Technische Universität München
Fakultät für Maschinenwesen
Lehrstuhl für Ergonomie

Non-Driving-Related Tasks in Conditional Driving Automation

Occurence and Effects of Passive Task-Related Fatigue on Take-Over Performance in Conditional Driving Automation

Oliver Jarosch, M.Sc.

Vollständiger Abdruck der von der Fakultät für Maschinenwesen der Technischen Universität München zur Erlangung des akademischen Grades eines Doktor-Ingenieurs (Dr.-Ing.) genehmigten Dissertation.

Vorsitzender: Prof. Dr. Markus Zimmermann

Prüfer der Dissertation: 1. Prof. Dr. Klaus Bengler
2. Prof. Dr. Mark Vollrath

Die Dissertation wurde am 19.11.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Maschinenwesen am 01.07.2020 angenommen.

—— Meinem Vater ——

Danksagung:

Liebe Kolleginnen und Kollegen,
liebe Familie,

nach letztendlich etwas über drei äußerst interessanten, spannenden und abwechslungsreichen Jahren konnte ich die vorliegende Arbeit vollenden. Während meiner Zeit bei der *BMW Forschung, Neue Technologie und Innovationen* konnte ich mein Wissen im Bereich Fahrerassistenz, Mensch-Maschine-Interaktion und Fahrsimulation erweitern und gleichzeitig neue Freundschaften knüpfen.

Die Aufgaben die hierzu zu bewerkstelligen waren, haben mich teilweise stärker beansprucht als ich es mir vor der Dissertation ausmalen konnte. Nur mit der Unterstützung einer Vielzahl von Personen konnte mein Projekt verwirklicht werden.

Mein erster Dank geht an Herrn Prof. Klaus Bengler, der meine Dissertation universitätsseitig betreut hat. Besonders die regelmäßige Rücksprachen mit Ihnen und ihr Input zu den jeweils anstehenden Studien habe ich sehr geschätzt. Ein weiterer Dank geht an Herr Prof. Mark Vollrath, der die Zweitbegutachtung der Dissertation übernommen hat. Vielen Dank auch für die Organisation ihrer Verkehrspsychologen-Doktoranden Workshops. Auf diesen Veranstaltungen konnte ich wichtiges Feedback erhalten und gleichzeitig haben sie mir viel Freude bereitet.

Vielen Dank weiterhin an Sebastian Hergeth, Ronee Chadowitz, Svenja Paradies und Sonja Rümelin. Ohne euch hätte ich dieses spannende Promotionsthema wohl nie bearbeiten können. Vor allem dir liebe Svenja, möchte ich einen besonderen Dank aussprechen. Ich habe die Zeit mit dir, als meine Betreuerin seitens BMW, sehr genossen!

Ein weiterer Dank geht an meine Praktikanten und Studenten, die mich während meiner Promotion tatkräftig unterstützt hatten: Mark-Kevin Doeker, Matthias Kuhnt, Corinna Köhler und Daniel Feiner.

Ein weiteres Dankeschön geht an Hanna Bellem. Danke für deine Doktoranden-Insider-Tipps und dem Input zu meinen Versuchen.

Auch dem Fahrsimulator-Team um Stefan Barowski möchte ich Danke sagen. Auch wenn es nicht immer einfach war, konntet ihr stets alle Probleme beseitigen, wodurch eine erfolgreiche Durchführung der Versuche gewährleistet werden konnte.

Vielen Dank weiterhin an Florian Fischer, der es mir ermöglicht hat, den Wizard-of-Oz Versuch durchführen zu können. Auch ohne den Fahrern Martin Liebich, Stefan Scheller und Steven Hofmann wäre dieser wichtige Versuch nicht umsetzbar gewesen. Danke!

Zu guter Letzt kommen die Leute, denen ich wohl am meisten zu verdanken habe. Meine Mutter Susanne, mein Zwillingsbruder Sebastian, meine Großeltern Hedwig, Emmy und Gerhard, mein Onkel Roland und meine Schwiegereltern in spe haben mich vom Anfang bis zum Ende unterstützt.

Liebsten Dank auch an dich liebe Milena, dass du mir immer positiv zugesprochen und mir Kraft für den nächsten Tag gegeben hast. Danke für deine Unterstützung!

Zusammenfassung:

In den kommenden Jahren, werden Fahrerassistenzsysteme und automatisierte Fahrfunktionen immer mehr Aufgaben übernehmen können und den Menschen zumindest teilweise in seiner heutigen Rolle als Fahrzeugführer ablösen. Für den menschlichen Fahrer bedeutet dies, dass er sich sobald das System aktiv ist, vollständig von der Fahraufgabe abwenden kann und sich während der automatisierten Fahrt mit anderen Dingen und Aktivitäten beschäftigen kann.

Anfangs wird die automatisierte Fahrfunktion nicht auf allen Strecken und in allen Situationen verfügbar sein und es wird Situationen geben, in denen das System die Fahraufgabe an den Fahrer zurückgeben wird. Diese Automationsstufe, in der der menschliche Fahrer als Rückfallebene agiert wird als *hochautomatisiertes Fahren* bezeichnet. In dieser Automationsstufe ist die Sicherheit der Fahrzeuginsassen somit auch abhängig von der Reaktion des Fahrers, der vom System zur Übernahme der Fahrzeugkontrolle aufgefordert werden kann. Da ein sicheres Eingreifen, unter allen Umständen, zwingend erforderlich ist, ergeben sich unterschiedliche Forschungsfragen im Hinblick auf die Kontrollübernahme. Im Rahmen dieser Dissertation wurde untersucht, (i) welchen Einfluss unterschiedliche Tätigkeiten, die der Fahrer während einer solchen automatisierten Fahrt bearbeitet, auf dessen Müdigkeitszustand haben, (ii) wie Müdigkeit beim hochautomatisierten Fahren detektiert werden kann, und (iii) inwiefern die durch die Aufgaben erzeugte Müdigkeit das Übernahmeverhalten beeinflussen können. (iv) Weiterhin wurde die Übertragbarkeit vom Müdigkeitsverlauf im Simulator auf den Realverkehr untersucht. (iiv) Abschließend wurde untersucht, durch welche Warnkonzepte der Fahrer bestmöglich in die Fahraufgabe zurückgeholt werden kann. Um diese Fragestellungen beantworten zu können, wurden insgesamt drei Fahrsimulator-Studien sowie eine Studie im Realverkehr, mit Hilfe eines Wizard-of-Oz Fahrzeugs, durchgeführt.

Als Ergebnis lässt sich festhalten, dass aufgabenbedingte Müdigkeit beim hochautomatisierten Fahren bereits nach weniger als 20 min aufgetreten ist. Das Auftreten von Müdigkeit ist hierbei stark abhängig von der jeweiligen Tätigkeit, die der menschliche Fahrer während der Fahrt bearbeitet. Monotone Überwachungsaufgaben scheinen besonders schnell zu Müdigkeit zu führen, wohingegen eine freie Beschäftigung mit selbstgewählten Tätigkeiten auftretende Müdigkeit nahezu komplett verhindern konnte.

Weiterhin konnte gezeigt werden, dass der Einfluss von aufgabenbedingter Müdigkeit auf das Übernahmeverhalten eher gering ist. Dahingegen scheint der Einfluss der automatisierten Fahrtdauer deutlich relevanter zu sein. Das Übernahmeverhalten war vor

allem nach längeren Versuchsfahrten im Vergleich zu kürzeren Versuchsfahrten mit dem automatisierten System, deutlich eingeschränkt. Nach längeren Automationsabschnitten, konnten einige Probanden die Kontrolle über das Fahrzeug nicht rechtzeitig bzw. nur unzureichend zurückerlangen, wodurch sie die Kontrolle über das Fahrzeug verloren haben oder einen Unfall nicht verhindern konnten.

Die Erfassung der Müdigkeit erfolgte sowohl subjektiv als auch objektiv. Als valider Müdigkeitsindikator konnte PERCLOS, welcher das Lidschlussverhalten über die Zeit wiederspiegelt, bestätigt werden.

Der Verlauf der Müdigkeit im Realverkehr war ähnlich zu dem, der während der Versuche im Fahrsimulator gemessen wurde. Bezüglich der Warnkonzepte, welche den Fahrer in den Übernahmesituationen in die Fahraufgabe zurückholen sollten, konnten mit einem blicklenkenden Konzept die besten Ergebnisse erzielt werden.

Abstract:

In the upcoming years, driver assistance systems and automated driving functions will be able to take over more and more functions and replace the human driver at least partially in his current role as a vehicle operator. Consequently, as soon as the system is active, the human driver will be able to completely turn away from the driving task and will be able to concentrate on other things and activities during the automated drive.

Initially, the automated driving function will not be available on all roads and in all situations and additionally, there will be situations where the system will return the driving task to the driver. This level of automation, in which the human driver acts as a fallback user, is named *conditional driving automation*. Thus, in this level of automation, the safety of the vehicle occupants is also dependent on the driver's reaction after he has been requested by the system to take over vehicle control. Since a safe intervention is absolutely essential under all circumstances, various research questions arise with regard to the taking-over of vehicle control. In the context of this dissertation it was examined (i) how different activities, the driver works on during such an automated drive, affect the driver's fatigue state, (ii) how emerging fatigue can be detected during conditional driving automation, and (iii) to what extent the task-related fatigue can affect the take-over behavior. (iv) Furthermore, the transferability of emerging fatigue from the driving simulator to real traffic on-road environment was investigated. (iiv) Finally, it was examined which warning concepts can be used to get the driver back into the driving task in the best possible way. In order to answer these questions, a total of three driving simulator studies as well as one experiment in real traffic environment was carried out with the help of a Wizard-of-Oz vehicle.

As a result, it can be stated that task-related fatigue in conditional driving automation was already measurable after less than 20 min. The occurrence of fatigue is strongly dependent on the type of task that the human driver performs while driving. Monotonous monitoring tasks seem to lead to fatigue rather quickly, whereas a free engagement with self-selected activities could almost completely prevent emerging fatigue.

Furthermore, it could be shown that the influence of task-related fatigue on the take-over behavior is rather small. On the other hand, the influence of automated driving time seems to be much more relevant. The take-over behavior was significantly reduced, especially after longer experimental rides compared to shorter experimental rides when

using the automated driving system. After prolonged periods of automation, some individuals were unable to regain control of the vehicle in time, which caused them to skid on the road. Others could not prevent an accident.

Fatigue was assessed both subjectively and objectively. PERCLOS, which reflects the eyelid closure behavior over time, was confirmed as a valid fatigue indicator.

The course of measured fatigue in real traffic environment on-road was similar to the course of fatigue measured in the driving simulator. With regard to the warning concepts, which were supposed to quickly get the driver back in the driving task, the best results were achieved with a visual warning concept.

Contents:

List of figures:

List of tables:

Abbreviations:

ABS .... Anti-lock Braking System

ACC .... *Adaptive Cruise Control*

$Acc_{\_lat\_max}$ .... *Maximum lateral acceleration (in m / s²)*

$Acc_{\_long\_max}$ .... *Maximum longitudinal acceleration (in m / s²)*

ACEA .... *Association des Constructeures Européens d'Automobiles*

ADAS .... Advanced Driver Assistance Systems

ADS .... *Automated driving systems*

BASt .... *Bundesanstalt für Straßenwesen*

CDA .... *Conditional driving automation*

CID .... *Central information display*

DDT .... *Dynamic driving task*

ECG - .... *Electrocardiogram*

EEG.. *Electroencephalogram*

ESC .... Electronical Stability Control

ESS .... *Epworth Sleepiness Scale*

FIZ.. *Forschungs- und Innovationszentrum*

HMI... *Human-machine-interaction*

HR .... *Heart rate*

HRV.. *Heart rate variability*

KSS .... *Karolinska Sleepiness Scale*

LDW .... *Lane Departure Warning*

LKA... *Lane Keeping Assistant*

MWT .... *Maintenance of Wakefulness Test*

NDRT .... *Non-Driving-Related-Task*

PERCLOS .... *Percentage of eye-lid closure over the pupil over time*

PVT .... *Psychomotor vigilance task*

RtI .... *Request-to-intervene*

SAE .... *Society of automotive engineers*

SDLP ....

Measurement of Standard Deviation of Lateral Position, *standard deviation of lateral position*

SSS .... *Stanford Sleepiness Scale*

SuRT .... *Surrogate reference task*

SWM .... *Measurement of Steering Wheel Movement*

$t_{brake\_maneuver}$ .... *First braking maneuver (in s)*

$t_{brake\_reaction}$ .... *First braking reaction (in s)*

TCS .... *Traction Control System*

t_{eyes} .... *Eyes-on-road time (in s)*

t_{hands} .... *Hands-on-time (in s)*

TOC-rating .... *Take-over controllability rating*

t_{steer} .... *First steering maneuver (in s)*

TTC .... *Time-to-collision*, *Time-to-collision*

$TTC_{\_MIN}$ .... *Minimal time-to-collision (in s)*

1. Introduction

Automation is one of the big trend topics of today. The origin of the word can be found in the ancient Greek adjective αὐτόματος (autómatos). This word is composed of the word-components αὐτός (autós) "itself, self-acting" and the participle μάτος (matos) "think, want". The term automation can either refer to work processes (i.e. to automate) or to the finished product (i.e. automated objects).

Thus, *automation* characterizes the (inherent) efforts of systems to autonomously achieve goals, to follow changing goals, to set and maintain goals, or, if goals are achieved, to develop activities to stabilize the system despite existing disturbances (Weller, 2008).

Another definition of automation is given by the Deutsches Institut für Normung e.V. (DIN) (1998): „Automation is […] equipping a facility so that it operates wholly or partially without human involvement."

The automation of processes and systems has a long history. Already in the 18th century people made use of automation of processes and work actions. For example in the patent of Edmund Lee (1745), it is described, how windmills can turn into the wind independently by machines that were powered by the windmill itself. Previously, this work had to be done by humans or animals. In order to make windmills more efficient, in the Middle Ages windmills were built in such a way that they could be turned around a vertical axis. This made it possible to turn the windmills with muscle power in the direction of the wind so that they could continue to work. With Lee's new invention, which attached an additional wind-wheel with a turning-mechanism to the windmill, the windmill reacted independently to the changing wind directions.

In the following decades, the age of the industrialization began. With advances in mechanics and new drive technologies (i.e. the steam engine) mass production in factories became possible whereby animal and human power became more and more replaceable by engines.

In 1787, Edmond Cartwright, for example, invented an automated weaving machine, the so-called *power loom* (Radcliffe, 1828). The development of these machines finally led to the first negative consequences for people: in 1811, unemployed weavers protested against the machines and their supporters.

In the following years, automation continued to progress. With the invention of electricity, increasingly efficient production techniques such as assembly line work were pushed forward. As a result, the working environment changed enormously: work became more monotonous and thus the demands on workers increased.

After automation had already progressed in the industry, at the beginning of the 20th century it also extended to private households. Refrigerators, which cool independently, replaced the iceboxes, which had to be filled with ice manually, and heaters with thermostats drastically simplified temperature regulation in the households.

Many other technical achievements were necessary to enable automation as we know it today. By the development of transistors it was possible to manufacture electrical circuits clearly smaller, whereby Boolean algebra could be used with less effort. Soon, integrated electrical circuits made it possible to equip devices with logic.

Through digital- and computer technology, the degree of automation could be further increased. Nowadays sensors and actuators communicate with each other and ensure a constant quality of the products even with fluctuations in the processes.

Industrial robots and automatic production lines became the state of the art in industrial countries. Techniques such as pattern recognition and artificial intelligence have not yet brought an end to automation.

Nowadays many activities can be carried out to a large extent self-actingly with the help of automation of machines. In addition to increasing productivity, this also leads to more accuracy and a higher speed in production. As a result, workers nowadays are often no longer exposed to the dangers that they were previously exposed to, and the physical health of the workers can be improved.

But increasing automation has not only led to positive consequences for society. The role of employees has often changed drastically. Where workmen used to work as manufacturers and operators of machines, today they are more likely to be involved in administrative, planning or purely monitoring activities.

Negative consequences, which can occur due to the new role of the human, were already described in the past. In *Ironies of automation* Lisanne Bainbridge (1983) describes how automation of industrial processes may rather expand than eliminate problems with the human operator. In this work she points out, that the more advanced a (automated) control system is, the more crucial may be the contribution of the human operator. She especially highlights two different ironies (which can be described as the direct opposite of what might be expected), that come along through the expectation of the system designer who wants to eliminate the error-prone human worker from the working system: Imperfections of (interaction) design of the automated systems can lead to operating problems and the operator, who should have been eliminated completely from the system, is still needed for tasks that the automated system cannot perform on its own.

A short example is used to illustrate this briefly:

In modern airplanes the autopilot controls the flight during most of the time, so that the human pilot normally does not have to fly manually. The main task of the human pilot is to monitor the functionality of the system. Under normal circumstances the autopilot takes over the routing and also the landing of the airplane. Only when the conditions for the autopilot are too difficult, however, the human pilot has to intervene and take over control of the airplane manually. This causes some problems:

- Manual control problems: if the pilot would have controlled the aircraft manually during the flight, the pilot would know how the aircraft reacts to manual inputs. Flying by autopilot, the pilot is inexperienced when taking over control in the first time and may have to wait for feedback from the system.

- Cognitive control problems: to generate strategies for unusual situations an adequate knowledge of the process is necessary. This knowledge depends on frequency of use and the knowledge develops only through use and feedback about its effectiveness.

- Problems due to system monitoring: it is well known that maintaining effective visual attention towards a source of information on which very little happens is impossible for even a highly motivated human (Mackworth, 1950). The longer the system works properly, the less efficient will be the reaction of the supervisor. It is also known that monotonous and passive tasks (like system supervising) can lead to fatigue.

This example should demonstrate that automation can reduce the humans' workload on the one hand, but on the other hand it is also associated with major problems, especially in unpredictable situations. Similar examples can be found for almost all sectors in which automation has already been implemented.

One segment that has been little automated to this date but will be affected by increasing automation in the coming years is that of the automotive sector. In the past, driving has been made much easier and safer for people through different driver assistance and safety systems. The human drivers' driving task has so far been considerably simplified. For example, there are systems that start the vehicle independently without the driver having to do much hands-on work, as it used to be in the past. Different control and regulation systems ensure increased safety. However, until today human drivers are responsible for the execution of the driving task. But it is expected, that also this task will be fulfilled by automation in the upcoming years.

One of the main reasons for automating the driving task is the savings in time. In the additional time saved through increasing automation, drivers are expected to be able to work on *non-driving-related-task*s (NDRTs) or to relax.

However, in the first level of driving automation, *conditional driving automation* (CDA) in which the driver can completely turn away from the driving task, sleeping will not be allowed. In this level of automation, the human driver still is responsible for the driving task when the system requests him to intervene and to take over control of the vehicle. In contrast to the highly trained pilots of airplanes, however, the human driver has considerably less time in such a takeover situation. In addition, the traffic scenarios can be considerably more complex, which can further affect the appropriate response of a driver. Furthermore, an adequate driver state seems to be appropriate for a good performance in such situations. One aspect of the driver state that seems to be important for a good drivers' take-over reaction is the energetic state or the fatigue state of the driver. The fact, that task-engagement can positively as well as negatively affect the fatigue state of humans is known from experiments from the past. Similar effects are expected to occur due to NDRT engagement in CDA.

In order to understand how the driver state can affect the take-over performance of the driver and how NDRTs can influence the driver state, four experiments were conducted within the framework of this dissertation.

2. Theoretical Background

Automated Driving Systems (ADS) consist of hard- and software that are collectively capable of performing the entire dynamic driving task (DDT) on a sustained basis, regardless of whether it is limited to a specific operational design domain (SAE, 2018). This causes, that the DDT is gradually transferred from the human driver to the vehicle – with enormous benefits for the driver himself and for the whole society:

- Improvement of driving safety (Hummel, Kühn, Bende, & Lang, 2011; Meyer & Deix, 2014; Trimble, R. Bishop, Morgan, & Blanco, 2014)
- Improvement of driving comfort (Payre, Cestac, & Delhomme, 2014; Strand, Nilsson, Karlsson, & Nilsson, 2014)
- Improvement of efficiency (Wallace & Silberg, 2012)
- Improvement of productivity (as the driver can engage in NDRTs)

But if you take a more differentiated look, it becomes clear that increasing automation will not solve all problems undoubtedly. One of the main arguments in favor of increasing automation in the driving task is that it will increase safety for the human driver. This is often accompanied by the argument that the human driver is the root cause of accidents. However this reasoning ignores that accidents are caused through multi-causality and are rather rare. Thus, the human driver is not cause, but often, the very last and important safety component in the system (Bengler, Winner, & Wachenfeld, 2017).

From research of the past it is also well known, that in addition to positive effects, automation can also lead to negative effects for the human operator. In her work *Ironies of Automation*, Bainbridge (1983) describes these possible effects more precisely. The irony is described as a combination of circumstances which leads to the direct opposite of what might be expected: “The classic aim of automation is to replace human manual control, planning and problem solving by automatic devices and computers.” But it is also known that even highly automated systems need human beings for supervision, adjustment, maintenance, expansion and improvement (Bibby, Margulies, Rijnsdorp, Withers, & Makarov, 1975). Thus, one can conclude, that these automated systems still are human-machine systems. Bainbridge (1983) adds that the more advanced such a control system is, the more crucial may be the contribution of the human operator.

In order to better understand the aim of this dissertation, in this chapter the following theoretical issues are presented in more detail:

In CDA, the type of driving automation that was investigated in this thesis, the human driver can completely turn away from the driving task, but if the system requests the driver to take over control, the driver must be able to recover the driving task in the available time budget. A detailed description of the history of driver assistance systems as well as an overview of the different levels of automation is given in sections 2.1. and 2.2.

As long as the system for automated driving is active, the human driver can engage in NDRTs. From research of the past, it is well known, that task-engagement can lead to positive as well as to negative consequences regarding the fatigue state of humans (Mackworth, 1950). Especially passive tasks are connected to a decrease in human performance. Similar effects are supposed to occur in the context when people either have to permanently monitor the system during automated driving or when they engage in tasks that rather tend to fatigue them. In other words, when tasks do not challenge the driver due to their passivity, which is a known causation for fatigue. How such aspects may affect the take-over performance in CDA is described in 2.3.

In CDA increasing monotony due to the system executing the DDT and the human driver monitoring the system can lead to fatigue. Both, increasing automation as well as monotony are well-known causations for *passive task-related fatigue*, which leads to the same consequences as all forms of fatigue: an increased crash risk and impaired driving performance. This may be a problem, when a take-over situation suddenly appears in CDA. In section 2.4. the different forms of fatigue are described in more detail. Also see section 2.4. for related work on this topic.

Due to the potential risk for the participants, for reasons of reproducibility, and for reasons of realization of the CDA function, the experiments were carried out in simulated environments. These environments are explained in section 2.5.

Also the human-machine-interaction (HMI) concept has a major influence on the human performance in take-over situations in CDA. Therefore, in section 2.7., the role of HMI concepts in CDA is explained.

In section 2.5. the different measurements for emerging fatigue are described. Therefore, the measurements were divided into different categories and are described in more detail. Advantages and disadvantages of the individual measurement-methods are addressed.

As one focus of the present thesis was to investigate how take-over performance of the human driver is affected through the driver’s state, in section 2.8. it is described how the performance of the human driver can be measured in such situations.

To close the theory section, in section 2.9., a full problem statement and the research questions that should be addressed in this dissertation are given.

2.1. From Driver Assistance Systems To Automated Driving

In the last decades, enormous developments in driver assistance systems have made driving safer and more comfortable. Classic driver assistance systems like the *Anti-lock Braking System* (ABS) or *Electronical Stability Control* (ESC) support the driver in emergency situations for many years now. Through improvements in the field of sensor technology and signal processing in the last decade, driver assistance systems changed to *Advanced Driver Assistance Systems* (ADAS). Nowadays even some parts of the driving task are already executed by automated systems. The introduction of further automated driving functions is only a matter of time right now. In this section a brief overview of the development of driver assistance systems to ADAS is given.

The present thesis was written according to the *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems* of the society of automotive engineers (SAE) (SAE, 2018). In this taxonomy, driver assistance systems and levels of automated driving are defined and classified. Therefore, the driver assistance systems presented in this section are also classified according to this taxonomy. The taxonomy is further explained in the following section 2.2.1.

All started in the 1950s, when the first driver assistance system, the *Cruise Control*, was developed to support the driver in the driving task. The system was able to control the speed of the vehicle and to keep it at a constant level. Under the name *Speedostat*, it was commercially introduced by Chrysler in 1958 (Holloway, 1966).

Some years later, in 1966, the *Jensen FF* was the first car that used an ABS. The ABS that was used in this car was a mechanical system. The first electronical ABS was introduced in 1978 by Bosch. The first vehicles that used this system were the *Mercedes S-Class W 116* and the *BMW 7 series E23*. Since the year 2004, all vehicles under a weight of 2.5 t standardly have to be equipped with an ABS due to an agreement of the *Association des Constructeures Européens d’Automobiles* (ACEA).

