe passenger cars* is made through five-star Euro NCAP rating (Amin, et al., 2015; 2019). Trafikverket also recommends infrastructure and vehicle technology development by following ISO 39001:2012. This standard has the performance indicators categorized under three broad topics: Risk exposure factors (such as distance travelled and traffic volume), Final safety outcome factors (such as number of deaths, serious injuries) and Intermediate safety outcome factors (such as road design and appropriate speed, appropriate route selection based on vehicle type, use of personal safety equipment like seatbelts, fitness of driver, appropriate speed selection in the given context, vehicle safety measures like occupant protection, collision avoidance, and post-crash response including emergency response and preparedness) (International Organization for Standardization, 2012).
These KPIs are aligned with the recommendations from the European Transportation Safety Council (2018) and the European Commission's working documents (2019) in the continued pursuit of Vision Zero. MacDonald et al. (2004), in an international study with the U.S. Department of Transportation noted that most transport agencies use "number of fatalities" as an important indicator and often use derivatives of this metric such as total societal cost.

5.2 Intelligent Transportation Services

The EU research project AdaptIVe (2017) classified the relevant indicators into three main categories, namely Vehicle, Driver and Environment. Manoeuvre duration, automation level, manoeuvring velocity, control force, time headway, trigger and coordination between vehicles were the vehicle related parameters which were collected from the vehicle. Kaparias et al. (2011) built a KPI framework for Intelligent Transport Systems based on four strategic pillars namely: traffic efficiency, traffic safety, social inclusion and land use, and pollution reduction (refer CONDUITS Deliverable 3.5 for a comprehensive list of KPIs).

AECOM (Payne, Hobbs, & Redfern, 2015) in a study commissioned by the EU Directorate-General for Mobility and Transport identified a total of 228 KPIs related to ITS, of which 61% were related to benefit or impacts of the ITS and 39% to deployment and implementation. These included KPIs such as network efficiency and congestion, environmental impacts, improved road safety, enhanced modal integration, security among others. The authors then scored these KPIs based on Clarity, Meaningfulness, Complexity and Transferability. These KPIs were also used by Marsili, et al. (2016) in their recommendations to the European Commission through the research project EU EIP.

5.3 Human Machine Interface

Usability measures are widely used as the key indicator to evaluate the impact of HMI and underlying user interactions. Usability metrics can be calculated through formal (ISO 9241-11:2018) and informal methods (heuristic evaluation method, Nielsen & Molich, 1990). In a heuristic evaluation method, HMI experts evaluate the interfaces based on the usability guidelines and principles while performing user tasks. Naujoks et al. (2019) developed a tool (detail checklist) to verify if the automated vehicle HMI complies with international standards and best practices, that could be used for heuristic assessment. The checklist items are shown below in **Fout! Verwijzingsbron niet gevonden.**:

- The system mode should be displayed continuously
- System state changes should be effectively communicated
- Visual interfaces used to communicate system states should be mounted to a suitable position and distance. High-priority information should be presented close to the driver's expected line of sight
- HMI elements should be grouped together according to their function to support the perception of mode indicators
- Time-critical interactions with the system should not afford continuous attention
- The visual interface should have a sufficient contrast in luminance and/or color between foreground and background
- Texts (e.g., font types and size of characters) and symbols should be easily readable from the permitted seating position
- Commonly accepted or standardized symbols should be used to communicate the automation mode. Use of non-standard symbols should be supplemented by additional text explanations or vocal phrase/s
- The semantic of a message should be in accordance with its urgency
- Messages should be conveyed using the language of the users (e.g., national language, avoidance of technical language, use of common syntax)
- Text messages should be as short as possible
- Not more than five colors should be consistently used to code system states (excluding white and black)
- The colors used to communicate system states should be in accordance with common conventions and stereotypes
- Design for color-blindness by redundant coding and avoidance of red/green and blue/yellow combinations
- Auditory output should raise the attention of the driver without startling her/him or causing pain
- Auditory and vibrotactile output should be adapted to the urgency of the message
- High-priority messages should be multimodal
- Warning messages should orient the user towards the source of danger
- In case of sensor failures, their consequences and required operator steps should be displayed

Figure 30. HMI checklist for heuristic assessment (Naujoks, Wiedemann, Schömig, Hergeth, & Keinath, 2019)

Forster et al. (2019) validated the above HMI checklist effectiveness through a driving simulator study which measured the impact of usability and acceptance of two HMI variants (designed for transitions between L3 to manual driving). Usability measures decreased significantly when the HMI requirements of the checklist were not met.