After the introduction of the *Traction Control System* (TCS) in about 1985 the next big milestone in the field of driver assistance systems was the ESC in 1995. A microcomputer that monitors the signals from different sensors detects when the vehicle starts to skid and can intervene if necessary. Through individual braking interventions for each wheel and throttling of the engine power, this system ensures stability so that the vehicle keeps the track.

The next driver assistance systems that were developed focused on an increased comfort for the human driver while driving. Since 1998, *Adaptive Cruise Controls* (ACC) support the driver in longitudinal vehicle control. Through acceleration and deceleration inputs, the vehicle keeps a before preset distance to the vehicle in front. Currently available systems are even able to brake to a standstill and start up again after an approval by the human driver. This system was actually invented for an increase in comfort, however, it also increases road safety as it can support the driver in sudden emergency braking situations, especially when the driver is inattentive. According to the SAE taxonomy this system can be seen as the first SAE Level 1 *Driver Assistance* system (SAE, 2018).

Two other ADAS that can also be considered as Level 1 *Driver Assistance* systems are the *Lane Departure Warning* (LDW) and the *Lane Keeping Assistant* (LKA). The difference between these two systems is that the LDW only warns the driver when the vehicle drifts from the current lane (through vibrations on the steering-wheel or the seat) whereas the LKA actively keeps the vehicle in the lane through steering inputs. Thus, both assistance systems support the driver in the lateral control of the vehicle.

A combination of these two Level 1 *Driver Assistance* systems, more detailed a system that supports the driver in both, the lateral and longitudinal vehicle control, is the next type of automation, SAE Level 2, *Partial Automation*. These systems are state of the art at most automobile manufacturers right now. Typical systems for this type of automation are parking assistant systems that execute longitudinal and lateral control in slow speed parking situations or ADAS that support the driver in longitudinal and lateral control like Tesla's *Autopilot*, Cadillac's *Super Cruise* or BMW's *Driving Assistant Plus / Professional*. However, in order to be more precise, these systems must permanently be monitored by the human driver. An engagement in NDRTs is not possible and not permitted.

The next levels of automation (i.e. SAE Level 3 – 5) in driving that will be available for the public market are the first systems that allow the human driver to completely turn away from the driving task and to engage in NDRTs. Whereas in Level 3, CDA, a human driver still has to be able to resume control of the car, from Level 4 on, an intervention by a human driver due to a *request-to-intervene* (RtI) is no longer necessary. Level 3 automation or CDA for roadway travel (e.g. on the Autobahn or on a highway) is expected to be introduced to the consumer market within the next years (Belz et al., 2017).

2.2. Taxonomies Of Automated Driving

Apart from general taxonomies on automation (e.g. Parasuraman, Sheridan, & Wickens, 2000; Riley, 1989), two have become widely accepted in the field of automated driving. On the one hand, there is the taxonomy of the German Bundesanstalt für Straßenwesen (BASt) (Gasser et al., 2012) which is widely used in Germany and on the other hand, the taxonomy of the SAE (SAE, 2018) which has established itself internationally. In the following two sections, these two taxonomies are further described.

2.2.1. The Taxonomy Of The BASt

In 2012, a group of experts from the German BASt assessed the legal consequences of increased vehicle automation according to the German law. The expert-group included experts from different domains like the German automotive industry, component suppliers and academia. In their final report, the experts distinguish between three levels of automation, partial automation, high automation and full automation. Next to these levels of automated driving functions, the level *Driver Only* and *Assisted* complete the taxonomy. The different degrees of automated driving and the tasks to be either performed by the system or the human driver are further described. See Table 1 for the BASt taxonomy.

Table 1: The different degrees of automated driving and their definitions according to the BASt (according to Gasser et al., 2012).

Nomenclature	driving tasks of the driver according to degree of automation
Full Automation	The system executes lateral and longitudinal guidance completely in a defined case. - The driver is not required to monitor the system. - Before leaving the application case, the system requires the driver with a sufficient time reserve to take over the driving task. - All system-boundaries are recognized by the system. - The system is always able to return to a minimum risk state.
High Automation	The system executes lateral and longitudinal guidance (for a certain period of time and / or in specific situations). - The driver is not required to monitor the system. - If required, the driver has to take over the driving task within a sufficient time reserve. - All system boundaries are recognized by the system. - The system is not able to return to a minimal risk state at all time.
Partial Automation	The system executes lateral and longitudinal guidance (for a certain period of time and / or in specific situations). - The driver is required to permanently monitor the system. - The driver must at all time be prepared to take over vehicle control. - The human driver is responsible for the driving task.
Driver Assistance	The driver permanently executes either the lateral or the longitudinal guidance and the other driving task is executed by the system within certain limits. - The driver is required to permanently monitor the system. - The driver must at all time be prepared to assume full responsibility for vehicle control.
Driver Only	The driver executes the longitudinal and the lateral guidance himself permanently during the entire ride.

Capabilities of Automation

2.2.2. The Taxonomy Of The SAE

The internationally most widespread taxonomy is the one of the SAE. In their information report J3016 *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems* (SAE, 2018), the SAE also defines different levels of automated driving functions. The first version of the document was published in 2014. In 2016 and 2018 some major adaptions followed.

In Table 2, an overview of the different levels according to the SAE is displayed. Next to the name of the level of the automated driving type, the table includes a narrative definition and identifies who (the driver or the system) is responsible for which subtasks of the driving task. These include the execution of the lateral and longitudinal motion control as well as the object and event detection, recognition, classification, and response (OEDR). Next to that, the table shows if the driver or the system is responsible as fallback-performance in the case the system reaches its limits.

In all experiments conducted as part of this thesis, the driving simulator systems as well as the *Wizard-of-Oz* vehicle simulated SAE Level 3, *Conditional Driving Automation* functions. The majority of the experiments focused on take-over situations, in which the human driver had to regain control of the vehicle due to a RtI issued by the system.

Table 2: The different degrees of automated driving and their definitions as defined by the SAE (SAE, 2018) (DDT: dynamic driving task; OEDR: Object and event detection, recognition, classification, and response; ODD: operational design domain; ADS: automated driving system)

Level	Name	Narrative definition	DDT		DDT fallback	ODD
			motion control	OEDR		
Driver performs part or all of the DDT						
0	No Driving Automation	The performance by the *driver* of the entire *DDT*, even when enhanced by *active safety systems.*	Driver	Driver	Driver	n/a
1	Driver Assistance	The sustained and ODD-specific execution by a driving automation system of either the lateral or the longitudinal vehicle motion control subtask of the DDT (but not both simultaneously) with the expectation that the driver performs the remainder of the DDT.	Driver & System	Driver	Driver	Limited
2	Partial Driving Automation	The sustained and ODD-specific execution by a driving automation system of both the lateral and longitudinal vehicle motion control subtasks of the DDT with the expectation that the driver completes the OEDR subtask and supervises the driving automation system.	System	Driver	Driver	Limited
ADS performs the entire DDT						
3	Conditional Driving Automation	The sustained and ODD-specific performance by an ADS of the entire DDT with the expectation that the DDT fallback-ready user is receptive to ADS-issued RtIs, as well as to DDT performance-relevant system failures in other vehicle systems, and will respond appropriately.	System	System	Fallback ready driver	Limited
4	High Driving Automation	The sustained and ODD-specific performance by an ADS of the entire DDT and DDT fallback without any expectation that a user will respond to a RtI.	System	System	System	Limited
5	Full Driving Automation	The sustained and unconditional (i.e., not ODD-specific) performance by an ADS of the entire DDT and DDT fallback without any expectation that a user will respond to a RtI.	System	System	System	Unlimited

In contrast to the before mentioned BASt taxonomy, the SAE distinguishes between six levels of automated driving functions compared to five in the BASt taxonomy. While the

SAE levels 0, *No Driving Automation*, 1, *Driver Assistance*, and 2, *Partial Driving Automation* correspond to the first three levels of the BASt (*Driver Only*, *Driver Assistance* and *Partial Automation*), differences in the higher levels of automation appear. In the SAE definition, the *High Automation* from the BASt is further divided into two subcategories:

- In SAE level 3, *Conditional Driving Automation* (CDA), the driver is required as fallback-performance and has to intervene if requested by the system. In such take-over situations a control transition from the system to the human driver (Driver Only or Assisted) has to be executed within the available time budget. This can either be the case due to known system boundaries (unurgent or long-term take-over situation; i.e. due to knowledge deposited in the map) or due to system-boundaries detected by on-board sensors (urgent or *short-term* take-over situations).

- SAE level 4, High Automation, instead is capable of executing minimal risk maneuvers if the human driver does not respond appropriately to a request-to-intervene (RtI).

Thus, inconsistencies between these two taxonomies become obvious. The SAE *High Automation* refers to another Level than the *High Automation* level of the BASt (which actually is *Conditional Automation* referring to the SAE).

The highest levels in these two taxonomies, SAE Level 5, *Full Automation* and the BASTs *Full Automation* represent a similar system status. In these highest levels of automation, the systems guarantee functionality in all situations.

The level of automated driving which was used in the experiments that are reported in this thesis correspond to the SAE Level 3, *Conditional Driving Automation* or to the BASts Level *High Automation.* Due to the international standard of the SAE taxonomy, this taxonomy is used in the further course of this thesis.

2.2.3. Particularities And Take-Over Situations In Conditional Driving Automation

Vehicles equipped with CDA functionality are short before being introduced to the consumer market. With the *Audi AI Staupilot* (traffic jam pilot), the German car manufacturer was the first that announced SAE Level 3 driving functions in a series production car (Netter, 2017). However, the *Audi AI Staupilot* is only available in slower speed driving situations up to 60 km/h (37.3 mph) on the highway. Also other vehicle manufacturers have announced the first vehicles with automated driving functions for the next few years.

BMW announced the first CDA vehicle for 2021 (BMW AG, 2016), Daimler for 2020 (Automotive News Europe, 2018) and all Tesla models are already equipped with hardware designed for higher automated driving functions (Tesla Motors, 2016). With the introduction of these CDA systems, new possibilities for the human driver will arise. However, there are some particularities that have to be considered.

CDA will first only be permitted on designated roads or areas (e.g. on the Autobahn or highway). When the human driver enters a road where CDA is permitted, the driver can activate the system. From then on, the system executes the lateral and the longitudinal control of the vehicle. The human driver is not responsible for the driving task anymore and not even has to monitor the driving environment. For the human driver this means, that he can engage in NDRTs or he can simply enjoy the automated ride and relax. However, in CDA the human driver represents the fallback for the DDT and therefore has to regain control if requested by the system due to take-over situations (see chapter 2.3. for further information). Thus, an adequate state of the driver is required for safety reasons (Damböck, 2013). Sleeping for example is not possible for the human driver in CDA.

All in all, there can be different reasons for such take-over situations. Gold (2016) differentiates between three reasons:

- End of the automation scenario: CDA will initially only be available on approved roads or highways. When the end of such a road (e.g. exit of the highway, end of the highway) will be reached, the human driver has to regain control of the vehicle. Such system-limits are stationary, can be stored in maps and are therefore permanently available to the system. Next to roads, weather conditions like fog, rain or snow must be suitable for CDA. If these conditions change to an improper state, this can also lead to an end of the automation scenario and consequently to a take-over situation. In these end-of-automation scenarios the human driver can be prepared for the take-over situation some time before he actually has to take over control of the vehicle. Thus, these situations can be categorized to *long-term* take-over situations.

- Failure of sensors: The functionality of automated driving functions like CDA requires accurate data merged from many different sensors. A lack of such sensor data, even if the data is redundant, may impair safety in automated driving. Thus, if the system detects a non-functionality of sensors the human driver will be requested to take over vehicle control. The vehicle may be guided based on the last sensor data until the human takes over control. The available time-budget in such situations depends on the amount of the afflicted sensors, the last detections of the sensors and the traffic environment. These situations correspond to *long-term* or *short-term* take-over situations.

- Situation-related take-over: The next reason for the system issuing an RtI may be due to driving-situations detected by on-board sensors that are not capable for the automated driving system. For example, such situations may be (unidentified) objects in the own lane in higher density traffic, persons on the road (e.g. after an accident) or missing lane-markings. If such circumstances are detected by on-board sensors, the human driver will be requested to intervene and take over control of the vehicle. The available time-budget in such situations is limited to just a few seconds. It has to be assumed that these *short-term* take-over-situations are most demanding for the human driver. Therefore, especially these situations were investigated in the context of this dissertation.

A further classification for take-over situations is given by Gold, Naujoks, Radlmayr, Bellem, and Jarosch (2018). In this work, concrete examples of take-over situations in CDA are presented and classified based on the four parameters *time budget / urgency*, *predictability*, *criticality* and *complexity of the drivers' response*. In the following section, these parameters are further described. This is intended to help to understand how take-over situations have to be designed to measure human performance in such situations.

Time budget / urgency: The time budget / urgency of a take-over situation indicates how much time the human driver has to react upon the RtI. A high time budget goes along with low urgency and a low time budget goes along with a high urgency. Gold (2016) differentiates between low, medium and high *temporal criticality*. Petermann-Stock, Hackenberg, Muhr, & Mergl (2013) assume that take-over times (i.e. time the human driver needs to react upon an RtI) will be at least between 5 - 10 s. The higher the available time-budget is, the more likely is that the human driver will successfully take over control.

Predictability: A scenario can either be high or low predictable. If imminent system boundaries are known to the automated driving system long before they actually appear, and thus the driver can be informed about the upcoming intervention a long time before he actually has to take over control, the situation is highly predictable. Such situations can for example be known based on map-material and include end of highways and construction zones. On the other hand, there are take-over situations with low predictability. These situations are detected by on-board sensors and involve *short-term* take-over situations. Examples for these situations are for example obstacles on the road that are detected by on-board sensors.

Urgency and *predictability* of a take-over situation often depend on each other: if a take-over situation suddenly appears and gets detected by on-board sensors this comes along with a rather small time-budget. On the other hand side, if a take-over situation is known from the backend the time-budget for the intervention of the driver is high.

Criticality: The criticality of a testing-scenario indicates the potential risks of a take-over scenario in the case that the human driver cannot take over control of the car in the available time-budget. A low critical scenario, for example, would be accidentally passing an exit ramp on the highway whereas a highly critical scenario would be an obstacle or accident on the own lane the ego-vehicle would crash in, if the driver does not react appropriately. Thus, the criticality of the scenario is determined by the situation.

Complexity of the drivers' response: The complexity of a take-over situation refers to the required responses of the driver. In different situations different driver reactions may be necessary or appropriate. They can either be complex (e.g. evasive maneuver or lane change-maneuver), or rather simple (e.g. stabilizing the vehicle in its lane). The more opportunities the driver has to react upon the RtI, the more complex the situation gets. Next to the number of different possible reactions of the drivers, the surrounding traffic and the traffic density also affect the complexity of a take-over situation.

According to these four parameters, different take-over scenarios were classified. An expert group assessed scenarios and classified how the particular scenario demands the human driver from low to high effort. Furthermore, certain test scenarios for different research questions have been identified. For testing of the maximum driver performance, a *highly urgent*, *poorly predictable*, *highly critical* and *rather demanding* scenario is suggested. Possible scenarios that are proposed are: total sensor failure, obstacle ahead detected by on-board sensors or the loss of the reference signal (i.e. lane markings) (Gold et al., 2018).

2.3. Aspects Of The Driver State in Conditional Driving Automation

A term that was first used in the context of CDA with regard to the driver is the German term W*ahrnehmungsbereitschaft* (engl.: readiness to perceive) that was introduced within the regulatory for automated driving functions (§ 1b StVG). The text of the regulation implies that the human driver can turn away from the driving task while CDA or higher automated driving functions are active. However, this regulatory also implies that the driver must remain perceptive that he can fulfill his obligation to be able to regain control of the vehicle if the system requests him to do so. Additionally the human driver has to recognize (due to obvious circumstances) that the prerequisites for the intended use of such automated driving functions no longer exist. More specifically, this means that the driver can take his hands off the steering-wheel, turn his gaze away from the road and engage in other activities. However, he must remain so perceptive that he can capture situations (i.e. acoustical and visual RtI) and then regain control of the vehicle. The circumstances which the driver must perceive must be so obvious that they can also be seen when turning away from the traffic situation. This could be assumed, for example, if the driver is made aware of driving errors or technical faults of the system by other vehicles honking their horns, or if the system has performed an emergency stop without external cause (Deutscher Bundestag, 2017).

In that sense, an adequate driver's state in CDA is necessary for safety reasons: when it comes to a take-over situation, the driver must regain control of the vehicle within the available time-budget. The remaining of control, the so-called *control transition process* can be affected by different aspects of the drivers' state. These include not only pure driver-related aspects, but also those that are more likely to be influenced by the activities and / or NDRTs the driver is engaged in while driving automated (Naujoks, Befelein, Wiedemann, & Neukum, 2018; Buld & Krüger, 2005).

In the *Take-Over Process Model* published by Marberger et al. (2018), it is displayed how a control transition from the system to the human driver should proceed in CDA. This model is introduced in section 2.3.1.

A central part of this model is the *Driver state transition*. In this phase a *task-switch* from the current activity (i.e. engaging in a NDRT) to manual driving is required for the human driver. This task switch is supposed to afford a time consuming re-configuration of different aspects of the drivers' sate. Which aspects can be involved and how these may further affect take-over performance is described in section 2.3.2.

2.3.1. The Take-Over Process Model - From Automated To Manual Driving

When it comes to a take-over situation in CDA a task-switch from automated driving to manual driving, has to be fulfilled by the driver. Therefore the human driver has to end the current NDRT and turn to the driving task. In the best case, the system maintains the automation until the driver has completely taken over vehicle control.

The tasks for the driver and the system can be further divided into sub-tasks and sub-states which should be performed in a predefined order. Such a subdivision of the transition process is also necessary when it comes to an evaluation of the performance of the drivers in a take-over situation. Therefore, to evaluate a human driver's reaction in such a take-over situation, reaction times based on the RtI can be used.

These different subtasks and states are specified and described in more detail in the following. Marberger et al. (2018) has taken a closer look at the procedure that is necessary when a driver is requested to take over vehicle control. In order to make the process easier to understand, a division between the tasks and states of the human driver and the CDA system was made. See Figure 1 for the transition process model according to Marberger et al. (2018). The numerals shown in the graphic are explained below.

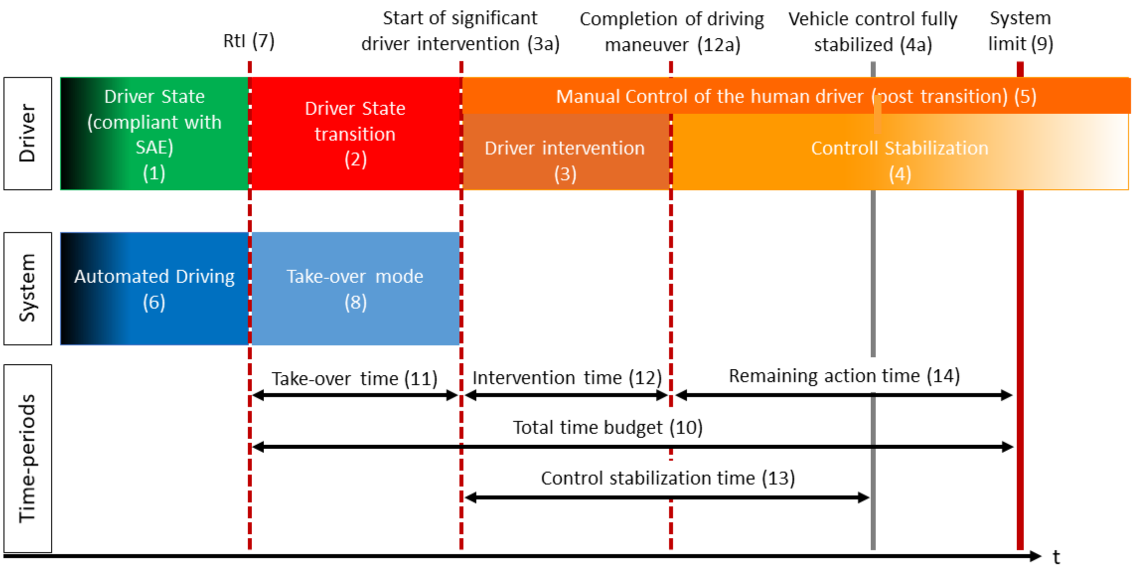


Figure 1: Take-over Process Model. Adapted from Marberger et al., 2018. Copyright 2018 by Springer International Publishing AG.

States related to the human driver in conditional driving automation:

(1) *Driver State (compliant with CDA):* In CDA, a fallback-ready driver is absolutely necessary, as the system hands over vehicle control to the human driver, when system limits are reached. Thus, falling asleep or other conceivable activities which would make a successful take-over less likely must be prevented in CDA. This state ends with the issuing of a RtI.

(2) *Driver state transition:* In this phase of the transition process, the human driver is required to end the current activity and to (re)focus towards the driving task. Therefore the driver must perceive, understand and react upon the RtI. This, for example,

includes mental reprocessing towards the driving task and an eyes-on-road reaction. This complex process is further described in 2.3.2.

(3) Driver intervention / Deactivation of AD: Due to a significant intervention [(*3a) Start of significant driver intervention;* e.g. steering > 2° or braking > 10%], the human driver deactivates the CDA system. This process also transfers the responsibility from CDA to the human driver. Depending on the system, different mechanisms to regain vehicle control exist: switch off the system via a button, executing a significant steering intervention or executing a significant braking intervention.

(4) Control stabilization: In this phase, the human driver must fully reestablish vehicle control of the automated driving system. If the human driver is capable of the control stabilization, this phase ends with *vehicle control fully stabilized (4a)*.

(5) *Post transition (manual) control (consists of phase (3) and (4):* In this phase of the transition process, the human driver is in charge to fully control the vehicle. Thus, this phase is in focus of research focusing on the human drivers' reaction in take-over situations. In the funding project Ko-HAF, it was distinguished between a first reflex-like *driver reaction*, the so-called preparation of driver readiness, and a first conscious *driver intervention*. The first *driver reaction* was initiated by a hands-on-steering-wheel detection (t_{hands}) or a first brake-pedal contact ($t_{brake\_reaction}$). A *driver intervention* on the other hand was characterized by an exceeding of the steering-wheel angle over 2° (t_{steer}) or a brake pedal shift of more than 10% ($t_{brake\_maneuver}$).

States related to the CDA system:

(6) *Automated Driving (CDA):* In this state, the system executes the CDA function. The system performs longitudinal and lateral control of the vehicle while the human driver can engage in NDRTs as he is not supposed to monitor the automated driving system or the traffic environment. When the vehicle is driving automated and detects an event or a condition which it cannot handle, an RtI is issued.

(7) *Request to intervene (RtI):* An RtI is a system generated warning message that requests the driver to remain vehicle control. When an RtI is issued, the system changes its status to the *take-over mode*.

(8) *Take-over mode:* The take-over mode is the status of the automated driving system after an RtI has been issued. For a CDA vehicle, this means, that the system must at least continue the execution of the driving task for a sufficient period of time to

bridge the gap between automated driving and the required manual driving of the human driver.

(9) *System limit:* Situation that cannot be handled by the automated driving system and which causes an RtI, when detected by the system. Possible examples for such system limits are sensor failures, detected obstacles on the road ahead or the end of the approved road for CDA.

Time-periods related to the take-over reaction

(10) *Total time budget*: This phase represents the maximum of the available time for the driver intervention. This time window starts with the RtI issued by the system and would end with the system reaching the system-limit. Of course, in the best case, the human driver can control the vehicle before the system hits it´s limits. The *total time budget* plays a decisive role in the take-over process as it determines how much time the human driver has for his intervention.

(11) *Take-over time*: Time-window in which the human driver has to realize the RtI and to decide for a reaction. This period starts with the issuing of the RtI by the system and ends with the first intervention of the driver.

(12) *Intervention time*: This phase starts with the first intervention of the driver (3a) and ends with the completion of the driving maneuver (12a).

(13) *Control stabilization time*: This phase starts with the significant driver intervention (3a) and ends with the phase vehicle control fully stabilized (4a).

(14) *Remaining action time*: Remaining time period after the completion of the driving maneuver (12a).

2.3.2. The Concept Of Driver Availability In Conditional Driving Automation

The most essential and challenging part in the before mentioned *take-over process model* from CDA to manual driving is the *driver state transition process* (Marberger et al., 2018). In this phase the human driver has to reconfigure all aspects of the drivers' state from the current NDRT or activity to the manual driving task. For this purpose, the driver has to fulfill a change from a current state (which is affected by the current NDRT) to a target state (which is necessary for manual driving). This target state can also be described as the *readiness* of the driver to fulfill a take-over reaction (ISO/TR 21959-

1:2018, 2018). The drivers' state is a complex construct and consists of several parameters. Marberger et al. (2018) categorized the following aspects: the *sensory state*, the *motoric state*, the *cognitive state*, the *arousal level (i.e. energetic state)* and *motivational conditions* that may affect the take-over time and / or take-over quality. A further explanation of these different aspects of the driver state and how these aspects may be influenced by different NDRTs is given in the following. A schematic representation of this process can be seen in Figure 2.

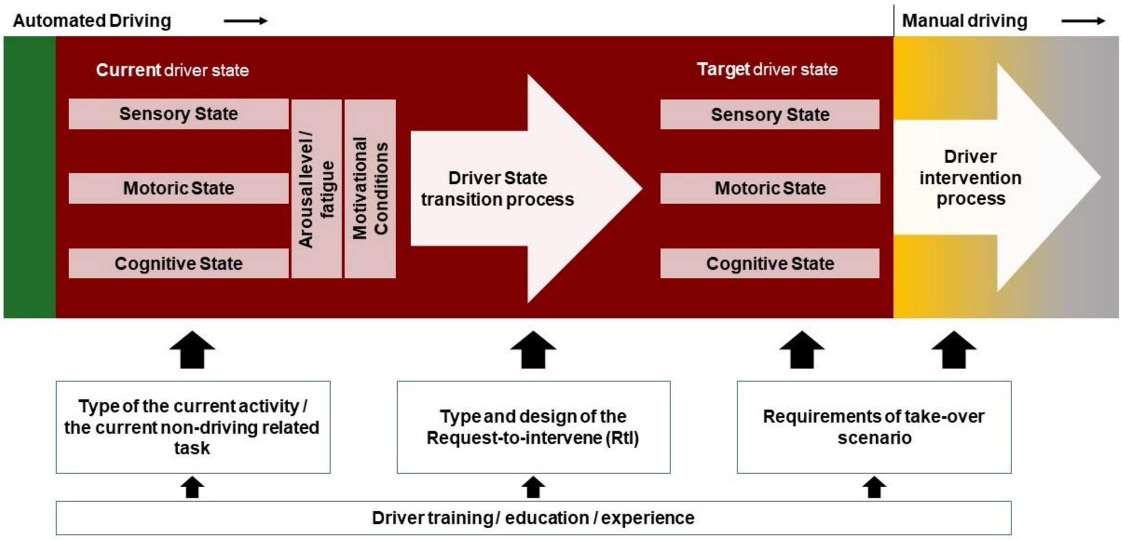


Figure 2: Driver State Model. Adapted from Marberger et al., 2018. Copyright 2018 by Springer International Publishing AG.

The different aspects of the driver state:

Sensory State: The sensory state describes what the driver can currently perceive with his sensory systems (i.e. current state) and the kind of information that is required in a specific take-over situation (i.e. target state). For example, watching a movie when driving with an activated CDA system would cause an additive visual load. Listening to music via headphones could result in the driver no longer being able to notice an acoustic RtI signal. Aspects of the sensory state have been investigated by Naujoks, Mai, and Neukum (2014) and Gold, Berisha, and Bengler (2015).

Motoric State: The motoric state of the human driver can further be distinguished between the overall driver posture on the one hand, as well as the occupation of the hands of the driver on the other hand. The posture of the driver is determined by the position of the driver in the vehicle: is the driver in a position whereby a fast take-over reaction is made possible (sitting in an upright position in direction of the ride) or is he lying back so he cannot even reach the steering-wheel for an intervention? Furthermore, not only the drivers' posture but also the occupation of the drivers' hands can affect the take-over performance. When an RtI is issued and the hands of the driver are occupied, the object must first be put away for taking over lateral guidance. Zeeb, Härtel, Buchner, and Schrauf (2017) investigated such aspects before.

Cognitive State: The cognitive aspect of the driver state relates to the drivers' reconfiguration of mental task sets or response rules upon an RtI (Kiesel et al., 2010). When an RtI is issued, the driver is required to perceive the situation in full context and then make decisions for a correct reaction. Cognitively demanding activities the driver is dealing with while driving automated may affect his ability to perceive the situation correctly. How cognitive tasks affect take-over performance was previously investigated by Gold, Körber, Lechner, and Bengler (2016) and Neubauer, Matthews, and Saxby (2012).

Motivational Conditions: If an activity strongly binds the driver during CDA due to its inherent motivation, this may also affect take-over performance when a RtI is issued. Therefore, lock-out approaches (i.e. black screen or take-over signal on the device), seem helpful for improving take-over performance. Of course this is only possible for electronic devices. Wandtner, Schömig, and Schmidt (2018) have conducted research in this area before.

Arousal level / fatigue: In CDA a monotonous and rather passive situation for the human driver may occur, when the system is performing the driving task (Neubauer, Matthews, Langheim, & Saxby, 2012). Increasing automation as well as monotony are well known causations for fatigue (May & Baldwin, 2009). In manual driving, fatigue leads to deteriorated reaction times and an increased crash risk. Transferred to CDA, this would mean, that fatigue could lead to an impaired take-over performance and increased crash risk. This assumption was examined on the basis of the present thesis. As the investigation of this driver state aspect was the focus of the present work, the construct of fatigue is explained in more detail in the following section 2.4.

2.4. Fatigue

In the field of sleep research the terms sleepiness, drowsiness and fatigue are not clearly distinguished and are therefore commonly used as synonyms. Besides the three mentioned terms, there are many more, which are supposed to describe rather similar constructions or states (Johns, 1998). Nevertheless, a brief introduction to the different terms is given.

Johns (2007) distinguishes between two constructs: on the one hand, there is drowsiness and sleepiness and on the other hand, there are the constructs fatigue and tiredness. The first two terms (i.e. drowsiness / sleepiness) are described as "inclined to sleep, heavy with sleepiness, half asleep, dozing", being synonymously with the adjective "sleepy" (Johns, 2007) whereas the other two terms (fatigue / tiredness), which can also be used synonymous, are described as "weariness resulting from bodily or mental exertion" (Johns, 2007). Thus, these two different constructs (drowsiness / sleepiness and fatigue / tiredness) clearly differentiate from one another in their attribution of effects, measures and consequences. Fatigue, which is rather a "subjective state of weariness" can be relieved by rest, whereas drowsiness can only be relieved by sleep (Johns, 2007).

Another model that can be used for explaining the different forms of fatigue is the model from May and Baldwin (2009): in their model, the two authors distinguish between task-related fatigue (which can be further divided in *passive* and *active task-related fatigue*) and *sleep-related fatigue*. The resulting fatigue of a person is thus composed of the different forms of fatigue. If these two models are being compared, one can assume that the sleep-related fatigue of May and Baldwin (2009) describes the construct that Johns (2007) describes as drowsiness / sleepiness and May and Baldwin's task-related fatigue corresponds to John' s fatigue / tiredness.

Since the model of May and Baldwin (2009) is contextually related on the development and consequences of fatigue in manual driving, and due to the fact that increasing automation is posted as one of the reasons for *passive task-related fatigue* in this model, the following section (2.4.1.) takes a closer look at the fatigue model of May and Baldwin (2009), as in the following this thesis refers to this model.

In this present work it was evaluated how different NDRTs affect task-related fatigue of the human driver in CDA, how these changes of the driver state can be detected, and in the end, how this affects take-over performance.

2.4.1. The Fatigue Model Of May & Baldwin

As fatigue is a well-known causation for road accidents in manual driving, it is supposed, that this state may also negatively affect driving performance of the human driver in take-over situations in CDA. In their fatigue model, May and Baldwin (2009) distinguish between three forms of fatigue: *passive task-related fatigue*, *active task-related fatigue* and *sleep-related fatigue*. They also point out possible causations for these forms of fatigue and identify the consequences that result from fatigue. The model is represented in Figure 3.

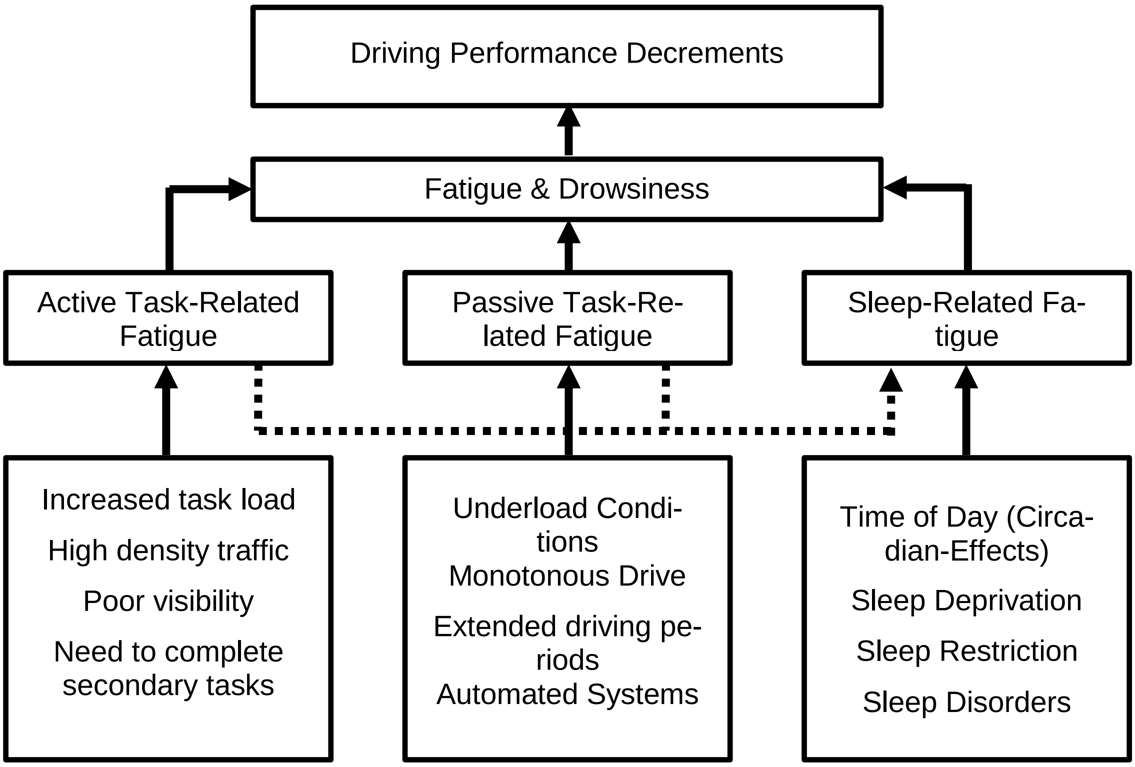


Figure 3: Fatigue model adapted from May and Baldwin (2009).

As can be seen in the model, the authors differ between three forms of fatigue which can occur in manual driving:

Active task-related fatigue: This form of task-related fatigue can be connected to prolonged and continuous perceptual-motor adjustment, mental overload and high demand driving conditions (Gimeno, Cerezuela, & Montanes, 2006). Such overload conditions can occur in manual driving when there are bad sight conditions (fog, snow), high traffic density and / or when the driver is additionally distracted from the driving task by the completion of secondary tasks (input in navigation system). According to Desmond, Hancock, and Monette (1998), *active task-related fatigue* is the most common form of fatigue in manual driving.

Passive task-related fatigue: In contrast to the *active task-related fatigue*, *passive task-related fatigue* occurs when there are only a few or no demands at all on the perceptual and / or motoric system. In manual driving, this can be due to prolonged driving periods in monotonous traffic environments, little traffic densities or when the driving task is too predictable. According to Gimeno, Cerezuela, & Montanes (2006) drivers may rely on simplified mental models of the driving task which leads to reduced effort for the task, which can further increase fatigue.

Sleep-related fatigue: Another form of fatigue, which in contrast to the before introduced forms of fatigue is independent from the current task or activity, is the sleep-related fatigue. This form of fatigue is caused through the circadian rhythm and is associated to sleep deprivation and sleep restriction. Due to the circadian rhythm, human beings get tired when it gets dark in the night and wake up again in the morning. In manual driving, driving performance decreases when sleepiness increases and the longer a person remains awake (May & Baldwin, 2009).

2.4.2. Negative Effects Of Fatigue On Human Performance

It is rather obvious that fatigue goes hand in hand with a reduction in human performance. Especially in higher states of fatigue. Such a reduction in human performance has already been scientifically proven. Also in the driving task fatigue is connected to deteriorated performance. In the next section such effects are further explained.

A commonly used test to examine negative consequences of fatigue is the psychomotor vigilance task (PVT). The PVT is a reaction time task that is about reacting as fast as possible upon a visual stimulus. This task was amongst others used by Jewett, Dijk, Kronauer, and Dinges (1999) to measure performance in relation to different levels of sleep deprivation. In their experiment, participants either slept for *0*, *2*, *5* or *8 hrs* during the night and afterward had to complete the PVT task in the morning at 10 am. In this experiment it could be demonstrated, that participants that slept for a longer period of time performed better in the PVT task. This is reflected in faster reaction times as well as in a fewer number of lapses.

In another experiment, Graw, Kräuchi, Knoblauch, Wirz-Justice, and Cajochen (2004) investigated effects of a 40-hrs sleep deprivation on performance in the PVT task. Performance, which was measured by frequency of lapses and reaction time, decreased when the time awake increased.

However, effects of fatigue on driving performance have also been investigated in the past. For example, Lenné, Triggs, and Redman (1997) examined how sleep-related fatigue affects driving performance in a driving simulator experiment. As results, they reported an impaired driving performance according to the time of the day and the circadian rhythm. In their experiment, participants showed the strongest deviations from their average speed in dependence to the circadian rhythm.

Task-related fatigue in driving was investigated by Matthews and Desmond (2002). In their driving simulator experiment, a high-workload signal detection task was used to induce *active task-related fatigue*. The subjective fatigue state as well as objective performance measures (i.e. heading error) suggest that a general state of fatigue emerged due to the task engagement. After the experiment, participants reported symptoms of fatigue and a decreased energetic state.

In another driving simulator experiment, Ingre, Åkerstedt, Peters, Anund, Kecklund & Pickles (2006) investigated the relationship between subjective fatigue and the probability of accidents and / or incidents. Results suggest, that participants that exceeded the state of being *tired* or *very tired, big problems to stay awake* (Karolinska-Sleepiness Scale) are much more likely to have an accident compared to lower levels of subjective fatigue.

In further experiments possible reasons for this impaired driving performance were investigated: A significant correlation between poorer tracking (e.g. measured by standard deviation of lateral position, SDLP) (Philip, Sagaspe, Moore et al., 2005; Rossi, Gastaldi, & Gecchele, 2011; Åkerstedt, Peters, Anund, & Kecklund, 2005; Anund et al., 2008) as well as impaired reaction times (Saxby, Matthews, Warm, Hitchcock, & Neubauer, 2013; Johns, Chapman, Crowley, & Tucker, 2008; Philip, Sagaspe, Taillard et al., 2005) could be demonstrated when participants were fatigued.

Next to these driving simulator studies, fatigue is also a well-known causation for an impaired driving performance and accidents on-road in the real world. In their report *Drowsy Driving 2015* (National Center for Statistics and Analysis, 2017) a total of

396.000 crashes involving drowsy driving occurred between 2011 and 2015 in the USA. Of these accidents 2.4% were *fatal*.

In Germany from 1975 - 2017, each year about 2000 road accidents occurred due to drowsy driving. In the year 2017, 3064 persons were injured in such road accidents. 46 persons died and 923 were severely injured (Statistisches Bundesamt, 2018).

2.4.3. Fatigue In Automated Driving – What Has Been Investigated So Far?

In CDA, the execution of the driving task, the monitoring of the driving environment as well as the monitoring of the automated driving system are no longer in the responsibility of the driver when the system is activated (SAE, 2018). This can lead to a monotonous and rather passive situation for the human driver, whereby the development of *passive task-related fatigue* may at least be increased (Neubauer, Matthews, Langheim et al., 2012). In May & Baldwin's fatigue model from 2009, increasing automation (as the model is originally referring to manual driving) is mentioned as a possible causation for task-related fatigue. In CDA, fatigue can become a serious problem because the human driver still represents the fallback performance and has to be able to react properly in the case of a system-initiated RtI.

In recent studies it has already been shown that fatigue can indeed become a problem when it comes to automated driving: in an experiment conducted by Feldhütter, Hecht, and Bengler (2018) n = 3 participants have dozed off when they drove automated for about 60 min and did not have to engage in a NDRT. Also Omae, Fujioka, Hashimoto, and Shimizu (2006), reported that some participants fell asleep when they drove conditional automated. In this experiment, eight participants from a total sample of N = 30 fell asleep when they were told to monitor the system for automated driving. The duration of the ride was preset to 60 min.

Also Jamson, Merat, Carsten, and Lai (2013) and Merat, Jamson, Lai, and Carsten (2012) have already conducted experiments on fatigue and the development of fatigue in CDA. They also reported that fatigue develops rapidly when driving automated and that this goes along with negative consequences when taking over control of the vehicles.

However, similar effects have been investigated much earlier. The following is a brief overview of the research that has been done on this topic so far:

First attempts for assessing effects of fatigue in combination with automated driving systems and situations in which the human driver had to regain control of the vehicle have been made in the 1990s. However, these first attempts were not aimed at CDA, but rather dealt with partial automated driving systems that had to be monitored for the entire ride.

Hahn (1993) was one of the first researchers who investigated effects of fatigue on take-over performance in monotonous environments in combination with automated driving functions. In a driving simulator experiment, he investigated take-over behavior of 32 participants. The main focus of this experiment was to investigate how long the drivers need for regaining control of the vehicle and how a human driver can best be informed that he has to take over control of the automated vehicle. Results suggested, that all participants took over control in under 2.5 s. However, there seemed to be an influence of the driving experience: younger and less experienced drivers needed more lateral space for their take-over reaction (Hahn, 1993).

Also Desmond, Hancock and Monette (1998) investigated effects of a partial driving automation system. In their experiment, they compared effects of partial automation and manual driving on driving motivation and fatigue. The participants ($N = 34$) had to perform both, an automated ride and a manually ride. During both 40 min rides, three incidents at defined points of time (beginning of the ride, in the middle of the ride and at the end of the ride) occurred, which forced the human drivers to intervene. In the manual driving scenario, wind gusts have led to a drift of the vehicle, whereas in the automated ride, the same drift was explained due to automation failures. The results suggested, that the driving performance in these incident situations was better when the participants had to drive manually for the entire ride compared to the automated ride. In this experiment the researchers reported that the development of fatigue and the task load in the automated ride was comparable to the manual driving ride. From these findings, the authors explained that driving strategies, in which the driver is still involved in the driving task, are superior to higher automation levels.

In the following years, the number of experiments on automation effects and fatigue in automated driving increased substantially. This also led to an increase in experiments focusing on further automation levels.

In an experiment conducted by Saxby, Matthews, Hitchcock, and Warm (2007) automated driving was used to induce passive fatigue (*passive conditio*n). Next to the automated driving condition, there was the *active condition*, in which participants had to drive manually and react upon wind gusts during the entire ride, and the *control condition* (manual driving without wind gusts). As a result of this study it can be noted that fatigue tended to be highest in the *passive condition*.

In a following experiment, Saxby et al. (2008) used similar manipulations as in the experiment mentioned above (Saxby et al., 2007), which could be used to induce active and passive fatigue states. In this experiment it could be confirmed, that passive fatigue of the drivers increased when driving automated. Another finding in this experiment was, that participants that were in the *passive condition* had slower reaction times upon an unexpected event. Furthermore, these drivers were more likely to crash than participants from the other conditions (*active* and *control*).

Neubauer, Matthews, Langheim, & Saxby (2012) investigated how driver-initiated 5 min periods of automated driving affected fatigue, stress and workload compared to a manual driving group in 35 min rides. According to the description of the system, it was most likely a *partial driving automation* system, which had to be permanently monitored. In the end of the ride, when the drivers had to drive manually in both groups, an emergency situation occurred to which the drivers had to react. Results of this experiment suggest, that the optional periods of automated driving did not positively affect subjective fatigue or stress states of the drivers. The participants that were already fatigued before the experiment were more likely to activate the full automation periods. Reaction times upon an emergency event were slower, when participants were in the automation periods condition. However, only 44 of 93 participants in the automation periods condition switched automation on for at least one time.

In a further experiment Neubauer, Matthews, and Saxby (2014) investigated how an engagement with media devices (*no media device*, *cell phone call* or a *quiz game*) during different levels of automated driving (*no automation*, *partial automation* and *total automation*) affects fatigue. The media devices had a positive effect on minimizing the loss of task engagement and the increased load caused by vehicle automation. However, media usage has not been associated with faster responses to events, indicating that such media devices cannot permanently increase drivers' attention.

Jamson, Merat, Carsten, Oliver, & Lai (2013) also investigated if drivers in automated vehicles are likely to engage in secondary in-vehicle tasks. The researchers' assumption was that such tasks can be a pleasant alternation from the monotony of system monitoring. Thus, in their experiment, participants were free to engage in entertaining tasks, when the automated driving function was activated. The experimental design involved two different levels of automation (*manual control* and *highly automated*) and two levels of traffic density (*low* and *high*) in a repeated-measures design. The following results were reported: participants that could switch on automated driving did further engage in the entertaining tasks compared to the manual driving group. In situations with higher traffic density, however, the drivers did less engage in the NDRTs but payed more attention towards the traffic-environment. The researchers also evaluated the fatigue states of the drivers by measuring the eye-lid closure behavior. Results suggest higher fatigue states in the *automated driving condition* compared to *manual driving*.

Another researcher that used eye-based metrics to investigate if and how fatigue emerges in automated driving was Körber, Cingel, Zimmermann, and Bengler (2015). In this experiment, participants had to observe a partial automated driving system for 42.5 min and simultaneously engage in a reaction time task. The results suggest, that objective fatigue measured with eye-tracking methods increases with the duration of the automated ride. However, the measured reaction times upon the RtI were not negatively affected through the automated driving.

Also the influence of different tasks during CDA on the drivers' state has already been investigated. In a driving simulator experiment conducted by Schömig, Hargutt, Neukum, Petermann-Stock, and Othersen (2015), drivers were randomly assigned to one of three groups. Depending on an eye-lid closure based metric, automated driving was available for 15 min from a certain fatigue level. Based on their group, participants either had to continue *manually driving*, to *drive automated without task engagement* or to *drive automated and engage in a quiz task*. After the 15 min interval of automated driving (or manually driving in the manually driving group), participants who had to engage in the *Quiz-task* had the lowest fatigue levels measured with the objective eye-lid based metric compared to the two other groups.

In an experiment conducted by Miller et al. (2015), effects of different NDRTs on drivers fatigue in CDA were investigated. Participants had to *monitor the automated driving system*, *read*, or *watch a video* in random order during a 40 min automated ride. A video analysis was conducted to measure signals for emerging fatigue for all groups. Results support, that a task engagement (*reading* and *watching a video*) can counteract emerging fatigue compared to a *system monitoring*. In the monitoring task group, longer periods of eye-lid closure as well as yawing was detected.

Also Feldhütter, Gold, Schneider, and Bengler (2017) investigated how task engagement affects the driver state in CDA. Therefore, effects of the *surrogate reference task* (SuRT), a visually and manually distracting task, were compared to a *no task condition*. As second variable, the duration of automated driving was manipulated (*five min* vs. *20 min*). Next to changes in gaze behavior during the automated driving period, the authors also investigated the take-over performance. Results suggest, that gaze behavior was significantly affected after 20 min of automated driving. Also reaction times were significantly impaired after the *longer automated ride* compared to the *5 min* ride. Next to the duration of the ride, also the task engagement negatively affected reaction times. However, neither the duration nor the task negatively affected measures concerning the quality of the drivers' input.

As can be seen from section 2.4.1., next to *task-related fatigue*, also the other form of fatigue, the so-called *sleep-related fatigue* is a known causation for accidents in manual driving. This form of fatigue and its effects on take-over performance in CDA have also already been investigated.

Vogelpohl, Vollrath, Kühn, Hummel, and Gehlert (2016) investigated the influence of *sleep-related fatigue* (May & Baldwin, 2009) on take-over performance in CDA. Therefore, participants were randomly assigned to either an *alert group* (day time ride, no sleep deprivation) or a *fatigued group* (night ride, sleep deprived). Out of these two groups, each 15 participants either had to drive manually or automated. The simulated rides lasted up to 60 min and external experts evaluated the fatigue state of the drivers. Depending on the level of fatigue, a RtI was triggered by the examiners. The results support other findings: fatigue can increase in automated driving, especially when the drivers are sleep deprived. However, in this experiment all participants that experienced the CDA function were able to take over control when requested without severe consequences.

One of the more recent research results related to the present thesis is the series of experiments conducted by Schmidt (2018). The researcher analyzed the reaction ability for take-over situations in prolonged and monotonous CDA rides, and how fatigue in CDA emerges compared to manual driving. For this purpose three driving simulator experiments were conducted. In the first experiment, N = 14 participants had to drive manually for about 2:30 hrs, whereas in the second (N = 41) and third (N = 20) experiment participants had to drive conditionally automated for about 3:00 hrs. During the CDA rides, the drivers had to respond to reaction requests all 30 or 180 sec. These punctual interrogations were intended to make the drivers aware of the fact that they were still responsible for the overall driving task, despite the CDA system. Next to these interrogations, the drivers were confronted with several situations in which they had to regain control of the vehicle. The analysis of the driving parameters in these situations indicated that the drivers reacted safely and adaptively to all RtIs even when they faced higher levels of fatigue in those challenging take-over situations. However, it must be kept in mind that the test persons had to prove their reactivity in relatively short sections during the automated ride by reacting to a signal within 5 s. In addition, the ride was interrupted by several take-over situations whereby a monotonous situation gets prevented. This may have affected the fatigue state of the drivers positively. Next to these findings, the author stated that emerging fatigue in manual driving is comparable to the emerging fatigue in CDA.