5.4 Driver and Occupant State

Considering the two driver states highlighted in MEDIATOR, i.e. distraction and fatigue, and given a universally agreed reference measure of driver state, the most common KPI for classification is found in signal detection theory – true positive, true negative, false positive and false negative and ratios between them, their ROC-curves (receiver operating characteristics) and AUC (area under curve) (Wikipedia contributers, 2019a; 2019b). However, the premise of a universally agreed reference, or ground truth, may well be the biggest challenge in this endeavour.

For distraction, visual inattention has been explored (by e.g., Victor, Ahlström, Kircher) with various criteria for classifying the driver as distracted following a driver's glance away from the road centre or area relevant for driving, counting down from a conceptual 'bank' of attentional resources. Yet the concept of attention/distraction is very elusive and lacks a definition that is generally agreed upon (e.g., MiRA paper by Kircher).

For fatigue, including sleep-related fatigue, the most straightforward method would be employing a self-rating scale, e.g., the KSS scale, which has the benefits of being easily deployed and immediately available for analysis. KSS is also the preferred sleepiness/fatigue ground truth in the upcoming amendment of Regulation (EU) 2019/2144 of the European Parliament and of the Council with technical requirements and test procedures for type-approval of motor vehicles with regard to driver drowsiness warnings. The problems associated with self-assessment includes inter- and intra-individual variability, ceiling effects, poor time resolution and the fact that the assessment itself may have local influence on the individuals' sleepiness level. In an attempt to obtain more 'objective' sleepiness data, various expert scoring methods could be employed (video, EEG, EOG). The drawbacks of these methods are that the rating/annotation is slow and consumes a lot of manual efforts, while still suffering from inter- and intra-individual differences and rater subjectivity, usually addressed by merging results from multiple raters. The use of EEG in dynamic settings and with practical applications is also a challenge due to motion artefacts

For driver discomfort, changes in Action Units (AU) related to the close approach situation showed that eyes were kept open (AU43) and eye blinks were reduced, indicating attentive visual monitoring of the situation. Observed raises of upper lids (AU5) and inner brows (AU1) are considered as essential component in all prototypes and major variants of surprise. In addition, the lips were pressed (AU24) and stretched (AU20) during the discomfort interval. Simultaneously, lower lip drop diminished (AU16) and upper lip raiser increased (AU10), resulting in an upward oriented lip pressing motion. This combination of lip movements could be interpreted as sign for tension. Thus, the combination of all AU changes points towards a reaction of surprise, tension and visual attention during the close approach situation.

5.5 References

AdaptIVe. (2017). *Deliverable D1.0: Final project results of EU FP7 project AdaptIVe (www.adaptive-ip.eu).*

Amin, K., Bengtsson, K., Berg, H.-Y., Skyving, M., Yamazaki, R., Forsman, Å., . . . Rizzi, M. (2019). *Analys av trafiksäkerhets utvecklingen 2018. Målstyrning av trafiksäkerhetsarbetet mot etappmålen 2020.* Trafikverket.

Amin, K., Berg, H.-Y., Bengtsson, K., Larsson, P., Forsman, Å., Vadeby, A., . . . Strandroth, J. (2015). *Analysis of road safety trends 2014: management by objectives for road safety work towards the 2020 interim targets.* Trafikverket.

Amundsen, F., & Hydén, C. (1979). Proceedings of the First Workshop on Traffic Conflicts, Oslo, September 26–27, 1977. *Accident Analysis & Prevention 11(4)* , 317-318.

European Commission . (2019). *Staff Working Document: EU Road Safety Policy Framework 2021-2030 – Next steps towards "Vision Zero".* Brussels: European Commission.

European Transport Safety Council. (2018). *Breifing: 5th EU Road Safety Action Programme 2020-2030.* Brussels: European Transport Safety Council (ETSC).

Forster, Y., Hergeth, S., Naujoks, F., Krems, J., & Keinath, A. (2019). Empirical Validation of a Checklist for Heuristic Evaluation of Automated Vehicle HMIs. *International Conference on Applied Human Factors and Ergonomics* (pp. 3-14). Springer.

Grayson, G., Hydén, C., Kraay, J., Muhlrad, N., & Oppe, S. (1984). *The Malmö study: a calibration of traffic conflict techniques. A study organised by ICTCT (The International Committee on Traffic Conflict Techniques.* The International Committee on Traffic Conflict Techniques.

Hayward, J. C. (1972). Near miss determination through use of a scale of danger. *Highway Research. Record, 384*, 22–34.

Hydén, C. (2016). Traffic conflict techniques: some data to supplement accident analysis. *Transport Planning and Traffic Safety: Making Cities, Roads, and Vehicles Safer*, 89-107.

Hydén, C., & Linderholm, L. (1984). The Swedish traffic-conflicts technique. *International Calibration Study of Traffic Conflict Techniques*, 133-139.

International Organization for Standardization. (2012). *ISO 39001: 2012: Road Traffic Safety (RTS) Management Systems: Requirements with Guidance for Use.* ISO.

Kaparias, I., Bell, M. G., & Tomassini, M. (2011). *Deliverable 3.5: Key Performance Indicators for traffic management and Intelligent Transport Systems.* CONDUITS consortium,

Kraay, J. H. (1983). *Review of traffic conflicts technique studies, R-83-9.* Institute for Road Safety Research SWOV.

Kraay, J. H., van der Horst, A. R., & Oppe, S. (2013). *Manual conflict observation technique DOCTOR (Dutch Objective Conflict Technique for Operation and Research).* Voorburg (NL): Foundation Road safety for all.

MacDonald, D., Yew, C., Arnold, R., Baxter, J., Halvorson, R., Kassoff, H., . . . White, W. (2004). *Transportation Performance Measures in Australia, Canada, Japan, and New Zealand.* U.S. Department of Transportation.

Marsili, F., Lüpges, C., Geißler, T., Schuster, G., Kulmala, R., Studer, L., . . . Strom, M. (2016). *EU EIP Activity 5: ITS Deployment and Benefit KPIs definitions - Technical Reference.*

Naujoks, F., Wiedemann, K., Schömig, N., Hergeth, S., & Keinath, A. (2019). Towards guidelines and verification methods for automated vehicle HMIs. *Transportation Research Part F: Traffic Psychology and Behavior, Volume 60*, 121-136.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 249-256). Association for Computing Machinery. doi:https://doi.org/10.1145/97243.97281

Payne, S., Hobbs, D., & Redfern, R. (2015). *Key performance indicators for intelligent transport systems.* AECOM.

Wikipedia contributers. (2019a, July 23). *Detection theory.* Retrieved from Wikipedia, The Free Encyclopedia: https://en.wikipedia.org/w/index.php?title=Detection_theory&oldid=907590846

Wikipedia contributors. (2019b, September 23). *Receiver operating characteristic*. Retrieved from Wikipedia, The Free Encyclopedia: https://en.wikipedia.org/w/index.php?title=Receiver_operating_characteristic&oldid=917442289

6 Functional requirements

6.1 Main conclusions

The results reported in this deliverable laid the foundations for the functional requirements for the driver state module. With respect to driver distraction and driver fatigue various algorithms were developed that can detect distraction and fatigue.

The preliminary prototype of the distraction detection algorithm showed promising results. The accuracy for eyes-off-road classification measured with the F-score was 92.75%. The accuracy of classifying a 40-second episode as distracted based on the simplified AttenD algorithm was 95.85%. And finally, the prediction for distraction severity had a Root Mean Square Error (RMSE) of 0.06 and MSD was 0.00.

The preliminary prototype of the fatigue detection algorithm based on video data and physiological data, showed a Mean Absolute Error (MAE) of about 0.5 points on the Karolinska Sleepiness Scale (KSS), or a binary classification accuracy of about 90%.

The conclusions with respect to comfort detection indicate that while it is possible to detect drivers' discomfort based on facial expression features, i.e., action units (AU) at the lab under controlled conditions, it might be too challenging to develop a real-time algorithm for real-life conditions within the MEDIATOR project. This is mainly because AU substantially differ between people and is affected by many factors such as personality and shift in time. Although a potential algorithm will be further explored in MEDIATOR, it is argued that identifying situations that make drivers uncomfortable offline is a more feasible approach for the Mediator system developed in this project.

Besides the development of driver state detection algorithms, the evolvement of fatigue and distraction over time and its effect on driving performance was evaluated under L2 driving conditions. The studies were performed both on the road as in a driving simulator. The main conclusion of these studies can be summarized as follows.

With respect to fatigue there was consistent evidence that fatigue was induced faster under L2 driving conditions than under manual driving conditions. However, the results of the on-road study revealed that the differences between manual and L2 driving conditions were mainly noticeable during night-time driving and less so during daytime driving. Instead, the driving simulator study, which was executed in morning and noon hours, did show significant effects of L2 on fatigue induction during day-time driving. It is currently not yet clear where these discrepancies between the on-road study and the simulator study stem from. One possible explanation is that during night time there is much less traffic on actual roads, which might have contributed to underload during L2 driving.