Therefore, the assumption of Schmidt (2018), that automation can prevent drivers from emerging fatigue, can therefore be rejected. In fact, as can be seen in the experiments that are reported in this thesis, it can be said, that automation cannot only not prevent drivers from getting fatigued, but that automation leads to fatigue even more quickly than it is the case in manual driving.

The topic of emerging fatigue in CDA was in the focus of Weinbeer, too. She mainly focused on the question of how an automated driving system should deal with a fatigued driver, and has therefore developed the so-called *Drowsiness-Management-Concept* (Weinbeer, Bill, Baur, & Bengler, 2018). In a further experiment different NDRTs and their potential for managing drowsiness were investigated in an on-road experiment (Weinbeer, Muhr, & Bengler, 2019).

2.5. Overview Of Measurements And Assessments Of Fatigue

As explained in the previous sections, driver fatigue is a major road hazard and a large number of accidents are caused by fatigued drivers or even drivers falling asleep (see section 2.4.2.). Also when driving automated, as long as the driver represents the fallback and thus has to be able to react appropriately in emergency cases, fatigue could be identified as a possible causation for road accidents (see section 2.4.3.). To detect fatigue at an early stage, researchers have been working for some time to detect drivers' fatigue as it emerges.

According to Čolić, Marques, and Furht (2014) methods for the assessment or detection of fatigue can be grouped into different categories: *subjective*, *physiological*, *vehicle-based*, and *behavioral*. In the following section, these individual measurement methods are described in more detail.

2.5.1. Subjective Methods

The assumption for subjective fatigue detection is that the more fatigued a person is, the higher is the person's need for sleep, which in turn can be classified in different levels. The method of subjective fatigue assessment is based on various questionnaires with which the test persons can describe their level of fatigue. Other methods use a subjective evaluation of experts that rate the fatigue of another person by observation. The following methods are examples of commonly used measurements in the context of subjective fatigue research:

- Maintenance of Wakefulness Test (MWT): The MWT is a test where participants get instructed to stay awake for at least 20 min without falling asleep. Hereby, the participants are being monitored by an experimenter. If a participant cannot stay awake for at least 15 min, it is considered the participant is to sleep deprived to drive. This test is mainly used for patients with somnolence (Mitler, Gujavarty, & Browman, 1982).

- Epworth Sleepiness Scale (ESS): The ESS is an eight-item self-report questionnaire, in which participants have to indicate their subjective tendency to fall asleep in less stressful environments on a scale from 0 (*no chance to fall asleep*) – 3 (*high chance to fall asleep*). Thus, subjects can achieve a total score from 0 – 24, whereby a state of < 10 is considered to be *awake* and a score of > 15 is considered to be *sleepy* (Johns, 1991).

- Stanford Sleepiness Scale (SSS): The SSS is a single-item questionnaire in which participants have to state their current level of sleepiness on a seven-point scale. The scale ranges from 1 (*feeling active, vital, alert, and wide awake*) - 7 (*almost in reverie and cannot stay awake. Sleep onset is imminent*). It could be demonstrated, that the SSS is sensitive to sleep loss and to performance metrics in a large number of experiments (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973).

- Karolinska Sleepiness Scale (KSS): The KSS is the most commonly used questionnaire for subjective fatigue assessment. It is just the single question *How tired are you right now?* which has to be answered by a nine-point scale. The scale ranges from 1 (*extremely alert*) – 9 (*very sleepy, great effort to keep awake, fighting sleep*) and presents verbal anchors for each step. As laboratory tests and field studies verified, the KSS is sensitive to sleep deprivation and sleep loss (Åkerstedt & Gillberg, 1990).

2.5.2. Physiological Methods

As physiological signals are sensitive to earlier stages of sleepiness, physiological methods can be used to detect sleepiness prematurely. The physiological methods are highly reliable and their accuracy may outperform other methods. However, due to the high effort involved in their application, these methods can only be applied to a limited extent, especially in driving simulator experiments with a large number of participants. Nevertheless, two of the most common ones are explained in more detail:

- Electrocardiogram (ECG): The ECG is used to record electrical activity of the heart of a human being. This method can be used to measure both heart rate (HR) as well as the heart rate variability (HRV). The HR represents the number of heart beats of a human being within a 1 min interval. The HR is depending on the mental and physical fitness level of the person and can vary between 50 – 100 beats per min (bpm). Bishop, Madnick, Walter, and Sussman (1985) detected, that the HR decreased when the subjects got tired during the driving task. In further experiments however, these results were mainly ambiguous. More sensitive for detecting sleepiness is the HRV. The HRV measures the time-interval between two heart beats. In this metric, fatigue goes along with an increasing variability of the time-interval between these heartbeats (Manzey, 1998). However, these ECG values should be treated with caution as they are subject to inter- and intraindividual differences.

- Electroencephalogram (EEG): The EEG is the most reliable signal for precisely detecting sleepiness of human beings. This measurement uses different frequency bands that are derived via electrodes from the human brain: the *delta band* corresponds to sleep activity, the *theta band* corresponds to fatigue, and the *beta band* corresponds to alertness (Akin, Kurt, Sezgin, & Bayram, 2008; Lin et al., 2010). However, this method is very complex and prone to errors. Additionally electrodes have to be attached to the head of the subjects and a liquid must be applied to ensure sufficient contact between the scalp and electrodes. Due to these unfavorable conditions and the lack of a possible applications in a production vehicle, this method was not considered for our tests.

2.5.3. Vehicle Based Methods

Fatigued drivers are likely to show specific driving patterns. Vehicle based methods are state of the art right now, as they can detect such specific driving behaviors that may indicate a fatigued driver. These methods are based on the principle that fatigued drivers tend to detect a deviation from the lane-center later than active and awake drivers. Thus, active drivers tend to steer more often compared to fatigued drivers, whereby active drivers tend to steer less pronounced than fatigued drivers. Two methods are further explained:

- Measurement of Steering-wheel Movement (SWM): The SWM detects micro-steering movements of the driver. A fatigued driver tends to steer less than attentive drivers. Thus, the number of micro-steering inputs is smaller compared to active drivers (Feng, Zhang, & Cheng, 2009).

- Measurement of Standard Deviation of Lateral Position (SDLP): An on-board camera system detects the lateral position of the vehicle in the lane. When the car leaves the designated lane, this can be detected by the sensors and the driver can be alerted (Ingre, Åkerstedt, Peters, Anund, & Kecklund, 2006).

2.5.4. Behavioral Methods

Behavioral methods seem adequate for measuring of fatigue in real-driving scenarios as they are non-invasive and the human driver is unaware of the measurement. Typical systems are mostly camera or sensor based methods. Some are explained in more detail in the following.

- Measurement of the head position: When getting fatigued, drivers tend to nod their head. This nodding can be detected by using 3D-video measurements (Murphy-Chutorian & Trivedi, 2010; Zhang, Zheng, Mu, & He, 2009).
- Measurement of yawning: Frequent yawning is a symptom of fatigue that is known to everybody. Thus, the detection of yawning can also be used as a fatigue detector. However, yawning does not only occur before falling asleep. Therefore, a redundant indicator or fatigue detector is appropriate (Saradadevi & Bajaj, 2008).

Eye-lid based measurements: Everyone knows it: when getting tired, it is hard to keep the eyes open. This mechanism, the eye-lid closure was soon identified as a reliable predictor for emerging fatigue (Erwin, 1976). First research attempts on measuring emerging fatigue by using eye-lid closure based data was conducted by Skipper, Wierwille, and Hardee (1984) and by Hardee, Dingus, and Wierwille (1986). Soon afterwards, the metric PERCLOS, which means percentage of eye-lid closure over the pupil over time, was mentioned for the first time in a publication (Dingus, Hardee, & Wierwille, 1987). In this publication PERCLOS is defined as a state in which the eye is covered by at least 80 % by the eye-lid. In addition, an extensive validation by Wierwille, Wreggit, Kirn, Ellsworth, and Fairbanks (1994) showed that fatigue can be measured by PERCLOS very accurately. In addition, the method is relatively simple to use. Due to the fact that fatigue is measured accurately and that the method is easy to use, this method was used in the experiments conducted as part of this thesis to measure fatigue objectively. In section 3.3.2. further details of the measurement method are described. Other examples for eye-lid based metrics for fatigue detection are *blink frequency* or *mean blink duration* (Stem, Boyer, & Schroeder, 1994; Körber, Cingel, Zimmermann, & Bengler, 2015).

The different fatigue detection methods presented in this section have both advantages and disadvantages, especially for the assessment in the driving simulator. These are summarized in Table 3.

Table 3: Advantages and Disadvantages of the different fatigue detection methods.

Method	+	-
Subjective	- Easy to assess - Low effort	- Accuracy is questionable - Can be influenced by the respondent (social desirability)
Physiological	- Accurate fatigue detection	- Effortful - Susceptible to disturbances such as movements
Vehicle based	- Imperceptible to the driver	- Not suitable for automated vehicles (driver does not drive)
Behavioral	- Imperceptible to the driver - Accurate fatigue detection	- Elaborate post-processing - Effortful

2.6. Test Environments For Assessing Of Human Performance In Conditional Driving Automation

In the experiments described in this thesis, human drivers had to engage in NDRTs to reach higher levels of fatigue in order to investigate whether fatigued drivers are capable to solve *short-term* take-over situations in CDA. As such *short-term* take-over situations always involve a potential hazard for the participants, and due to the current legislation which not permits vehicles with CDA functionality on public roads in Germany, the experiments were carried out in simulated test environments. In the following section of this thesis, the general advantages of such simulated conditions are discussed. For the technical specifications and information about the implementation of these methods see section 3.1.1. for the driving simulator and 3.1.2. for the on-road *Wizard-of-Oz* approach.

2.6.1. Driving Simulator Experiments

Three of the experiments (Experiment 1, 2 & 4) carried out in the context of this thesis included *short-term* take-over situations with rather low time-budget to investigate how *task-related fatigue* due to monotonous monitoring tasks affects take-over performance in CDA. Due to the following advantages, the experiments that involved such take-over situations were conducted in a motion base driving simulator (Thoma, 2010):

- *Reproducibility*: The RtIs must be experienced in the context of a take-over situation in order to test the human reaction. For assessing comparable results upon the RtI, it must be able to repeat these situations in the same form for all participants, including the same weather conditions, a similar traffic situation and a similar behavior of the other road users. This can easily be done in a driving simulator.

- *Endangerment for participants*: Only in the driving simulator participants can be brought into situations in which an accident is not always unavoidable for some of the participants. Due to the fatiguing NDRT and the *short-term* take-over situation, the participants had to solve, the situations investigated in the experiments have to be classified as potential safety hazards.

- *Technical feasibility*: Automated driving functions can only be safely implemented with perfect environmental knowledge which requires a large number of sensors and artificial intelligence. So far, this can only be implemented fully functional in a simulated environment. Another advantage is, that predefined parameters for the assessment of the take-over reaction can be recorded with less effort compared to experiments on the road (e.g. reaction times and quality measurements).

However, such driving simulators also come along with some disadvantages. Above all, especially one thing is serious: the participants know that they are in a simulated environment in which nothing can happen to them. This can quickly lead to the situation that the participants do not behave as they would do in a real vehicle on-road equipped with CDA functions. This makes it very difficult to distinguish between effects caused by the simulator and a behavior that could also be observed in reality. Due to the absence of realistic danger and resulting consequences, a false sense of safety and responsibilities may occur (Käppler, 1993).

In addition, other aspects can contribute to the fact that the situation for the human driver in the simulator can be more monotonous than it would be in a real driving environment.

The lighting conditions in the driving simulator differ significantly from those in a real driving environment. Furthermore, the simulated environment in the driving simulator does not reproduce conditions as they exist in reality (traffic density, weather, etc.). This can contribute to the fact that the measured effects are more due to the environment of the driving simulator and less to the automated driving in connection with the NDRTs.

Therefore, it is necessary to verify the findings found in the simulator in reality. Since this is not legally possible with a real CDA vehicle, a *Wizard-of-Oz* vehicle (see section 3.1.2.) was used. This vehicle was used in one experiment (Experiment 3).

2.6.2. Wizard-Of-Oz Experiments

The method *Wizard-of-Oz* is named after the children's book *The wonderful Wizard of Oz* by the US author Lyman Frank Baum (1900).

The *Wizard-of-Oz* method is a commonly used method in different areas of research and science. It is particularly often used in psychology, ergonomics and human factors. The basic principle of this method is always the same: a system behavior that is not yet available due to technical difficulties or legal requirements is simulated by a hidden experimenter, the so called *Wizard*. This *Wizard* simulates the artificial behavior of an automated system. Of special importance is that the existence of the *Wizard* is not known to the participants, so that they think that the technical function runs functionally and / or automatically. Therefore, the *Wizard* is hidden for the participant and controls the system from another room or at least a separated area.

Aim of this method is that the participants assume to communicate with, or to control an autonomous system, but in reality the automated system is only simulated by another person (i.e. the *Wizard*). For the testing of new technical solutions which are conceivable in the near future, but for which there are no technical solutions yet, or which are still prohibited by law, this method is a good way to explore future user behavior with such systems.

In the past, the *Wizard-of-Oz* method was first used to simulate a speech-based interaction with a computer. The test persons were told to communicate via speech with a computer which would process their language and react to it. In fact a hidden *Wizard* reacted to the speech input of the test persons and answered by entering a speech output (Salber & Coutaz, 1993; Klemmer et al., 2000).

Soon, the method was also applied in the context of vehicles: for testing of novel operating concepts in vehicles, a *Wizard-of-Oz* approach was used. Among others Alpern & Minardo (2003), have investigated multimodal interaction such as gesture control. Manstetten, Krautter, Grothkopp, Steffens, and Geutner (2001) used a *Wizard-of-Oz* approach to investigate to which extent voice control can be applied in vehicles. Both technologies were not yet fully developed at that time.

For some years now, the *Wizard-of-Oz* method has also been used to investigate innovative ADAS, which are not yet existent, or whose prototypical implementation would be too costly for the purpose of the experiment. For example Schieben, Heesen, Schindler, Kelsch, and Flemisch (2009) and Flemisch et al. (2010) used a *Wizard-of-Oz* approach to simulate automated driving functions in a driving simulator. For this purpose, a hidden *Wizard* driver took over the vehicle control in the moment the participant activated the automated driving system. The test participants thought that the simulated vehicle had been driven automatically.

In on-road environment in the real world a *Wizard-of-Oz* method was used for the first time in 2010. In order to investigate interactions of human drivers with various levels of automated driving systems and / or ADAS on the German Autobahn, Petermann and Schlag (2010) built up a vehicle with a second driver's work place with redundant steering-wheel, pedals and all other required instruments for driving in the rear of the vehicle. The second driver seat in the rear of the vehicle was separated from the front seats by a partition wall as well as by a darkened glass.

A similar vehicle was used in the Experiment 3 as part of this thesis, to verify the findings from the simulator experiments. For further details about the technical specifications and more detailed information about the vehicle that was used see section 3.1.2.

2.7. Aspects Of Human-Machine-Interfaces In Conditional Driving Automation

Due to the changed role of the human driver in CDA, there may also be changes in the HMI design and the control elements with which the driver controls the vehicle. In previous experiments, for example, the classic instruments including the steering-wheel and the pedals were replaced by a side stick including grip force measurement (Damböck, Kienle, Bengler, & Bubb, 2011). Also Kienle (2015) has investigated alternative operating elements for automated driving vehicles. He investigated a two-axis yoke operating element similar to those used in an aircraft. Transverse guiding was similar to a standard steering-wheel but longitudinal guiding was also controlled with the element by pulling or pushing it.

In another approach Kerschbaum, Lorenz, and Bengler (2014) could demonstrate that a decoupled steering-wheel can be helpful in take-over situations and additionally, that a steering-wheel without spokes can help to improve take-over performance. In a further attempt, the same experimenters investigated if a shape-changing steering-wheel can help to improve the participants' awareness of the driving mode and experienced comfort when driving with CDA (Kerschbaum, Lorenz, & Bengler, 2015). Take-over performance could be improved with this concept.

However, not only the control elements can be adjusted in CDA, but also the design of the HMI concept displayed in the instrument cluster or the head-up display can help to improve the experience for the user (van den Beukel & van der Voort, 2011). A good HMI concept in an automated vehicle should be designed to support the human driver in many different ways. On the one hand, the current system status must be immediately visible for the driver and must not cause any false conclusions, as mode awareness (e.g. mode of the automated driving function) was identified as a potential hazard in CDA in experiments before (Petermann & Schlag, 2010; Petermann & Kiss, 2010).

Next to the aspect of mode awareness for the user of such automated driving system, the RtI and the design of the RtI play a crucial role when considering the HMI concept. The purpose of a RtI is to quickly signal the driver that he must regain control of the vehicle. Therefore, in the most experiments on take-over performance in CDA, a redundant concept which warned the driver both acoustically and visually was chosen. However, there were also some attempts in which the driver just got a pure acoustically RtI (Merat & Jamson, 2009; van den Beukel & van der Voort, 2013) or a pure visual RtI (Naujoks et al., 2014). Based on the results from this experiment, a purely visual warning concept is not recommended.

2.8. Measurement Of Take-Over Performance In Conditional Driving Automation

In CDA take-over situations, in which an intervention of the human driver is necessary, can occur. As an adequate reaction of the driver in such situations is not self-evident and furthermore dependent on many factors like the drivers' state (e.g. motoric state or fatigue state, see section 2.3. for more information) many experiments have been conducted so far, to assess human performance, when it comes to a take-over situation.

Especially in *short-term* take-over situations that are not known to the human driver or the automated driving system until detected by the on-board sensors, it is uncertain whether the human driver can react adequately. For that reason, such situations are in the focus of research right now and are examined in a variety of different studies. In order to evaluate how good the participants' responses in these experiments are and in order to interpret the results, it is necessary to identify metrics that allow such a statement. In previous experiments, two categories of metrics have been used frequently (Gold, 2016): *time based metric*s (see 2.8.1.) and *metrics that measure the quality of the responses of the drivers* (2.8.2.).

Since an individual consideration of the results only has a limited power of validity, it is advisable to always refer to aspects of both these categories for a take-over performance assessment (i.e. Gold, & Bengler, 2014; Lorenz, Hergeth, Kerschbaum, Gold, & Radlmayr, 2015).

Nevertheless, additionally a subjective video rating, the take-over controllability rating (TOC-rating) is introduced, which can also be used to measure take-over performance holistically (see section 2.8.3.).

2.8.1. Time Based Metrics For The Assessment Of Take-Over Performance In Conditional Driving Automation

As taking over control from an automated vehicle consists of a sequence of required actions the human driver has to fulfill (within the available time budget), the required time-intervals, the human driver needs to fulfill these actions, can be measured. In a *short-term* take-over situation a fast reaction indicates a better reaction compared to a slower reaction. An overview of the recorded reaction times that were used to assess human performance in the take-over situations in this thesis can be seen in section 3.4.1.

2.8.2. Quality Metrics For The Assessment Of Take-Over Performance In Conditional Driving Automation

Quality is defined as the degree to which a set of inherent characteristics of an object meets predefined requirements (DIN Deutsches Institut für Normung e.V., 2015). Thus, a variety of metrics (i.e. each required action of the human driver) can be used to assess the quality of the driver intervention in the take-over process. Most of them concentrate on accelerations or on the quality of the lateral control of the human driver after he regained control of the vehicle. An overview of the recorded quality measurements that were used to assess the performance of the human driver in the different take-over situations as part of this thesis is displayed in section 3.4.2.

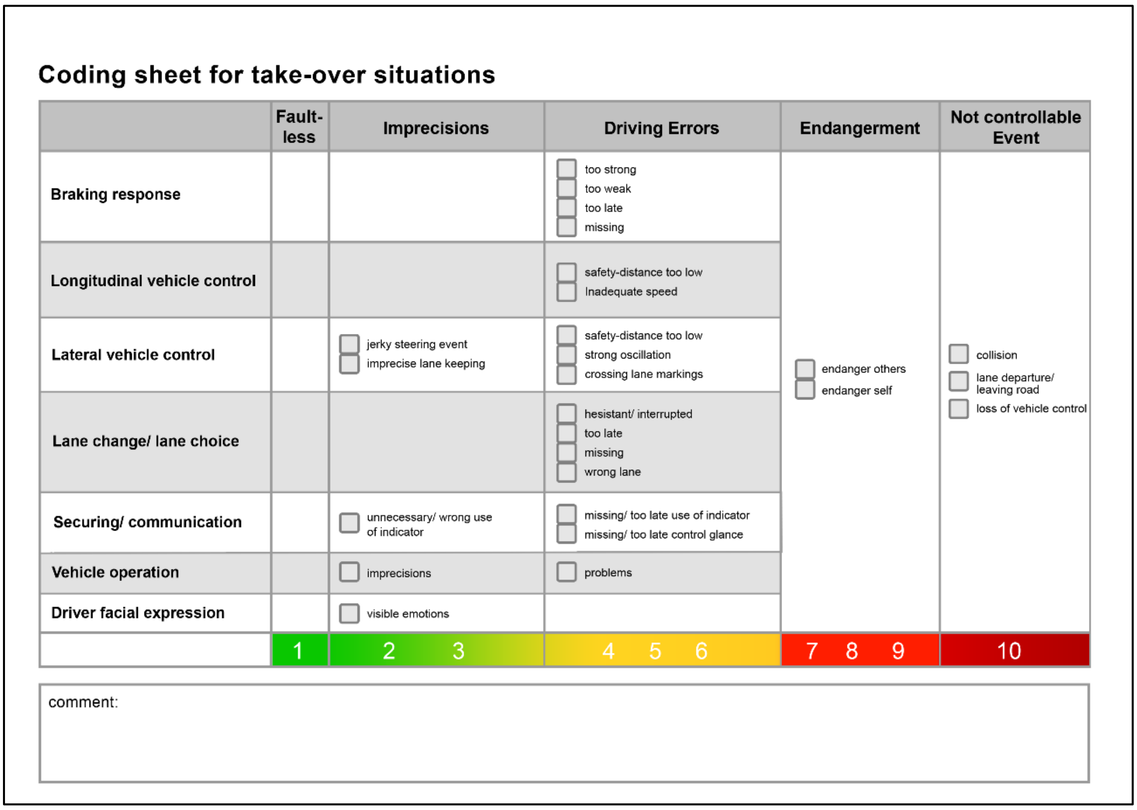
2.8.3. The TOC-Rating – A New Approach For The Assessment Of Take-Over Performance In Conditional Driving Automation

This chapter is based on a previous publication:

Naujoks, Wiedemann, Schömig, Jarosch, & Gold (2018).

The above mentioned metrics (*time based metrics* and *quality metrics*) were used in most of the experiments conducted so far for the assessment of take-over performance. However, even the simultaneous usage of both assessment categories does not imply a correct result either, as only those parameters can be reported which have been recorded and thus are visible in the data. For example, problems of the human driver with vehicle operation or dangerous driving behavior may not always be apparent from the available data set. For this purpose, a video-based expert rating, the TOC-rating, can potentially help to better understand human behavior in such take-over situations (Naujoks et al., 2018).

The TOC-rating was developed as a holistic assessment for control transitions from automated to manual driving. By a standardized rating scheme, at least three trained experts rate the take-over reactions of all participants of a driving simulator experiment. Therefore, the experts use video-data from different points of view and rate different aspects of the take-over performance of each participant. The main advantage of the TOC-rating scheme is that many different dimensions of the driving task are taken into account into one overall result.



Coding sheet for take-over situations

	Fault-less	Imprecisions	Driving Errors	Endangerment	Not controllable Event
Braking response			☐ too strong ☐ too weak ☐ too late ☐ missing		
Longitudinal vehicle control			☐ safety-distance too low ☐ Inadequate speed		
Lateral vehicle control		☐ jerky steering event ☐ imprecise lane keeping	☐ safety-distance too low ☐ strong oscillation ☐ crossing lane markings	☐ endanger others ☐ endanger self	☐ collision ☐ lane departure/ leaving road ☐ loss of vehicle control
Lane change/ lane choice			☐ hesistant/ interrupted ☐ too late ☐ missing ☐ wrong lane		
Securing/ communication		☐ unnecessary/ wrong use of indicator	☐ missing/ too late use of indicator ☐ missing/ too late control glance		
Vehicle operation		☐ imprecisions	☐ problems		
Driver facial expression		☐ visible emotions			
	1	2 3	4 5 6	7 8 9	10

comment:

Figure 4: TOC-Coding sheet used for video-based analyzing the take-over situations.

Process of the rating: in the coding scheme (see Figure 4) it is distinguished between different levels of hazards which are caused from particular actions of the driver: *faultless* (in vehicle handling) (1), *imprecisions* (2 - 3), *driving errors* (4 - 6), *endangerments* (7 - 9) and *not-controllable events* (10). The overall objective of this tool is for each expert to give an overall score for each take-over of each subject. In order to obtain this overall score for each participant, a hierarchical evaluation procedure is applied. The TOC-rating process is displayed in Figure 5.

- First of all, it must be checked whether an uncontrollable event has occurred (e.g. collision or loss of vehicle control). If this is the case, the take-over reaction must be rated with 10 - *not controllable event* by the raters.

- If the situation was not a *not controllable event*, the rater has to evaluate whether the situation has been safety-critical for the driver himself or for any other road user in the current situation. If the situation was safety critical, the rating results in a 7 – 9 - *endangerment*. Depending on the estimated risk which depends on the driving error, the rater has to assign a score.

- If the situation is not an *endangerment* for the driver or for other road users, the rater has to decide about the quality of the take-over. If the driver did not react correctly and *driving errors* occurred, the rater hast to decide (e.g. by frequency and severity of the driving errors) between a score of 4 – 6 - *driving errors*.

- If there were no *driving errors* at all, the rater has to evaluate if there were *imprecisions* during the take-over performance. If there are *imprecisions*, the overall rating of the expert would lead to a 2 – 3 - *imprecisions*, depending on severity and frequency of the imprecisions.
- Only if the take-over performance was perfect a rating of 1 is justifiable.

For the rating of the take-over performance, different dimensions of the driving task are taken into account. For an overall assessment of the take-over performance, it is necessary that all these dimensions are assessed:

- *Braking response*
- *Longitudinal vehicle control*
- *Lateral vehicle control*
- *Lane change / lane choice*
- *Securing / communication*
- *Vehicle operation*
- *Driver facial expression*

Rating material: For the TOC-rating, video-material of take-overs from automated driving to manual driving is required. Hereby it does not matter whether the video-material from already conducted experiments is used or whether a new experiment is carried out with the purpose to use the TOC-rating. For a strong evaluation of the take-over performances, the videos should include different perspectives of the scenario. This necessarily includes the following perspectives:

- *View of the driving scene:* ideally both from a bird's eye perspective and from the point of view of the driver.
- *View on the HMI elements:* to be able to determine when the RtI was triggered.
- *View on control elements:* to estimate the intervention and vehicle operation (i.e. intensity of steering) of the driver.
- *View on the driver:* to evaluate if the driver engaged in NDRTs and to see the facial expression of the driver.

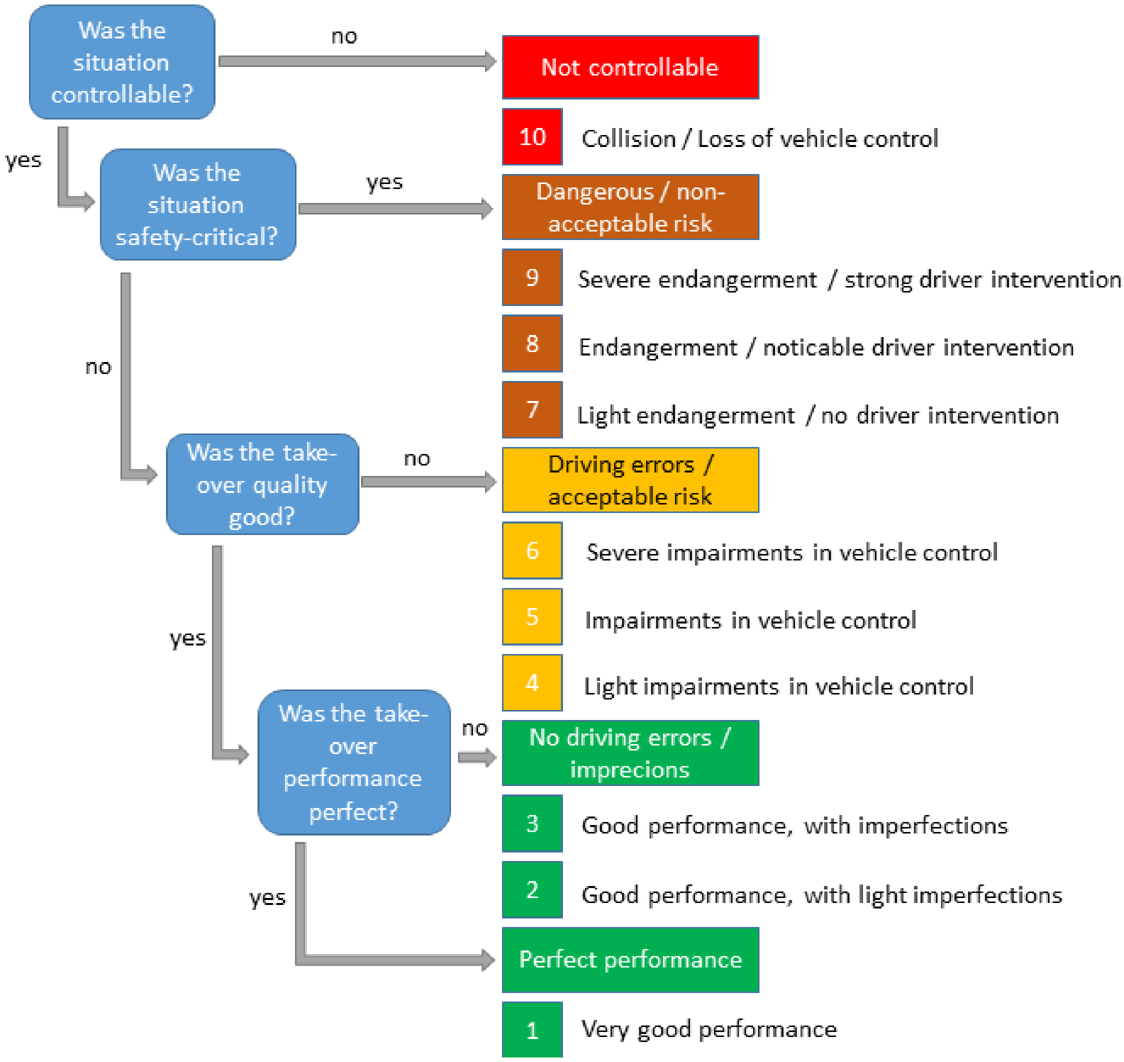


Figure 5: TOC-rating process.

In all experiments that contained take-over situations the take-over performances were, next to time based metrics (i.e. reaction times) and aspects concerning the take-over quality, evaluated by using the TOC-rating tool. Therefore, three trained raters have viewed all take-over reaction videos and rated the take-over performances. The results are each reported in the results section of the different experiments (see sections 4. – 7.).

An overview of the metrics used for the measurement of take-over performance in the different experiments that are reported in this thesis can be found in section 3.4.

2.9. Problem Statement

When CDA will be available for the consumer market within the next years, it will represent the first level of automated driving which allows the human driver to completely turn away from the driving task and to engage in NDRTs. However, the human driver still needs to be in an appropriate state for driving for the entire ride as the human driver represents the *fallback level* and thus has to be able to regain control if requested by the CDA system. In some situations, the so-called *short-term take-over situations* the time budget for such a control transition process will be as short as 5 – 10 s (Petermann-Stock et al., 2013).

As explained in section 2.3.2., the driver state consists of many different aspects. The task-switch for the human driver in CDA, from the active driver to a passenger who has to intervene if being requested by the system, also changes some conditions for the driver with direct relevance on the driver state. One aspect in particular will change as a result of CDA: in manual driving, the human driver is responsible for the longitudinal and lateral guidance as well as for constant securing behavior. With the introduction of CDA, however, the driver has no specified task until he is requested to take over the vehicle control. Furthermore, he is supposed to relax or to engage in NDRTs.

Therefore, it is expected that a monotonous situation for the human driver can occur in CDA, whereas conditions for *passive task-related fatigue* may emerge. Previous experiments could already demonstrate, that some tasks or activities can support the driver in such monotonous situations. Positive effects of different NDRTs counteracting fatigue states of the drivers have been proven. However, it has to be expected, that not all NDRTs or activities have the same impact on the drivers' fatigue state. Especially, when it is not a free choice task, but rather a task that has to be done (i.e. supervise the automated driving system like it is the case in SAE level 2, *Partial Driving Automation*). This permanent task engagement can lead to fatigue, especially to one form of fatigue: *passive task-related fatigue* (May & Baldwin, 2009).

Apart from the *passive task-related fatigue*, which can arise during the automated ride due to increasing monotony, it also has to be assumed that human drivers which already are *sleep related fatigued* will use such automated driving functions.

Concluded it has to be stated, that fatigue can occur in CDA. For safety reasons it is of great importance, that also these fatigued drivers are capable of taking over the vehicle control when requested by the system.

Therefore, to better understand the interaction between NDRTs, the drivers' fatigue state and take-over performance, this dissertation was written as part of the German Funding

Project Ko-HAF. Based on four experiments conducted in a driving simulator or in a *Wizard-of-Oz* vehicle, the following research questions were addressed:

1. How do NDRTs affect the drivers' state during CDA?

2. How can emerging fatigue be detected in CDA?

3. How does fatigue affect take-over performance of the driver, when a short term take-over situation occurs?

The following research questions came up after the first experiments conducted in the driving simulator:

4. Does fatigue emerge in real traffic environment on-road as quickly as it is in the driving simulator when driving CDA?

5. How can the human driver be best supported in *short-term* take-over situations?

The four experiments are explained in detail in sections 4. - 7. In the following section, section 3., the overall method is explained in detail.

3. Overall Method

To address the central research questions mentioned in section 2.9., four experiments with a total of *N* = 235 participants have been conducted. Three of these experiments were conducted in the motion base driving simulator of the BMW Group in Munich, in the Research and Innovation Centre (German: Forschungs- und Innovationszentrum; FIZ) and one experiment was conducted in real driving environment on the German Autobahn A92 by using a *Wizard-of-Oz* approach. In section 3.1., these two test environments are described in more detail.

In all experiments, NDRTs were used to affect the drivers' state in CDA. These tasks were used either to provoke drivers' fatigue (*Pqpd-task*) or to prevent drivers from fatigue (*Quiz-task*, *Free-choice-activity*). Information on the individual NDRTs that were used in the experiments can be found in section 3.2.

As the detection of fatigue was a major objective of this work, different methods were used to measure emerging fatigue. Both subjective and objective methods were applied in this context. See section 3.3. for an overview of the fatigue-measurements used in the different experiments.

Another main objective of the experiments was to investigate how fatigue affects take-over performance in CDA. To assess take-over performance, objective measures (in detail *time based metrics* and *quality based metrics*) as well as a subjective measurement, the so-called TOC-rating, were used. See section 3.4. for more detail about these measurements.

Information regarding the HMI design that was used in the experiments is provided in section 3.5. This includes information about the presentation of the different system states (*automation available* / *automation active* / *RtI*) and information on how the system can be operated by the driver.

An overview of the individual experiments can be found in section 3.6. In this section, information about the number of participants, the experimental design, the research focus as well as the test-environment is provided.

Further details as well as the research questions and objectives of the respective studies are given in section 3.7.

3.1. Test Environments

Due to the current legislation which prohibits CDA on-road and for safety reasons (in the case the participant cannot react appropriately) the participants experienced the vehicles equipped with CDA functionality in simulated environments. Three experiments were conducted in the driving simulator and one experiment was conducted in the *Wizard-of-Oz* vehicle. In the following section, further details about these two methods are provided.

3.1.1. Driving Simulator Experiments

The main advantages of a driving simulator over a real on-road experiment in real traffic environment are the ability to make precise settings that create the same conditions for all test participants during the test and the high reproducibility of the scenarios to be evaluated. Another enormous advantage of driving simulators is the possibility of investigating situations that would not be possible in real road traffic due to safety reasons for the test persons.

Since three out of four of the experiments listed in this thesis involved take-over situations that required a driver intervention, these experiments were carried out in a driving simulator. All driving simulator experiments were conducted in a motion base driving simulator at BMW group laboratories in the FIZ in Munich (see Figure 6).

Technical specifications of the simulator: Nine visual channels are rendered at 60 frames / s at a resolution of 1920 x 1200 pixels. Seven forward channels provide a horizontal field of view of 240° x 45°. The two rear channels, providing the same resolution, project the rear view and thus can be seen through the side mirrors. The simulator uses a hexapod system with six degrees of freedom. Maximum translation is + / - 1.2 m and rotation + / - 30°. Maximum acceleration is 7 m / s².

In all three experiments the same Mock-Up, a *BMW 5 Series Touring* with automatic transmission, was used. Steering-wheel, pedals and all necessary instrumentation was identical to a production car. A servo-motor and a steering model was used for realistic steering torque simulation for manually driving or in the case of a driver intervention following an RtI.

Automated driving was implemented as follows: The CDA system could be activated by pressing a button on the steering-wheel. After activating the system, the car directly took over longitudinal and lateral control. In this moment the drivers had to take their hands off the steering-wheel and their feets from the pedals. The target speed of CDA was set to 130 km / h (80.8 mph). The system was set up to overtake slower moving vehicles automatically. In all driving simulator experiments, participants were told to switch on the CDA system after entering the highway and to drive automated for the entire ride.

The system could be deactived by a braking input or by pressing the button on the steering-wheel again. When deactivated, the vehicle decelerates with drag torque and drifts in the current direction of the steering-wheel input. In situations the system cannot handle, an RtI is indicated.

Figure 6: Motion base driving simulator of the BMW Group (Copyright © BMW AG).

3.1.2. Wizard-Of-Oz Approach

In order to verify that the results, especially the results related to the drivers' fatigue state, are not only attributable to the driving simulator, but can also occur in reality on-road, a *Wizard-of-Oz* approach was used additionally to the experiments in the driving simulator. This allowed the test persons to experience (simulated) CDA in real traffic environment. Therefore, a standard production type car (*BMW X5*) was equipped with a second driver's seat and corresponding instruments in the rear of the vehicle. Thus, the vehicle could be driven from a driver in the rear of the vehicle, who was covert for the actual participant. A privacy glass divided the front and the rear seat row of the vehicle. The modified vehicle that was used for the experiment is shown in Figure 8. In Figure 7, the layout of *Wizard-of-Oz* vehicle is displayed whereby the method should be clarified.

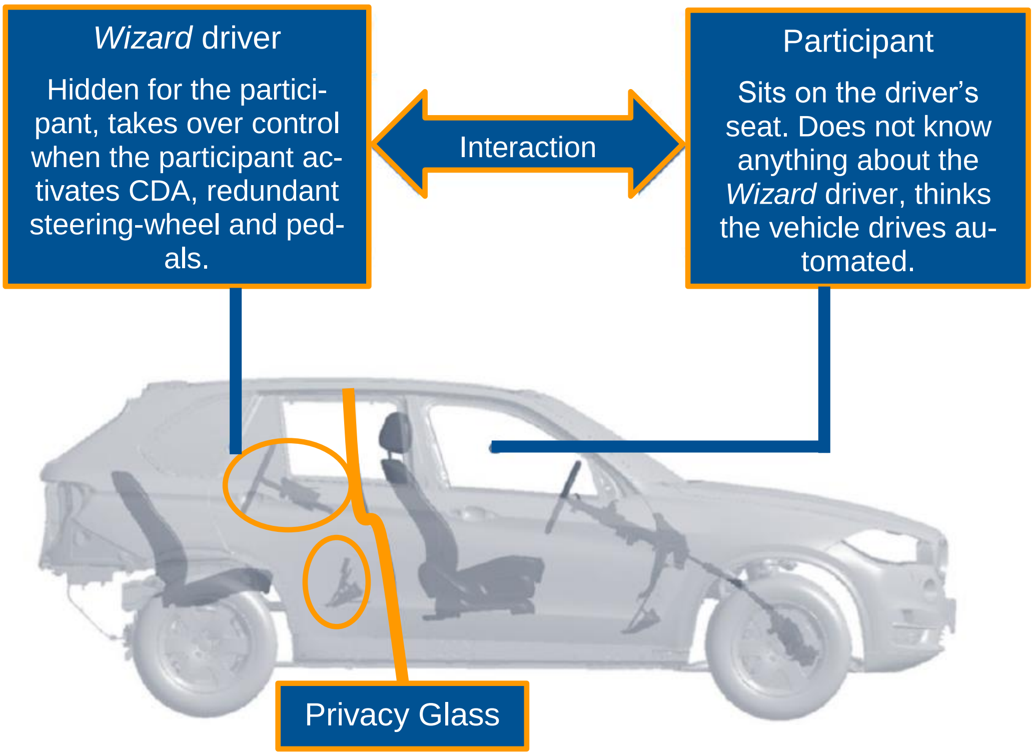


Figure 7: Schematic layout of Wizard-of-Oz vehicle.

Figure 8: Wizard-of-Oz vehicle side view with the two driver workplaces.

The second driver's seat (for the *Wizard* driver) was placed in the center of the rear and was equipped with a second steering-wheel, pedals and all other instruments that are necessary for safe driving. The *Wizard* was able to secure the driving environment via a camera-monitor system. The *Wizard* was already sitting in the rear of the vehicle when the participant was brought to the vehicle.

With this method, the participant who is sitting in the front of the vehicle on the driver's seat cannot recognize, nor hear or see the *Wizard* driver in the rear. From the very beginning of the experiments and with the aim to avoid unnecessary questions about the glass pane and the separated rear space the participants were told that the vehicle is equipped with automated driving functions and that the rear area is required for computing hardware and sensors. The participant could activate and deactivate vehicle automation (i.e. the *Wizard*) via two buttons on the steering-wheel. When the participant activated the system (i.e. the *Wizard*) the *Wizard* driver in the back decoupled the steering-wheel and the pedals from the real driving seat in the front and then controlled the vehicle from the rear. In specific situations (e.g. exit of the Autobahn) the vehicle control was returned to the participant in the front of the vehicle through *long-term* transitions (time-budget of about 60 s).

After the experimental rides, the test persons were informed about the *Wizard-of-Oz* method and the second driver in the rear. The method was used for Experiment 3.

3.2. Non-Driving-Related Tasks – A Way To Affect The Fatigue State Of The Driver?

One focus of all experiments conducted as part of this thesis was to investigate how different NDRTs affect the drivers' state during CDA. Therefore, in Experiment 1 (see section 4.) and Experiment 2 (see section 5.), two standardized tasks had to be processed by the participants during the entire experimental rides. One of the two tasks, the so-called *Pqpd-tas*k (see section 3.2.1.) was designed to lead to fatigue of the driver, whereas the other task, a classic *Quiz-task* (see section 3.2.2.) was designed to prevent the drivers from emerging fatigue. In Experiment 3 (see section 6.), which was conducted in the *Wizard-of-Oz* vehicle, participants again either had to deal with the *Pqpd-task* or could freely choose their activity (*Free-choice-activity* – see section 3.2.3.) while driving conditionally automated. In the last experiment, Experiment 4, all participants had to deal with the fatiguing *Pqpd-task*. The NDRTs used in the experiments to affect the drivers' state are further explained in the following sections.

In previous experiments, effects of different NDRTs on take-over performance were investigated. An overview as well as a classification of these NDRTs can be found in the publication by Naujoks, Befelein et al. (2018).

3.2.1. Pqpd-task

The *Pqpd-task* was used to cause *passive task-related fatigue* (May & Baldwin, 2009) to the participants while driving automated. The *Pqpd-task* was designed inspired by the *d2 test of attention* (Brickenkamp, 1962), which is a well-established test for the measurement of attention. However, since the *d2 test of attention* is a demanding task and therefore rather leads to *active task-related fatigue*, the test was transformed into a passive task. As this task should lead to *passive task-related fatigue* due to its monotony and the limited stimuli for the participants, a task with a low event rate and a fixed event location was designed (Warm, Parasuraman, & Matthews, 2008).

Another reason for using a task for fatiguing the participants instead of a *no-task engagement* was to create similar conditions for all participants. In other experiments, a *no task condition* was successfully used to fatigue the participants. However the course of the emerging fatigue was not comparable between all participants. As possible reason for this phenomenon, the so-called *mind-wandering* (Helton & Warm, 2008) was named (Körber et al., 2015). Therefore, in the experiments reported here, NDRTs were used to get comparable results from all participants that can be attributed to the emerging fatigue due to the NDRT-engagement.

In the *Pqpd-task* the four letters *P*, *q*, *p* and *d* were randomly presented for a variable amount of time between 10 – 15 s on a Windows-Surface tablet computer. The task for the participants was to react every time the small *p* was displayed on the screen by a touch input. The touch input was confirmed by a slight coloring of the background. On average, a *p* was presented once a min. The performance in this task was not further evaluated as the task was used to increase the monotonous situation for the driver during the automated ride and thus to provoke fatigue. The fatiguing effect of this task was confirmed in a pretest. See Figure 9 for a detailed illustration of the task.

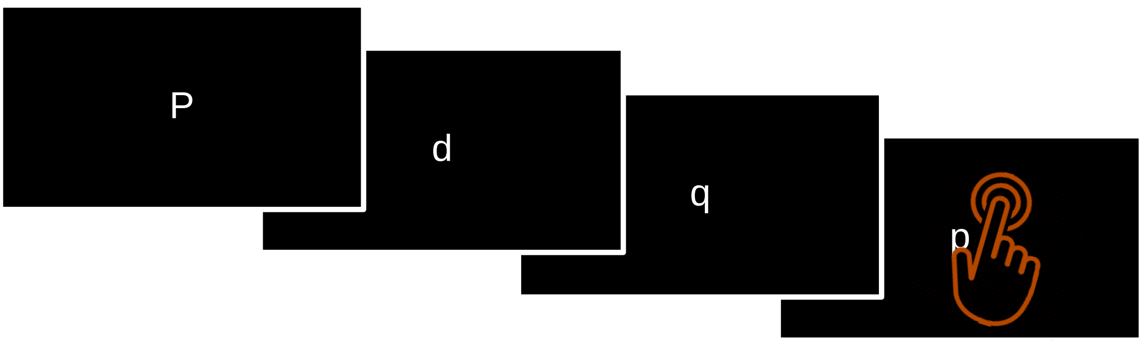


Figure 9: Pqpd-task.

3.2.2. Quiz-task

In contrast to the *Pqpd-task*, the *Quiz-task* was designed to prevent the human driver from getting fatigued while driving conditionally automated. The task looked similar to popular quizzes like *Quizduell* or *Quizup*. The *Quiz-task* consisted of a total of 500 questions. In a randomized sequence always one of these questions was displayed, which had to be answered with the help of four multiple choice answers by a touch-input on a Windows-Surface tablet computer. Always one answer was the right one, and the other three answers were wrong.

Correct or incorrect answers were highlighted with green or red animations on the screen. To further increase the activating effect of the task, a total score was displayed in the right side of the screen. With every right answer, the total score further increased. A similar *Quiz-task* had been used previously and successfully to prevent fatigue in CDA in a driving simulator experiment by Schömig et al. (2015). See Figure 10 for further details about the *Quiz-task* that was used in the experiments.

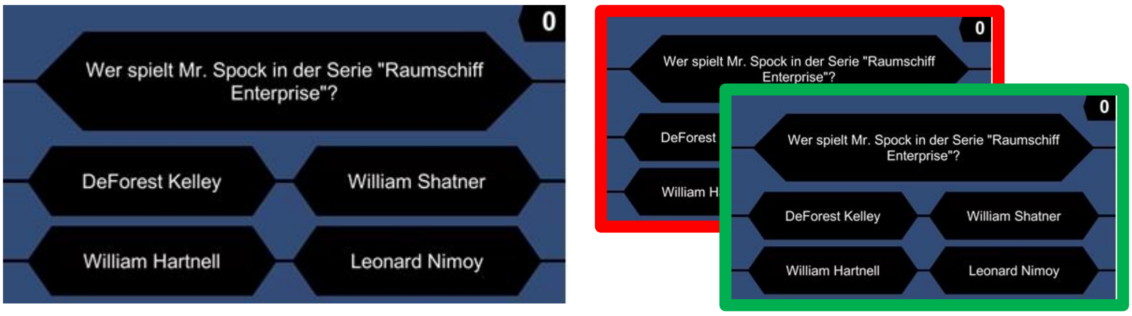


Figure 10: *Quiz-task.*

Thus, the two NDRTs that were used in Experiment 1 and Experiment 2 were designed in such a way that they differ in their effect on the fatigue state of the driver, but require the same resources. Both tasks were visually and motorically (touch-input) demanding.

3.2.3. Free-choice-activity

In Experiment 2, it could be shown that a prolonged engagement in the *Quiz-task* also led to emerging subjective fatigue (see section 5.). Therefore, in the *Wizard-of-Oz* experiment (Experiment 3) a freely selectable activity (*Free-choice-activity*) was used to investigate if a freely selectable activity can be used to prevent participants from getting fatigued while driving conditionally automated. The idea was that the participants should be engaged in activities that they enjoy to do while driving automated. Therefore, in the invitation to the experiment, the participants were told to bring an activity or a task, they would also like to do when traveling by train or airplane. The participants were allowed to work on their laptop, read a book or to do phone calls. Additionally, in the *Wizard-of-Oz* vehicle, a Windows Surface computer with a collection of games like Tetris and jump & run games was installed in front of the central information display (CID), which could be used by the participants. In addition, magazines and newspapers were available in the side pocket of the door which the participants were allowed to read during the automated ride.