The results from the driving simulator study also revealed that drivers who played a game of Simon under L2 driving conditions reported lower KSS scores than drivers who did not play it. This finding is encouraging as it demonstrates the importance of engagement with a secondary task under driving conditions that are likely to induce fatigue. This also emphasizes the complex trade-off between being distracted versus being fatigued under L2 driving conditions. In terms of driving performance, it was shown that under L2 driving conditions, regardless of whether drivers engaged with a secondary task, drivers had fewer number of glances on road hazards, which may indicate they had poorer situation awareness (SA) as compared to manual driving.

Next, with respect to driver distraction under L2 driving conditions it was found that drivers who were asked to engage with a secondary task during L2 driving conditions had a much smaller probability of identifying a hazard (~0.54) and deactivated the automation due to hazardous situations less often (probability of deactivating the automation =~0.116) than drivers who were not engaged with a secondary task (0.91, 0.22 respectively) under L2 conditions. These findings indicate that when distracted under L2 drivers' SA was poorer than drivers who were not distracted. Although drivers in this study were instructed to engage with a secondary task at specific time windows, the results have shown that drivers do regulate their behaviour (to some extent) and allocate more attention to the road in urban environments than in highway environments under L2 driving conditions.

Ultimately it can be concluded that, to make a safe and comfortable trade-off between driver and automation fitness, driver states should be monitored and predicted while driving with and without automation functions. The results described in this deliverable also show that driving with partial automation can result in degraded driver states, which emphasizes the importance of the Mediator system functions that focus on maintaining driver fitness.

6.2 Functional requirements for the driver module

The functional requirements define the function of the system and its components. Table 10 (next page) summarizes the functional requirements of the Driver State module with regard to fatigue, distraction and discomfort. The functional requirements for the driver module are based on the research described in this deliverable as well as on the knowledge gained during the analysis and experimentation phase of the MEDIATOR project. These functional requirements provide input to guide the further design and development of the Mediator driver state module.

Table 10. Functional requirements for the Mediator driver module

Functional Requirements

The system shall estimate worst, likely and best-case *time to driver (un)fitness* based on driver fatigue and distraction estimates for the current driving context
• The system shall *estimate* driver *fatigue* and *distraction*
  ○ The system shall detect heart rate, respiratory and facial features
  ○ The system shall estimate KSS score
  ○ The system shall detect eyes on/off road
  ○ The system shall estimate distraction severity based on eyes off/on road
  ○ The system shall detect non-related driving task
• The system shall *predict* driver *fatigue* and *distraction*
  ○ The system shall predict fatigue progression based on current KSS score
  ○ The system shall predict loss of situation awareness based on distraction severity
  ○ The system shall predict time to driver fitness based on non-related driving task involvement
• The system shall estimate when the driver is *unfit to drive*
  ○ The driver is deemed unfit to drive if it can no longer execute the manual driving task in a sufficiently safe manner due to degraded cognitive abilities (fatigue) or loss of situation awareness (distraction).
• The system shall estimate *time to driver fitness* as the longest estimate based on distraction or fatigue
• The system shall estimate *time to driver unfitness* as the shortest estimate based on distraction or fatigue
• The system shall estimate *worst*, *likely* and *best* cases using reliability of the underlying estimates
• The system shall request *context relevant information* from the context module
  ○ Possible information can be time of day, type of road and situation complexity
• The system shall *personalise* these estimations to improve accuracy
  ○ Detect *driver ID*
    ▪ Possibly request sleep/driving history and driver age via HMI
  ○ Estimate driver fitness with *individualized algorithms* tuned to specific drivers
  ○ Estimate comfort based on *historical data* on user acceptance or rejection of the takeover suggestions

The system shall determine the *driver state class* as fit, distracted or fatigued

The system shall estimate *worst*, *likely* and *best*-case *time to driver discomfort*
• The system shall compare upcoming *driving situations* with the identified uncomfortable driving situations
  ○ The system shall request relevant information from the *context module*
  ○ The system shall request information from the driver on uncomfortable driving situations
• The system shall estimate the *time* until the uncomfortable situation will occur
• The system shall estimate *worst*, *likely* and *best* case using the probability that comfort will be increased
• The system shall *personalise* the probability of a situation being uncomfortable to improve accuracy