3.3. Measurement Of Fatigue

The NDRTs explained in the previous section were used to affect the fatigue state of the drivers. Thus, one focus of the experiments was to investigate if and how fatigue emerges in CDA depending on different NDRTs. Therefore, it was necessary to measure emerging fatigue of the human drivers in the different test environments and experiments.

As already described in section 2.5., there are different methods of fatigue assessment which have already been used in experiments. However, for the purpose of a fatigue assessment in a CDA vehicle not all methods seem to be adequate. In addition, the method should detect the fatigue state of the participant non-intrusive and the system should be easy to install. In terms of a future implementation in a production vehicle such a system should not restrict the driver in his mobility and in the best case, the driver should not be aware of the measurement.

Based on these requirements, the following measuring instruments were used for the fatigue measurements in the experiments.

3.3.1. Objective Assessment: Electrocardiogram

For the measurement of ECG activity, the same measurement method was chosen as that was used by Schmidt (2019) in her series of experiments on thermal stimulation during passive driving periods. For this measurement a three-channel medical ECG system (G.tec Medical Engineering GmbH, 2018) was used. Based on the ECG signal, which is recorded at a frequency of 512 Hz, HR and different HRV measures can be calculated.

However, the results of Experiment 1 did not show any interpretable results. Rather, it seemed that a magnetic coupling of the steering-wheel in the simulator, which was necessary for the simulation of CDA, had interfered with the ECG signal. Consequently, the measurement of the ECG was omitted in the following experiments. Therefore the ECG measurement will not be discussed further and no results will be reported.

3.3.2. Objective Assessment: PERCLOS

For objective fatigue assessment over the courses of the different experimental rides, PERCLOS measurements were recorded.

The calculation of PERCLOS: To get PERCLOS, a fixed period of time β is considered (e.g. 60 s). All periods of time within β in which the eye is covered by the eye-lid to more than 80% are added to a sum α (e.g. 0.5 s + 0.3 s + 1.2 s + … = 6.3 s). The quotient of these two parameters α / β (e.g. 6.3 s / 60 s) gives the percentage of closed eyes (i.e. PERCLOS) within α (e.g. 10.5 %).

$$\frac{\alpha\ (sum\ of\ time\ intervals\ in\ which\ the\ eye\ is\ closed > 80\%)}{\beta\ (time\ period\ of\ consideration)} = PERCLOS$$

Furthermore, PERCLOS can also be used to classify different stages of fatigue. If the measured value of a person is above a defined threshold, a person’s state of fatigue is categorized as *tired*, *questionable* or as *awake*. Wierwille et al. (1994) proposes the following threshold values:

- 0.0 % – 7.5 % *awake*
- 7.5 % – 15.0 % *questionable*
- 15 % – 100.0 % *fatigued*

For the measurement and post-processing of PERCLOS, a head-mounted eye-tracker Dikablis Professional (Ergoneers GmbH, 2015) (see Figure 11) and the software D-Lab 3.0 (Ergoneers GmbH, 2014) was used.

The eye-tracker consists of a spectacle frame and a total of three cameras. One of the cameras is positioned between the eyes of the participant and records the panoramic view from the participant’s point of view. The other two cameras are each pointing at one eye of the participant. By combining the images of the eye-cams, overlaying it with the scene-cam and calibrating of the gaze of the participant it is possible to trace where the participant looked at, for each time. This was used for measurement of resulting eyes-on-road reaction time (i.e. t_{eyes}).

The main purpose of eye-tracking, however, was to measure the course of the fatigue state of the drivers during the experimental rides using PERCLOS. For this purpose, the camera images of the two cameras directed at the eyes, the so-called eye-cams, were used. These pictures each show the complete eye, including the pupil and eye-lid. With help of the D-Lab 3.0 software (Ergoneers GmbH, 2014), the recordings can be evaluated semi-automatically. The software should automatically recognize when the pupil is visible or when it is covered by the eye-lid. Each frame in which no complete pupil is visible, is evaluated as closed eye-lid, whereby PERCLOS increases. The sampling rate of all cameras is 60 Hz, resulting in 3600 frames per min per eye.

Figure 11: Dikablis Professional head-mounted eye-tracker.

Limitations of the eye-tracker and resulting consequences: It was already determined during the first experiment that the automated analysis of PERCLOS could not be carried out error-free by the software. False pupil detections, although the pupil was covered by the eye-lid at that time, or non-detections of the pupil, although it was clearly and completely visible, were the most common problems (see Figure 12). An exclusion of test persons with impaired vision who had to wear glasses and could not use contact lenses helped to improve the PERCLOS post-processing in the following experiments, but could not completely solve the problem. In order to obtain valid PERCLOS and eye-lid based fatigue data, all reported PERCLOS data were manually checked and adjusted if necessary. For this purpose, all relevant recordings were reviewed in slow motion and corrected in case of an error detection or missing detection. Only through this elaborate post-processing valid and interpretable eye-tracking data could be generated.

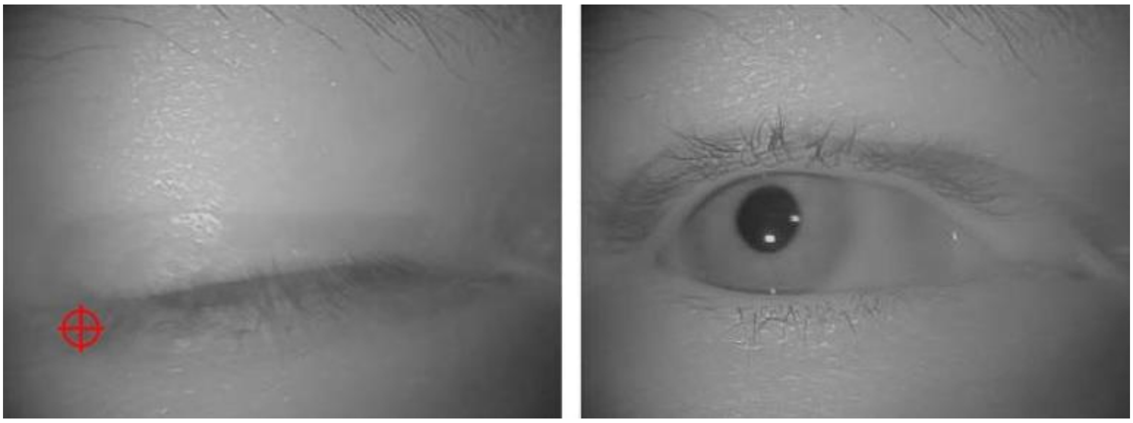


Figure 12: Errors in PERCLOS measurement: incorrect detection (left) & missing detection (right).

3.3.3. Subjective Assessment: KSS

Next to the objective fatigue assessment, subjective fatigue was assessed. Therefore the KSS, which was developed by Åkerstedt and Gillberg (1990) was used. A more recent review of the KSS was provided by Åkerstedt, Anund, Axelsson, and Kecklund (2014). In this review, the authors state, that the KSS highly correlates to driving performance and EEG indicators of sleepiness across individuals. They further state that high KSS values (> 6) are associated with an impaired driving performance. As already described in section 2.5.1., for fatigue assessment by the KSS, participants have to state their subjective fatigue during the 5 min period before the actual rating. For their answer a 9-point Likert scale with verbal anchors is provided. See Table 4 for the KSS-scale and the verbal anchors.

Table 4: KSS and corresponding verbal anchors.

Verbal anchor	KSS score
Extremely alert	1
Very alert	2
Rather alert	3
Alert	4
Neither alert nor sleepy	5
Some signs of sleepiness	6
Sleepy, but no effort to keep awake	7
Sleepy, but some effort to keep awake	8
Very sleepy, great effort to keep awake, fighting sleep	9

Both, the subjective and objective fatigue measurements were assessed for different points of time during the experiments in order to get a course of emerging fatigue depending on the duration of the automated ride and the engagement in the NDRT.

3.4. Assessment Of The Take-Over Performance In CDA

Another focus of this thesis was to investigate how *passive task-related fatigue* resulting due to NDRT-engagement affects take-over performance in CDA. For this purpose, the take-over performances of the participants were evaluated.

Two well-established and frequently used methods to assess take-over performance of the individuals are on the one hand the analysis of the achieved reaction times (e.g. hands-on reaction time) upon the RtI and on the other hand the analysis of the quality of the reaction of the drivers (e.g. longitudinal / lateral accelerations).

In the following, the metrics used for the assessment of take-over performance in the experiments conducted as part of this thesis are presented in more detail.

3.4.1. Time Based Aspects Of The Take-Over Performance

A system initiated RtI results in a series of different required actions the driver has to fulfill. All these actions must be executed within the available time-budget to make a successful take-over reaction possible. The reaction times required by the driver starting from the RtI-signal can be used to compare the take-over reaction time to the other drivers that participated in the experiment. In Table 5 the different reaction times that were assessed in the experiments (Experiment 1, Experiment 2 & Experiment 4) are explained in more detail. As far as possible, the reaction times are given in the correct order of their sequence.

Table 5: Overview of the different reaction times reported in the experiments.

Reaction time measurement (starting from the RtI)	Shortcut	Description
Eyes-on-road time (in s)	t_{eyes}	Time-interval the driver needs for turning his gaze towards the road scenery.
Hands-on-time (in s)	t_{hands}	Time-interval the driver needs for grabbing the steering-wheel with at least one hand.
First-touch-on-braking-pedal (in s)	$t_{brake\_reaction}$	Time-interval the driver needs for his first measurable touch on the braking pedal.
First-steering-intervention (in s)	t_{steer}	The first steering intervention is the point of time when the steering-wheel angle firstly exceeds 2°.
First-braking-maneuver (in s)	$t_{brake\_maneuver}$	The first braking intervention represents the point of time, the braking pedal shift firstly exceeds 10 % of the whole possible shift.

3.4.2. Aspects Concerning The Quality Of The Take-Over Behavior

The different drivers' reactions following an RtI can, next to the *time based categorization*, be evaluated concerning the *quality of the input of the drivers*. For the evaluation, different metrics can be used in order to evaluate the take-over reactions. However, the relevance of the different metrics strongly depend on the specific take-over situation and thus the maneuver required for handling the take-over scenario. For example, when the take-over scenario requires a lane-change maneuver, this goes along with higher values in the lateral acceleration, whereas a braking situation goes along with higher values in the negative longitudinal acceleration. Another metric that can be used is the time-to-collision ($TTC_{\_MIN}$). However, in some situations, there is no $TTC_{\_MIN}$ measureable. For example in sensor failure situations on a straight road. Therefore, each of these metrics can only be considered for one experiment or for experiments that contain the same take-over situation.

See Table 6 for a more detailed description of the measurements used in the experiments reported in this thesis.

Table 6: Dependent variables for takeover performance measurement.

Variable	Shortcut	Description
Maximum longitudinal acceleration (in m / s²)	**Acc_long_max**	Maximum absolute value of the acceleration in longitudinal direction measured in the take-over situation. In the case of a take-over situation, this are mostly braking reactions to avoid a collision. Acceleration is measured in m / s².
Maximum lateral acceleration (in m / s²)	**Acc_lat_max**	Maximum absolute value of the acceleration in lateral direction measured in the take-over situation. Acceleration is measured in m / s².
Minimal Time-to-collision (in s)	**TTC_MIN**	Time (in s) remaining until a potential collision (ISO International Organization for Standardization, 2013-07-23).

These two measurements (i.e. *time-based metrics* as well as *measurements indicating the quality of the drivers' input*) are the measurements that were used in most of the experiments on take-over performance so far, to assess the performance of the human drivers in the respective situations. However, these two metrics rarely evaluate a holistic take-over reaction, but rather individual aspects of the take-over reaction. Thus, many individual aspects can indicate a good take-over reaction, even if the take-over itself is of poor quality when viewed completely. For example, a fast take-over reaction is desired when it comes to reaction times and is therefore rated as a good take-over reaction concerning reaction time aspects. However, this does not mean that the take-over reaction was of good quality, too. It is the same the other way around. A take-over can be from good quality regarding the reaction of the driver, but if the driver cannot react within the available time-budget the whole situation can quickly become dangerous.

Therefore, to be able to rate the take-over performances in its entirety, next to the *time-* and *quality based metrics* also the *TOC-rating* was used for evaluation of the take-over reactions.

3.4.3. TOC–Rating – A Video-Based Expert Rating

All take-over situations were additionally rated by three trainer raters by using the TOC-rating tool, too. The tool is further described in 2.8.3.

For this rating process, all take-over reactions of all experimental rides conducted as part of this thesis were cut to 30 s videos, handed to three trained raters who independently rated the take-over reactions with the help of the rating sheet. Subsequently, mean values of the ratings of the three raters were calculated, which enabled group comparisons.

3.5. HMI Design Used In The Experiments

In the experiments reported in this thesis, a basic HMI concept was used to inform the driver about the current system status and to request him to intervene in case of a system initiated take-over situation. As this thesis was elaborated within the German funding project Ko-HAF, in which several automobile manufacturers, academia and other research institutions were involved, a standardized HMI concept was agreed upon. To make results comparable over all institutions, in all experiments the drivers were requested by a visual signal displayed in the central instrument cluster (see Figure 13). Additionally an urgent acoustic warning signal (beeping sound) was played simultaneously.

Figure 13: Visual states of the HMI concept.

Three different states of the CDA could be displayed in the central instrument cluster. When CDA was available, a grey steering-wheel was presented (see Figure 13, left). With activation of the system the steering-wheel icon was displayed in blue (see Figure 13, middle). When an RtI was issued, the steering-wheel icon turned red and two hands grabbing the wheel were displayed (see Figure 13, right). Next to the steering-wheel, text notices that informed the driver about the current system status were presented: *Autopilot verfügbar* (Eng.: *Autopilot available*), *Autopilot aktiviert* (*Autopilot activated*) or in the case of a RtI *Bitte manuell fahren* (*Please drive manually*) were displayed in regard to the current status. When a RtI was issued, simultaneously an auditory signal was presented and the NDRT on the Windows Surface was stopped (black screen).

3.6. Overview Of The Different Experiments

As part of this thesis, four experiments were conducted. Three of these experiments were conducted in the motion base driving simulator of the BMW group (see section 3.1.1.) and one was conducted in a *Wizard-of-Oz* vehicle (see section 3.1.2.). To some extent the experiments were based on each other but the experiments all had different main objectives. Next to that, different experimental designs were used for the experiments. In Table 7, an overview with detailed information about the number of participants, the experimental design, the focus of the experiment and the test-environment can be seen. In sections 4. – 7., the four experiments are described in detail.

Table 7: Overview of the four experiments conducted.

Name	Number of participants	Experimental design	Focus of the experiment	Environment
Experiment 1 (section 4.)	*N* = 56	Within-subjects	Impact of NDRTs on the drivers' state in automated driving in two 30 min rides; Take-over behavior after a RtI.	Driving simulator
Experiment 2 (section 5.)	*N* = 73	Between-subjects	Impact of NDRTs in prolonged automated driving (50 min) on the drivers' state; Take-over behavior after a RtI.	Driving simulator
Experiment 3 (section 6.)	*N* = 42	Between-subjects	Impact of NDRTs in prolonged on-road automated driving on the drivers' state.	*Wizard-of-Oz* vehicle
Experiment 4 (section 7.)	*N* = 64	Between-subjects	HMI concepts to support the fatigued driver in a *short-term* take-over situation	Driving simulator

3.7. Aims And Objectives

Based on the theoretical models and the knowledge generated so far, this thesis aims at a number of objectives. One of the main research aims was to investigate if fatigue emerges in CDA due to the passive situation for the human driver. Closely related to this topic was to investigate how different NDRTs affect the fatigue state of the drivers during the automated ride. In order to be able to assess the resulting fatigue, it was also necessary to identify suitable measuring instruments. Additionally, the transferability of the results from the driving simulator on real traffic environment on-road should be examined. Finally, the knowledge gained from the first experiments was used to design a possible HMI concept to support the driver in a *short-term* take-over situation.

For this purpose, a total of four studies were carried out. The main objectives of each experiment are explained in the following section:

3.7.1. Experiment 1

The first study was conducted in the motion base driving simulator at BMW facilities in Munich. In this first driving simulator experiment, the following research questions should be answered:

1. How do different NDRTs (*Pqpd-task* vs. *Quiz-task*) affect the drivers' fatigue state in CDA?
2. How can emerging fatigue be best assessed in CDA? Therefore different fatigue measurements (KSS, PERCLOS, ECG) were used and compared.
3. How does fatigue affect the take-over performance after a system initiated RtI in CDA?

3.7.2. Experiment 2

The second experiment was based on the first study. However, in contrast to the first experiment, effects of prolonged automated driving were investigated in this second experiment. Main objectives of the experiment were:

1. How do different NDRTs (*Pqpd-task* vs. *Quiz-task*) during prolonged CDA (50 min) affect the drivers' fatigue state?
2. How can emerging fatigue be best assessed in CDA? Therefore different fatigue measurements (KSS, PERCLOS) were used. How does the measured data change in comparison to the 25-min ride of Experiment 1? Does fatigue continue to increase?
3. How does the resulting fatigue (that emerged due to the NDRT) affect take-over performance after a system initiated RtI after prolonged CDA after 50 min?

3.7.3. Experiment 3

In the third study, the transferability of the results from Experiment 1 and Experiment 2 from the driving simulator to real driving environment should be examined. Therefore, a *Wizard-of-Oz* approach (see section 3.1.2. for more information about the *Wizard-of-Oz* method) was used to conduct an experiment with a comparable experimental design in real driving environment. Main research questions of this experiment were:

1. How do different NDRTs (*Pqpd-task* vs. *Free-choice-activity*) in CDA in on-road driving environment affect the drivers' fatigue state?
2. Are the measurements that were suitable for fatigue measurement in the driving simulator also suitable on-road?
3. Are the results from the driving simulator experiments (concerning the drivers' fatigue state) comparable to the course of the drivers' state on-road?

3.7.4. Experiment 4

The scope of the fourth and last experiment was to investigate how to best support a fatigued driver after a prolonged CDA ride in a take-over situation. For safety issues, this experiment was again conducted in the motion base driving simulator. The following research questions should be answered:

1. Which HMI concept can best support the driver in a *short-term* take-over situation? Therefore three different concepts with different characteristics were used and tested against each other.

2. Are there certain concepts that help the driver to perceive the situation correctly and more quickly?

3. Are there differences in the subjective experience between the three concepts?

In the following four sections, the experiments are described in more detail. See section 4. for Experiment 1, section 5. for Experiment 2, section 6. for Experiment 3 and section 7. for Experiment 4. In section 8., the results of the experiments are compared. A discussion of the findings is given in section 9.

4. Experiment 1 - Effects Of NDRTs On Drivers' Fatigue In Conditional Driving Automation

In the first experiment that was conducted as part of this thesis it should be investigated how different NDRTs during CDA affect drivers' fatigue. The NDRTs were used in a within-subjects design to either prevent the human driver from fatigue (*Quiz-task*) or to provoke fatigue (*Pqpd-task*). The *Pqpd-task* is a low-irritant task and should lead to *passive task-related fatigue (*May & Baldwin, 2009; see section 2.4.1. for more details*)*. Another aim of Experiment 1 was to investigate how emerging fatigue in CDA can be measured. Therefore, the drivers' subjective fatigue was assessed over the course of the two experimental rides by using the KSS. Next to the subjective assessment, PERCLOS and ECG data were recorded over the course of the experiment. In order to investigate how fatigue affects the take-over performance in a *short-term* take-over situation, a RtI was triggered after 25 min in both rides.

The results of Experiment 1 and some of the following statements have already been published in:

Jarosch, Kuhnt, Paradies, & Bengler (2017)

4.1. Summary

With introduction of conditional automation in vehicles the driver can engage in NDRTs and only has to intervene in case of a RtI. Therefore, *active task-related fatigue*, which is the most frequent form of fatigue in manual driving, is assumed to be replaced by *passive task-related fatigue*, due to increasing monotony and a rather passive situation for the driver in CDA. To investigate effects of different NDRTs on drivers' fatigue and take-over capability a driving simulator study was conducted. In total, 56 participants experienced two CDA rides on a highway. In each one of the two rides, participants either had to fulfill a monotonous monitoring task or an activating task. In each ride, they were assigned to one of these two tasks. During the rides, fatigue was measured with PERCLOS and KSS. Results suggest that a monotonous monitoring task in CDA provokes fatigue. Furthermore, PERCLOS could be confirmed as a valid measurement for detecting fatigue. Take-over performance in a system-initiated take-over situation after 25 min of CDA was not affected with regard to the randomly executed NDRT.

4.2. Introduction

In order to avoid repetitions and redundant explanations and derivations of theoretical models, this section summarizes what was introduced in the original publication and where this information can be found in this thesis.

In the introduction- and theory part of this paper first the taxonomy of automated driving referring to the SAE was described (see section 2.2.2.). The link between passive activities (i.e. NDRTs) and the resulting *passive task-related fatigue* was explained (see section 2.4.1.). Furthermore, the consequences of fatigue in road traffic were considered (see section 2.4.2.).

Summarized, consequences of fatigue in manual driving are increased crash risk due to deteriorated reaction times and reduced driving performance. Transferred to CDA, fatigue affects reaction times and take-over quality upon RtIs. It is also assumed, that CDA reinforces monotony whilst driving and that NDRTs can potentially affect drivers' fatigue.

Objective of this study was to investigate (i) how *passive task-related fatigue* can be measured in CDA, (ii) how different NDRTs affect the drivers' fatigue state in CDA, and (iii) whether fatigue affects reaction time and take-over quality upon a RtI.

4.3. Method

4.3.1. Participants

Fifty-six employees of the BMW Group voluntarily participated in the study. The sample consisted of nine female and 47 male participants. Mean age was 30.10 years (*SD* = 9.00 yrs, minimum = 20 yrs, maximum = 56 yrs). The participants were experienced drivers with mean driving experience of 12.29 yrs (*SD* = 9.36 yrs). The majority of the sample had experienced at least one driving assistance system (78.57 %). ACC was the system most participants had experienced before (75.00 %).

4.3.2. Experimental Design

All participants experienced two different NDRTs and two different take-over situations in a counterbalanced order during two experimental rides, resulting in a two-factor within design. Two types of NDRTs, which were presented on a tablet installed in the central console of the car, were used to affect drivers' fatigue. A monotonous monitoring task

(*Pqpd-task*) was used to induce passive task-related fatigue. In reference to Warm et al. (2008), a task with a low event rate and a fixed event location was selected. Different letters *P*, *q*, *p* and *d*) were presented on the screen for a variable time between 10 s – 15 s in a mixed order. Subjects had to touch the screen every time the *p* was displayed. The other task instead had the purpose to keep the drivers on an adequate level and prevent them from emerging fatigue (*Quiz-task*). Therefore, in reference to Schömig et al. (2015) a *Quiz-task* was used. Participants had to choose the right answer to a question out of four possibilities.

To test effects of resulting fatigue on take-over performance two different take-over scenarios (*accident on ego-lane* / *sensor failure in a bend*) occurred in the end of the two rides. Both scenarios were not predictable and highly critical but differed in the complexity of the required intervention by the drivers.

Driver fatigue, affected through the two NDRTs, was assessed using the self-report measurement KSS, developed by Åkerstedt and Gillberg (1990). As objective method for measuring fatigue, PERCLOS (Wierwille et al., 1994) was used. Self-reported and objective sleepiness was measured repeatedly over the course of the two experimental rides for each seven defined times (t1 – t7; see Figure 14). KSS was assessed after eye-tracking to not affect data.

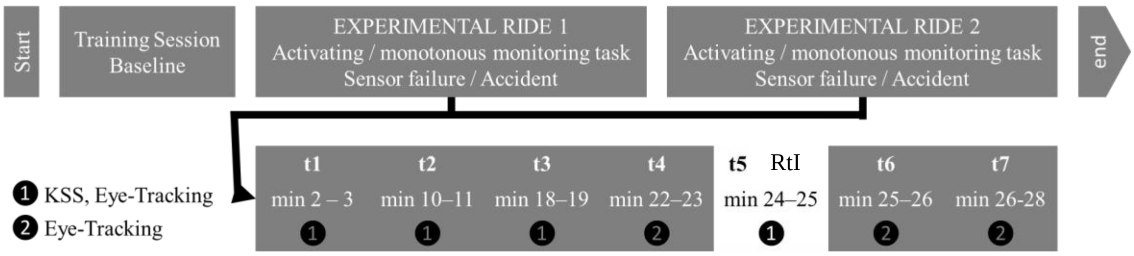


Figure 14: Experiment 1: Experimental Procedure.

Take-over performance was assessed using take-over reaction times (braking > 10 % of pedal position / steering input > 2º) and driving performance related parameters (longitudinal acceleration, lateral acceleration, steering-wheel-angle, steering-angle velocity and SDLP).

Data were collected in single 105 min experiments. The virtual driving scenario for all sessions was a three-lane freeway with a hard shoulder. At the beginning of each experiment, participants were briefed on the driving simulator and the CDA system which provided lateral and longitudinal control, including lane changes and overtaking. In a first 10 min training session, participants were familiarized with the simulator, the CDA and the

RtI. A 1 min baseline for eye-tracking data was recorded when the car drove with CDA. The following two experimental rides were identically concerning the route (~ 30 min, 59 km), the surrounding traffic (low to middle traffic) and weather conditions (cloudy, no rain). The RtI situation happened in min 25 (after 50 km). After the first experimental ride, participants had to leave the simulator for a break to regularize fatigue affected through the first NDRT. In the second ride, each the other NDRT was presented and the other take-over scenario occurred.

4.3.3. Apparatus

The experiment was conducted in the motion base driving simulator at BMW Group in Munich. In order to avoid repetitions see section 3.1.1. for information concerning the driving simulator that was used.

4.4. Results

A significance level of α = .05 was set for all hypothesis tests. In a first step the results of the fatigue state of the drivers as a dependency of the NDRTs are reported. In a second step the take-over performance is addressed.

4.4.1. The Course Of The Drivers' State Over The Two Rides Depending On The NDRT

All drivers have experienced two different rides. In one of the rides they had to engage in the *Quiz-task* whereas in the other ride they had to engage in the *Pqpd-task*. In the following sections it is reported how the subjective (KSS) and the objective fatigue (PERCLOS) differed in these two rides.

4.4.1.1. Subjective Fatigue (KSS)

KSS was highest in the *Pqpd-task* group after 18 min of CDA, *M* = 6.25, *SD* = 1.97 and lowest in the *Quiz-task* group after 2 min of the CDA ride, *M* = 3.70, *SD* = 1.50. See Table 8 for further information.

Table 8: Experiment 1: Descriptive KSS data for the two NDRTs.

Time of assessment	Task	*N*	*M*	*SD*
min 2	Quiz	53	3.70	1.50
	Pqpd	53	4.13	1.95
min 10	Quiz	53	3.77	1.44
	Pqpd	53	5.28	2.07
min 18	Quiz	53	4.11	1.69
	Pqpd	53	6.25	1.97

For further analysis of the subjective KSS data, a repeated measures ANOVA with the factor task (*Quiz-task* and *Pqpd-task*) and the different times of measurements (min 2, 10 and 18) was calculated. A Greenhouse-Geisser correction was used for correction of violations of sphericity.

Results showed a significant main effect of the NDRT on subjective fatigue over time, F (1, 104) = 17.95, $p < .001$, $\eta p^2 = .15$. There was also a significant main effect of time of measurement on KSS, F (1.47, 152.55) = 60.28, $p < .001$, $\eta p^2 = .37$, and a significant interaction effect between the factor of the NDRT and the time of the measurement, F (1.47, 152.55) = 27.83, $p < .001$, $\eta p^2 = .21$. Thus, the *Pqpd-task* induced a higher level of subjective fatigue compared to the *Quiz-task* over the course of the experimental rides. Self-reported sleepiness (KSS) increased significantly when participants had to deal with the monotonous monitoring task (i.e. *Pqpd-task*) whereas it did not change significantly during the *Quiz-task*. See Figure 15 for further information.

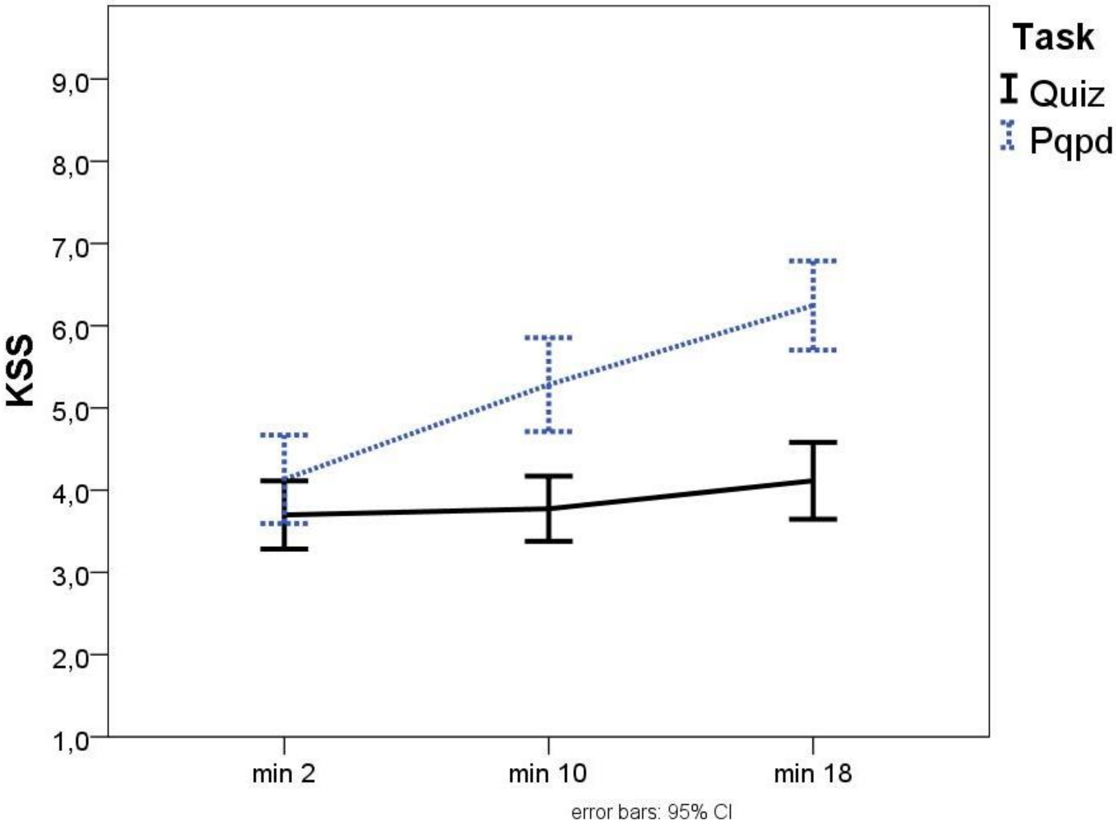


Figure 15: Experiment 1: KSS over the course of the experimental rides for the two NDRTs.

4.4.1.2. Objective Fatigue (PERCLOS)

PERCLOS was highest in the *Pqpd-task* group after 18 min of the ride, M = 11.47, SD = 9.80, and lowest in the *Quiz-task* group after 2 min of the ride, M = 3.19, SD = 1.94. In Table 9 you can see all mean PERCLOS values including the SD values for the three times of measurement (min 2, 10 and 18).

Table 9: Experiment 1: Descriptive PERCLOS data for the two NDRTs.

Time of assessment	Task	*N*	*M*	*SD*
min 2	Quiz	54	3.19	1.94
	Pqpd	54	5.08	3.20
min 10	Quiz	54	3.51	2.14
	Pqpd	54	7.84	5.54
min 18	Quiz	54	3.56	2.26
	Pqpd	54	11.47	9.80

Results showed a significant main effect of the NDRT on PERCLOS over time, F (1, 106) = 35.4, p < .001, ηp^2 = .25. The monotonous monitoring task (*Pqpd-task*) induced a higher level of PERCLOS compared to the activating task (*Quiz-task*). There was also a significant main effect of the factor time on PERCLOS, F (1.32, 139.73) = 25.24, p < .001, ηp^2 = .19, and a significant interaction effect between the tasks and time, F (1.32, 139.73) = 20.24, p < .001, ηp^2 = .16. As Figure 16 illustrates, PERCLOS increased with time spending on the NDRT only in the monotonous monitoring task (*Pqpd-task*) condition.

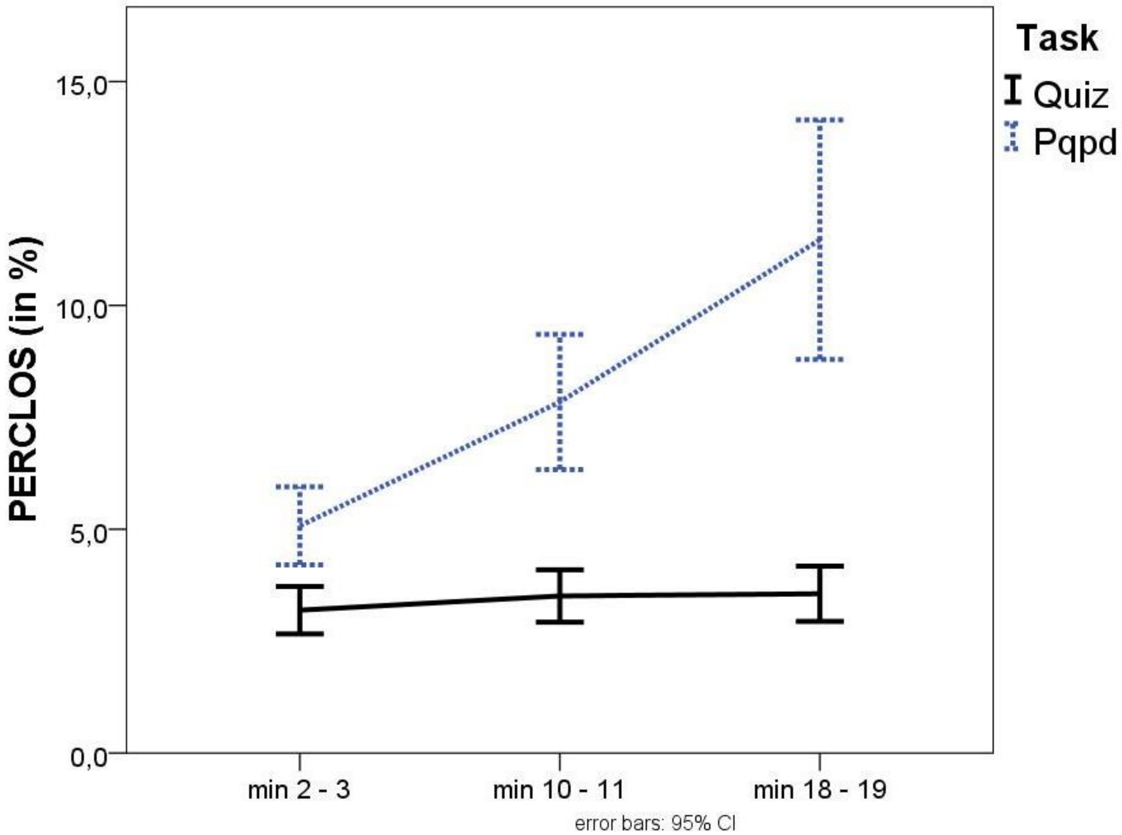


Figure 16: Experiment 1: PERCLOS over the course of the experimental rides for the two NDRTs.

4.4.2. Take-Over Reactions Of The Drivers' Depending On The NDRT

4.4.2.1. Take-Over Reaction Times

Differences in the reaction times due to the engagement in the different NDRTs were calculated with one-way unpaired t-tests for both situations starting from the RtI. A Welch-correction was used if the homogeneity of variances was violated. The factor of the independent variable was the NDRT and the reaction times were the dependent variables.

Accident on ego lane situation: None of the evaluated reaction times differed significantly between the two different task groups (*Pqpd-task* or *Quiz-task*). The first drivers' reaction in both groups was t_{eyes}, the eyes-on-road reaction time. This reaction was followed in both groups by a hands-on steering-wheel reaction (t_{hands}). After the hands-on reaction, in both groups a first braking reaction ($t_{brake\_reaction}$) was the next reaction. After his reaction, a first steering maneuver (t_{steer}) was observed in both groups. The last reaction was the first braking maneuver ($t_{brake\_maneuver}$). See Figure 17 and Table 10 for further information. Implausible values (e.g. $t_{hands} < 0.5$ s) were excluded from the analysis.

Table 10: Experiment 1: Reaction times (in s) for the two task groups in the *accident on ego lane* situation.

Reaction-time measure	Task	*N*	*M*	*SD*	Min	Max
t_{eyes}	Quiz	27	0.76	0.38	0.19	1.84
	Pqpd	25	0.94	0.47	0.32	2.31
t_{hands}	Quiz	23	1.14	0.64	0.66	3.40
	Pqpd	23	1.00	0.42	0.60	2.62
$t_{brake\_reaction}$	Quiz	28	2.37	1.04	0.74	5.10
	Pqpd	28	2.42	1.17	0.80	5.46
t_{steer}	Quiz	28	2.92	1.48	1.16	8.34
	Pqpd	28	2.71	1.79	1.12	7.44
$t_{brake\_maneuver}$	Quiz	28	3.25	1.30	0.74	5.36
	Pqpd	28	3.05	1.39	0.82	6.48

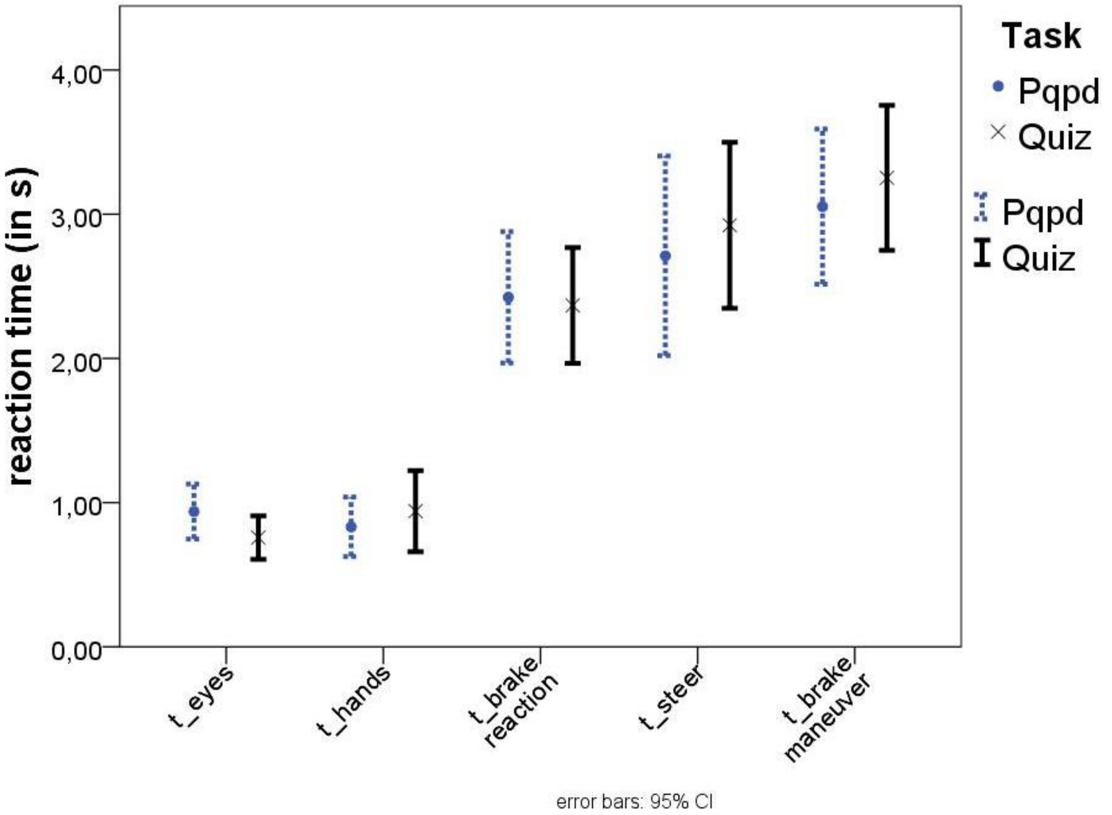


Figure 17: Experiment 1: Reaction times for the two task-groups in the *accident on ego lane* situation.

Sensor failure in a bend situation: In the *sensor failure in a bend* situation t_{eyes} differed significantly between the two task groups, t (33.57) = 2.77, $p < .01$, d = 0.76, r = 0.35. With M = 1.03 s (SD = 0.59 s) the participants who had to deal with the *Pqpd-task* needed approximately 0.3 s (95 %-CI [0.09 s, 0.59 s]) longer for their first eyes-on-road reaction compared to the participants that had to engage in the *Quiz-task* (M = 0.69 s, SD = 0.23 s). Also the reaction time for the first braking reaction ($t_{brake\_reaction}$) differed significantly between the two task-groups in this situation, t (12.48) = -2.86, p = .014, d = 1.3, r = 0.55. The participants that had to engage in the *Quiz-task* (M = 1.98 s, SD = 0.75 s) braked about 0.75 s (95 %-CI [-1.31 s, -0.17 s]) slower compared to the participants that had to engage in the *Pqpd-task* (M = 1.23 s, SD = 0.32 s).

All other reaction times did not differ significantly between the two different task-groups (*Pqpd-task* or *Quiz-task*). The first reaction of the drivers in both groups was the eyes-on-road reaction (t_{eyes}). This reaction was followed in both groups by a hands-on steering-wheel reaction (t_{hands}) and a first touch on the braking pedal ($t_{brake\_reaction}$). The first steering maneuver (t_{steer}) was the last reaction in both groups. Since a hard braking reaction was not necessary in this situation and only n = 8 reacted by a braking meneuver, $t_{braking\_maneuver}$ is marked grey. See Figure 18 and Table 11 for further information. Implausible values (e.g. hands-on times < 0.5 s) were excluded from the analysis.

Table 11: Experiment 1: Reaction times (in s) in the *sensor failure in a bend* situation according to the two tasks.

Reaction-time measure	Task	*N*	*M*	*SD*	Min	Max
t_{eyes}	Quiz	27	0.69	0.23	0.27	1.14
	Pqpd	27	1.03	0.59	0.32	2.72
t_{hands}	Quiz	22	1.06	0.23	0.6	1.54
	Pqpd	20	1.1	0.35	0.6	1.68
$t_{brake\_reaction}$	Quiz	10	1.98	0.75	1.02	3.12
	Pqpd	9	1.23	0.32	0.84	1.78
t_{steer}	Quiz	28	2.33	0.81	1.14	4.04
	Pqpd	28	2.49	0.96	0.86	4.88
$t_{brake\_maneuver}$ (n = 8)	Quiz	2	1.28	0.03	1.26	1.30
	Pqpd	6	1.49	0.43	0.94	2.06

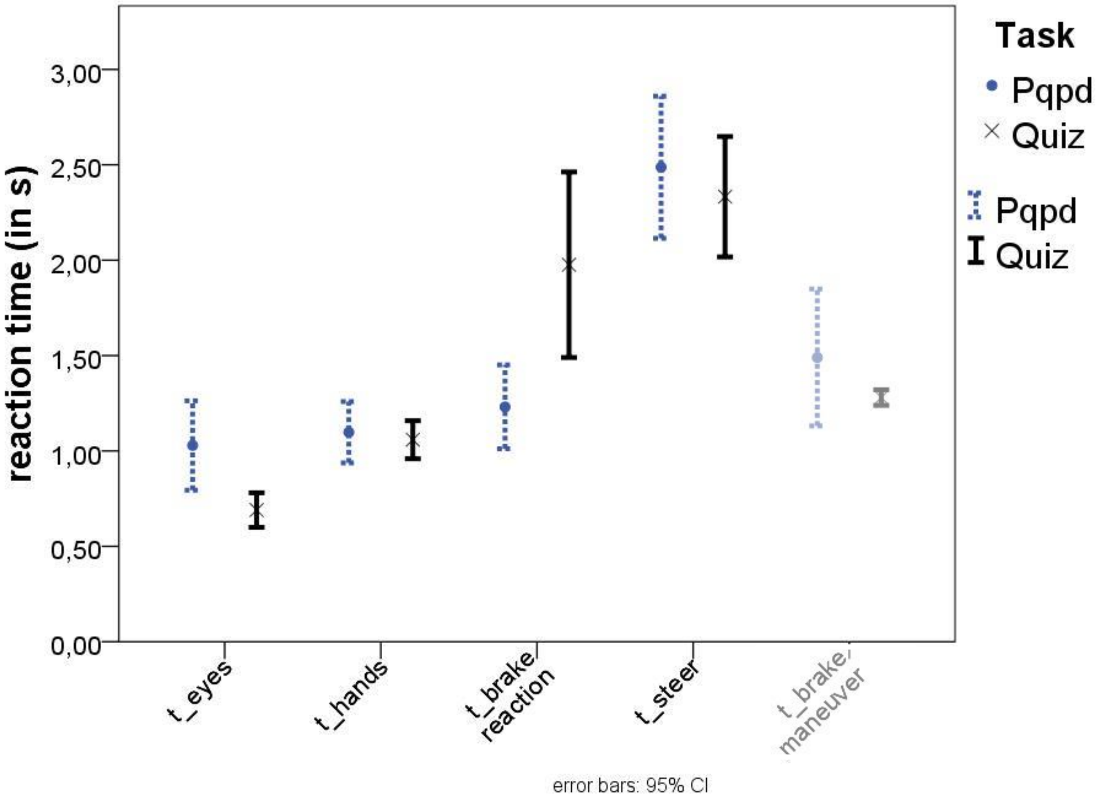


Figure 18: Experiment 1: Reaction times in the *sensor failure in a bend* situation.

4.4.2.2. Quality metrics of the take-over reaction

Metrics indicating the quality of the take-over reaction were calculated with unpaired t-tests for each of the two situations. The factor of the independent variable was the NDRT and the quality metrics were the dependent variables. Mean minimum times-to-collision ($TTC_{\_MIN}$) as well as mean maximum longitudinal accelerations ($Acc_{\_long\_max}$) and mean maximum lateral accelerations ($Acc_{\_lat\_max}$) were analyzed.

Accident on ego lane situation*:* $TTC_{\_MIN}$ differed significantly between the two task groups, t (51) = 2.09, p = .04, d = .57, r = -.27, whereas the participants from the *Pqpd-task* (M = 2.63 s, SD = 1.0 s) had higher $TTC_{\_MIN}$ compared to the *Quiz-task* group (M = 2.14 s, SD =0.7 s). In mean, the difference was 0.5 s (95 %-CI [-0.96 s, -0.02 s]. Also the mean $Acc_{\_lat\_max}$ differed significantly between the two different task groups in the *accident on ego lane* situation, t (54) = 2.62, p = .011, d = 0.698, r = 0.33. In this situation, the participants that had to deal with the *Quiz-task* steered 0.6 m / s² (95 %-CI [0.14 m / s², 1.04 m / s²]) stronger (M = 1.89 m / s², SD = 0.85 m / s²) compared to the participants that had to deal with the *Pqpd-task* (M = 1.3 m / s², SD = 0.84 m / s²). $Acc_{\_long\_max}$ did not differ significantly between the two task groups. See Table 12 and Figure 19 for further details.

Table 12: Experiment 1: Quality metrics for the input of the drivers in the *accident on ego lane* situation.

	Quality measure	Task	*N*	*M*	*SD*	95%-CI		Min	Max
*	$TTC_{\_MIN}$	Quiz	28	2.14	0.70	1.87	2.41	1.00	3.30
		Pqpd	25	2.63	1.00	2.22	3.05	1.04	4.89
	$Acc_{\_long\_max}$ (in m/s²)	Quiz	28	5.58	3.59	4.19	6.97	1.01	10.47
		Pqpd	28	6.22	3.68	4.79	7.65	0.43	11.19
*	$Acc_{\_lat\_max}$ (in m/s²)	Quiz	28	1.89	0.85	1.56	2.22	0.49	4.63
		Pqpd	28	1.30	0.84	0.97	1.62	0.35	3.55

Note. * $p < .050$, ** $p < .010$

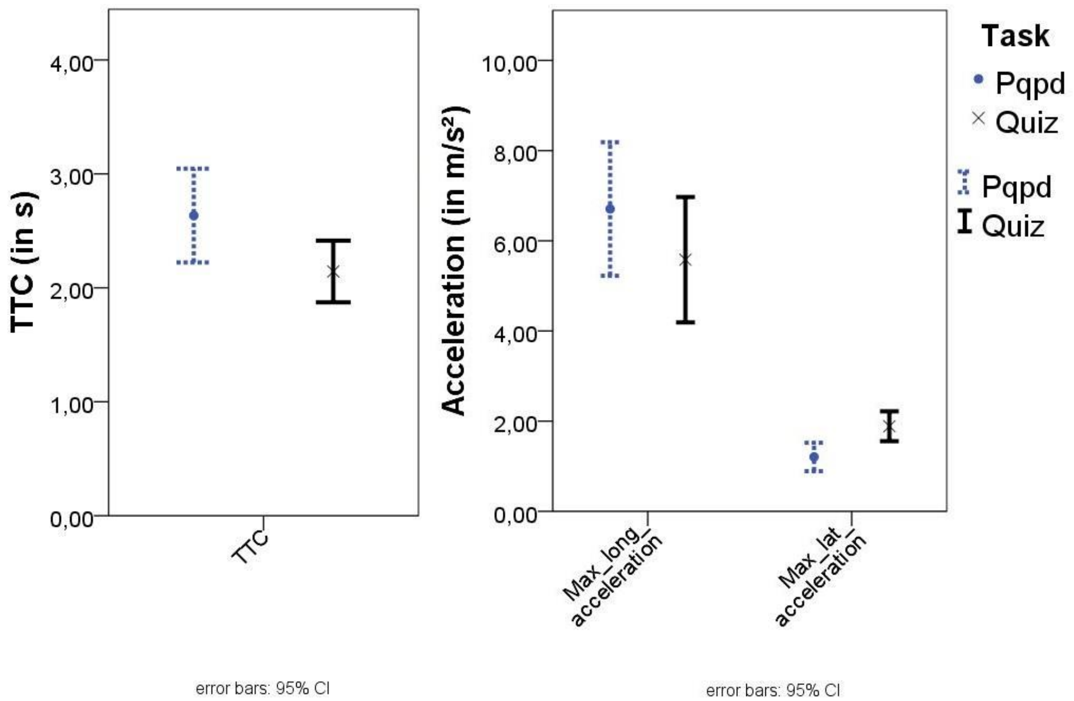


Figure 19: Experiment 1: Quality metrics upon RtI in the *accident on ego lane* situation.

Sensor failure in a bend situation: Mean $Acc_{\_long\_max}$ as well as mean $Acc_{\_lat\_max}$ did not differ significantly between the two different task groups in the *sensor failure in a bend* situation. $TTC_{\_MIN}$ was not analyzed in the *sensor failure in a bend* situation, as no braking maneuver was required. See Table 13 and Figure 20 for further details.

Table 13: Experiment 1: Quality metrics for the input of the drivers in the *sensor failure in a bend* situation.

Quality measure	Task	*N*	*M*	*SD*	95%-CI		Min	Max
$Acc_{\_long\_max}$ (in m/s²)	Quiz	28	0.85	0.85	0.52	1.18	0.43	4.03
	Pqpd	28	1.43	2.17	0.59	2.27	0.43	10.24
$Acc_{\_lat\_max}$ (in m/s²)	Quiz	28	2.33	0.82	2.00	2.65	1.04	4.43
	Pqpd	28	2.46	1.07	2.04	2.88	1.26	7.07

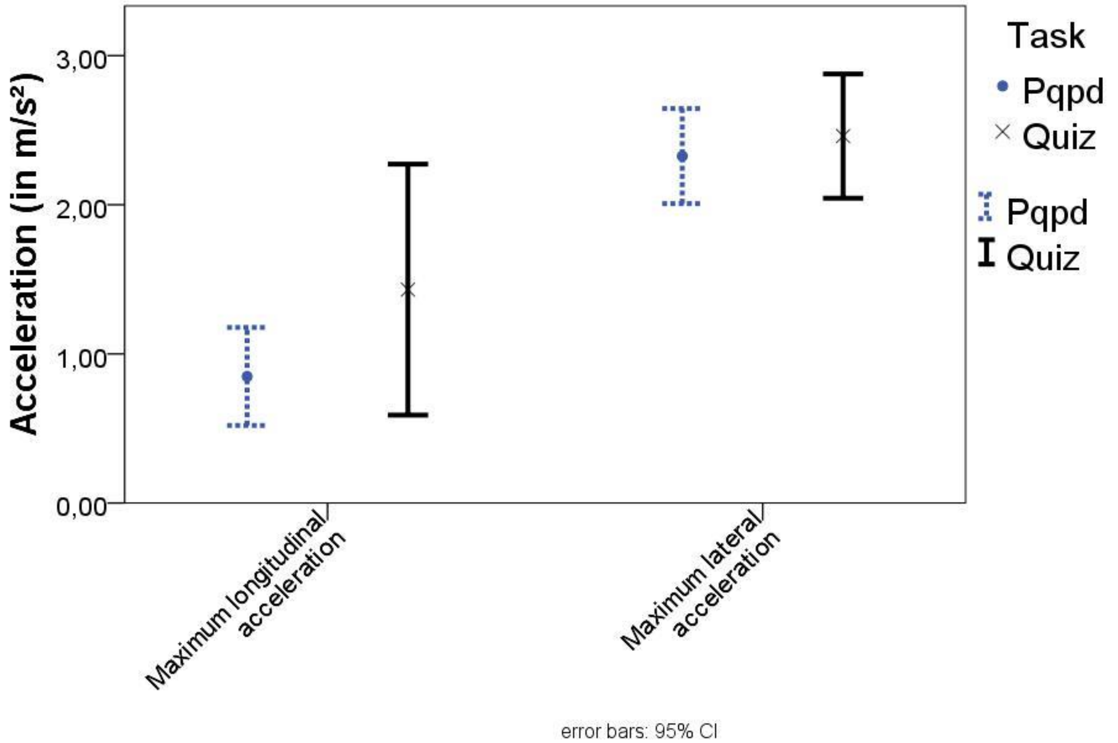


Figure 20: Experiment 1: Quality metrics for the input of the drivers in the *sensor failure in a bend* situation.

4.4.3. TOC-Rating

Next to reaction times and measurements indicating the quality of the drivers' input, the TOC-rating was used to assess take-over performances.

Accident on ego lane situation: According to the TOC-rating, the take-over performances did not differ between the two task groups significantly. In the *Quiz-task* group as well as in the *Pqpd-task* group a *Median* = 5 was achieved, exact Mann-Whitney-*U*-test: *U* = 332.50, *p* = .93. The effect size is *r* = .01 and corresponds to a very low effect (Cohen, 1988). See Figure 21 for further details.

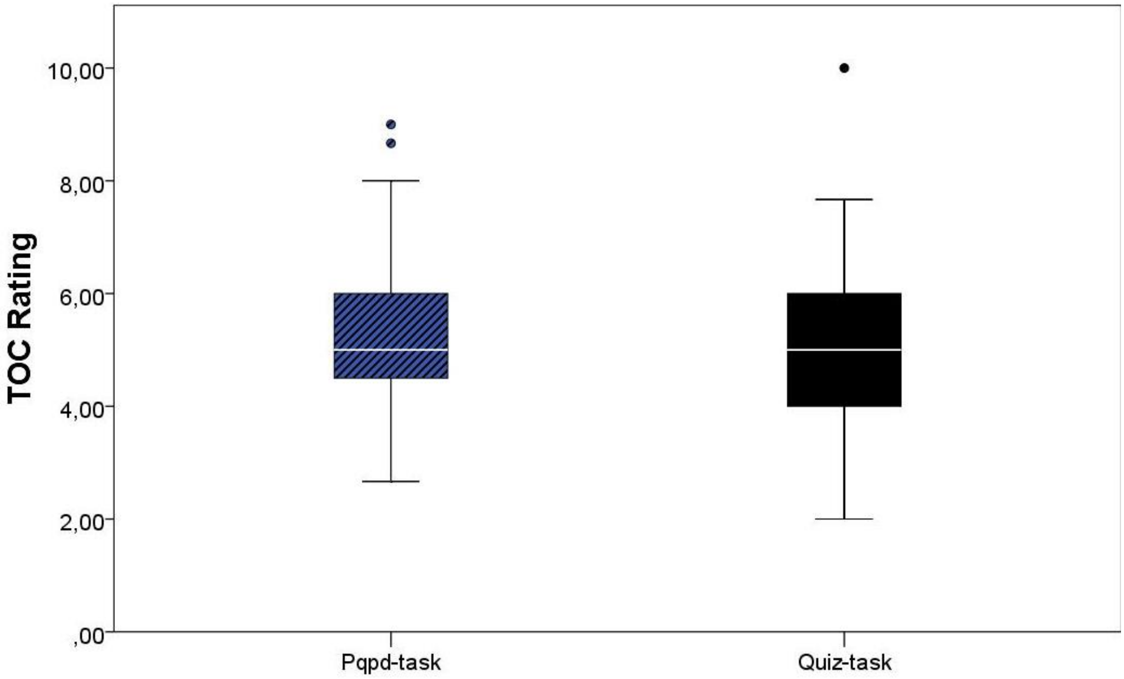


Figure 21: Experiment 1: TOC-rating for the two task-groups in the *accident on ego lane* situation.

Sensor failure in a bend situation: According to the TOC-rating, the take-over performances in the *sensor failure in a bend* situation differed significantly between the two task groups. In the *Quiz-task* group participants reacted better (*Md* = 4.00; smaller values indicate a better take-over performance compared the *Pqpd-task* group (*Md* = 4.66), exact Mann-Whitney-*U*-test: *U* = 246.00, *p* = .01. The effect size according to Cohen (1988) is *r* = .34 and corresponds to a medium effect. See Figure 22 for more details.

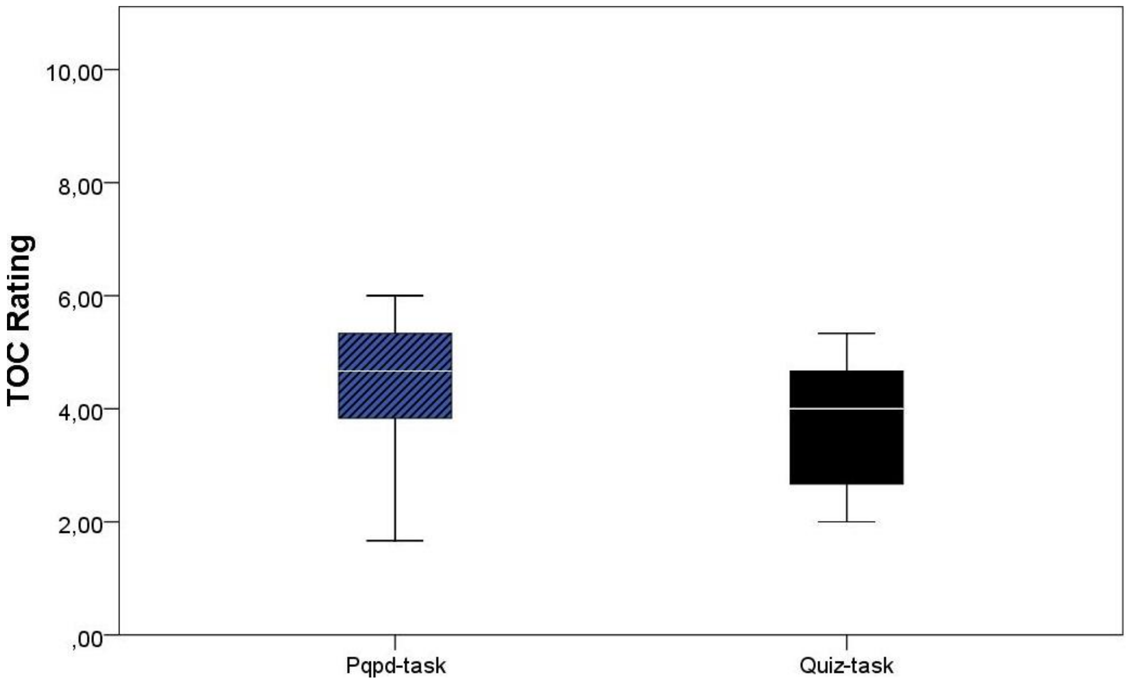


Figure 22: Experiment 1: TOC-rating for the two task-groups in the *sensor failure in a bend* situation.

4.5. Discussion

The objectives of this study were (i) to investigate how emerging fatigue can be measured in CDA, (ii) to investigate effects of different NDRTs in CDA on subjective and objective fatigue, and (iii) to investigate if resulting fatigue affects take-over performance.

As hypothesized NDRTs could be used to provoke fatigue in a 25 min conditionally automated ride. When participants had to deal with the *Pqpd-task*, a monotonous monitoring task, a significant increase in the self-reported KSS could be found compared to the participants that had to engage in the *Quiz-task* that was used to prevent participants from getting fatigued. Next to the subjective measurement, also objective eye-tracking data supported this finding. PERCLOS increased with time-on-task whilst participants dealt with the monotonous monitoring task (*Pqpd-task*). When participants had to deal with the *Quiz-task*, instead, PERCLOS stayed on a significant lower level. If the average values of PERCLOS after about 20 min of the CDA ride are considered, it becomes clear that the fatigue level is in the range *questionable*, especially for the *Pqpd-task* (Wierwille, 1994). In the *Pqpd-task* group n = 12 test persons achieved values above 15 %, which classifies them as *tired*, whereas no participant in the *Quiz-task* group achieved values above 15 %.

Thus, the *Pqpd-task* that was used to provoke *passive task-related fatigue*, fulfilled its purpose. KSS and PERCLOS indicate, that fatigue increased during this monotonous monitoring task. This has shown that *passive task-related fatigue* can emerge through passive tasks or NDRTs and that these effects occur after less than 20 min of a conditional automated ride.

Objective PERCLOS and the subjective KSS could be confirmed as reliable fatigue measurements as the values changed significantly over time and depending on the two NDRTs. PERCLOS increased only in the *Pqpd-task* condition that should provoke *passive task-related fatigue*. Applicability of eye-tracking as drowsiness detection in CDA for prolonged CDA rides, different races and light conditions should be further investigated in the future.

The physiological measures recorded during the experiment did not show any results regarding increasing fatigue. In addition, many data were affected by noise signals (driver movement or magnetic coupling of the steering-wheel), which made further evaluation much more difficult. In the following experiments no physiological data was recorded.

After 25 min of the two rides and engaging in either the *Pqpd-task* or in the *Quiz-task* a take-over situation with a time-budget of 7 s occurred. There were a few significant differences in the take-over reaction between the two task groups (*Quiz-task* vs. *Pqpd-task*). In the *accident on ego lane* situation, no significant differences in the reaction times between the two tasks could be found. However, when considering the measurements indicating the quality of the drivers' input, $TTC_{\_MIN}$ as well as $Acc_{\_lat\_max}$ differed significantly in this situation. The $TTC_{\_MIN}$ in the *Pqpd-task* group was significant lower compared to the *Quiz-task* group, indicating that participants from the *Quiz-task* group got closer to the accident or drove faster. In the same way $Acc_{\_lat\_max}$ is significantly different between the two task groups, indicating a stronger steering input in the *Quiz-task* group. Regarding the effect sizes, both effects are to be classified as small to medium effects which do not majorly affect the take-over performance. In the other take-over situation, the *sensor failure in a bend* situation, t_{eyes} and the $t_{brake\_reaction}$ differed significantly. All other measurements did not result in significant differences between the two task-groups. The participants that had to engage in the *Quiz-task* reacted about 0.3 s faster considering t_{eyes} reaction time but needed about 0.75 s longer until they braked compared to the participants from the *Pqpd-task* group. Both effect sizes indicate a medium effect. However, a braking reaction was not necessary in this situation and is therefore not relevant.

The TOC-rating showed a significant difference between the two task groups in the *sensor failure in a bend* situation. Accordingly, the participants from the *Quiz-task* group reacted better compared to the *Pqpd-task* group. In the accident situation, there were no differences between the task groups in the TOC-rating. Overall, it is noticeable that the *sensor failure in a bend* situation was solved in a safer way than the *accident on ego lane* situation. This is also evident when considering the TOC-rating.

4.6. Précis

- In a driving simulator experiment effects of two NDRTs in CDA were examined.
- One task was designed to provoke fatigue (*Pqpd-task*) and the other task was designed to prevent fatigue (*Quiz-task*).
- The participants experienced both tasks in a within-subjects experimental design for each 25 min while driving conditionally automated.
- Emerging fatigue was assessed by subjective KSS and objective PERCLOS measurements to defined points of time during the rides.

- When participants had to deal with the monitoring task (*Pqpd-task*) fatigue did emerge, whereas fatigue did not emerge when participants had to deal with the *Quiz-task*.
- The resulting fatigue did not affect take-over performance negatively.

5. Experiment 2 – NDRTs And Effects On Drivers' Fatigue In Prolonged Conditional Driving Automation

In the first experiment it could be demonstrated that different NDRTs during CDA affect the drivers' fatigue state. When the participants drove conditionally automated and simultaneously had to engage in a monotonous monitoring task (*Pqpd-task*), fatigue did emerge. A *Quiz-task* on the other hand, could prevent emerging fatigue while driving conditionally automated. PERCLOS and KSS could be demonstrated to be reliable for the measurement of emerging fatigue. Negative effects of *passive task-related fatigue* on take-over performance due to the *Pqpd-task* could not be demonstrated.

However, it is assumed that CDA will be applicable for longer than 25 min periods at a time. As CDA will first be available on the Autobahn or on highways, rather a long-term use of CDA is expected (Bishop, 2005). Therefore, in Experiment 2, effects of NDRTs in combination with prolonged CDA should be investigated.

The following section is based on a previous publication:

Jarosch, Bellem, & Bengler (2019)

5.1. Introduction And Theoretical Issues

Since some major parts of the introduction and the theoretical issues section would overlap with those of the present thesis, only a brief summary of what was described in the two chapters in the original publication, and where this can be found in the present thesis, is given here. All experiment-specific issues such as hypotheses, description of the participants or the exact procedure of the experiment are presented in detail. For the full paper please refer to the reference.

Summary of issues that were discussed in the original publication: First the SAE taxonomy of automated driving was explained. The focus was on the clarification of the CDA concept and the particularities of CDA (e.g. take-over situations, see section 2.2.3.). In addition, possible problems of increasing automation and the fatigue model of May & Baldwin (see section 2.4.1.) were explained. Further, consequences of fatigue in CDA were discussed (see section 2.4.3.) and measurements for fatigue were illustrated (see section 2.5.).

5.1.1. Hypotheses

One aim of the study was to investigate the effects of task-engagement in two different NDRTs during CDA on driver fatigue. Participants either had to deal with a monotonous monitoring task or an activating *Quiz-task*. We hypothesized that (i) an engagement in a monotonous monitoring task (*Pqpd-task*) leads to increased subjective and objective fatigue compared to an activating task (*Quiz-task*) and that (ii) an activating task can prevent emerging fatigue. The second aim of the study was to investigate the effects of fatigue provoked through the monotonous monitoring task on take-over performance. We expected that (iii) fatigue leads to an impaired take-over performance when it comes to a *short-term* take-over situation.

5.2. Materials and Methods

5.2.1. Participants

In this experiment *N* = 73 employees of the BMW Group voluntarily participated. Of those, *n* = 4 participants were excluded from further analysis due to simulator problems (e.g. missing RtI signal in the take-over situation). Another *n* = 2 participants were excluded from further analysis because of missing eye-tracking data (PERCLOS). One participant was excluded as he did not engage in the NDRT and was obviously waiting for a take-over situation, resulting in a sample size of 66. The sample consisted of 14 female (21.21 %) and 52 male (78.78 %) participants. The mean age of the participants was 31.36 yrs (*SD* = 9.86 yrs, minimum = 20 yrs, maximum = 60 yrs). The participants were drivers with a mean driving experience of 13.83 yrs (*SD* = 9.68 yrs). The sample quoted their driving experience as *experienced* (83.33 %) or *very experienced* (15.15 %). Just one participant was *inexperienced* (1.5 %) in driving.

Most participants had already experienced at least one driver assistance system (83.56 %). Table 14 shows the experience with driver assistance systems for the two conditions and results of a chi-square analysis indicating that there were no differences in their experiences with ADAS between the two NDRT groups before.

Table 14: Experiment 2: Experiences of the participants with driver assistance systems.

	ACC yes / no	Lane Keeping Ass. yes / no	Emergency Braking yes / no
Pqpd	27 (81.81 %) / 6 (18.18 %)	20 (60.60 %) / 13 (39.39 %)	11 (33.33 %) / 22 (66.66 %)
Quiz	24 (72.72 %) / 9 (27.27 %)	21 (63.63 %)/ 12 (36.36 %)	12 (36.36 %) / 21 (63.63 %)
χ^2 / *p*	χ^2 = 0.78; *p* = .38	χ^2 = .06; *p* = .8	χ^2 = .06; *p* = .8

Before the experimental ride, the majority of participants indicated to be *very awake* to *awake* (60.60 %). A state between *awake* and *tired, without problems staying awake* was indicated by 34.85 % and 4.54 % quoted to be *tired, without problems staying awake*. KSS did not differ between the two conditions before the experimental ride with *M* = 3.55 (*SD* = 1.52) for the *Pqpd-task* and *M* = 3.72 (*SD* = 1.4) for the *Quiz-task*, *t* (64) = -.48, *p* = .63.

5.2.2. Apparatus

The study was conducted in the motion base driving simulator at BMW group laboratories (see section 3.1.1.). PERCLOS was measured using the Dikablis head-mounted eye-tracker (see section 3.3.2.).The NDRTs were presented on a Windows Surface tablet (10.8 in) mounted in front of the CID.

5.2.3. Experimental Setup

Five participants were tested per day (8:15 am, 10:15 am, 1:00 pm, 3:00 pm and 5:00 pm). Each trial took about 1:45 hrs. Upon arrival, participants were given a written description of the CDA system. They were informed about the characteristics of the system as well as system boundaries including the possibility of RtIs. A confidentiality statement and a demographic form with questions about driving experience, age, gender and experience with ADAS had to be filled out by the participants. Afterwards, the head-mounted eye-tracker was calibrated for each participant.

This was followed by a familiarization ride to accustom participants with the driving simulator. In this session, participants first had to drive manually and then had to activate the CDA. After 2 min of automated driving, the examiner verbally explained that RtIs can occur. After this, an RtI was triggered by the examiner and participants had to take over control of the car. Thereupon, participants had to activate the system of CDA two more times and take over control of the system by a braking reaction and by pushing the button on the steering-wheel. According to Hergeth, Lorenz, and Krems (2017), prior familiarization with the system of CDA and its system boundaries strongly influences take-over performance. Due to the fact that some participants had previously participated in CDA experiments, the familiarization was necessary to get comparable data of all participants.

The training session was followed by the experimental ride. In the experimental ride, drivers had to deal with an NDRT for the entire ride. After 50 min of automated driving, the take-over situation occurred.

5.2.4. Non Driving Related Tasks

NDRTs were used to affect the drivers' fatigue level. Participants either had to deal with a monotonous monitoring task (*Pqpd-task*), which should induce *passive task-related fatigue*, or an activating *Quiz-task*, which should prevent participants from fatigue. To ensure task engagement, participants were told that the processing of the task has priority during CDA. See sections 3.2.1. and 3.2.2. for further details about the two NDRTs that were used in this experiment.

5.2.5. Specifics Of Conditional Driving Automation

Participants were told to switch on the CDA after entering the highway and to drive automated for the entire ride. The system of CDA is further explained in 3.1.1.

5.2.6. Human Machine Interface

See section 3.5. for information about the HMI used in this experiment. The HMI design was identical in all simulator experiments.

5.2.7. Scenario

The experimental ride was conducted on a three-lane highway with a hard shoulder. Traffic density was low and guidance was mostly straight. There were two elongated curves and hardly any overtaking situations during the whole ride. Weather conditions were set to a clouded sky to create an ideally monotonous situation.

After 50 min of automated driving, a take-over scenario occurred. For testing of human performance in take-over situations, Gold et al. (2018) recommends high urgency, low predictability, high criticality and medium to high complexity of the driver's intervention for the scenario.

The take-over was requested due to an unpredictable *accident on ego lane* situation: in the moment of the RtI, the ego vehicle is traveling on the right lane of a three lane highway. A hard shoulder is located to the right of the ego vehicle. Left of the car, there are two additional lanes with two cars in the lane left of the ego vehicle. As there was a 100-meter gap between the two vehicles, a braking or lane-change maneuver was possible to avoid an accident.

At the time of the RtI, the ego vehicle drove with a speed of 130 km / h (80.8 mph) and time-to-collision (TTC) with the accident on the ego-lane in front was highly urgent at 7 s. In the event that the driver did not intervene, the system would not deactivate and crash into the accident. Thus, the scenario is highly critical.

5.2.8. Dependent Variables

Effects on the drivers' state: PERCLOS (see section 3.3.2.) as well as KSS (see section 3.3.3.) were assessed for different points of time during the experimental ride. See Table 15 for the different time points of measurement.

Table 15: Experiment 2: Defined points of measurement (� PERCLOS measurement, afterwards KSS).

t1	t2	t3	t4	t5
Min 2-3	15 – 16	30 – 31	45 – 46	50
�	�	�	�	RtI

Effects on take-over performance: To measure the performance of the drivers in the take-over situation, *time based aspects* (i.e. reaction times) and *aspects concerning the quality of the take-over behavior* were recorded. Additionally the TOC-rating was used for the quality-assessment of the take-over reactions. See section 3.4. for further details about these methods.

5.3. Results

The main objective of the study was to examine how engagement in NDRTs during prolonged CDA affects drivers' fatigue over the course of the ride and how task-induced fatigue affects take-over performance.

5.3.1. Effects Of Prolonged CDA And The Engagement In NDRTs On The Drivers' State

To examine the effects of the prolonged conditional automated ride in combination with the execution of different NDRTs, PERCLOS and KSS were evaluated using mixed ANOVAs. The factor time of measurement was used as the within-subjects factor and the NDRT was used as the between-subjects factor. To evaluate how the NDRTs affected the drivers' fatigue state, times of measurement (t1, t2, t3 and t4) before the RtI were examined.

Effects on PERCLOS: PERCLOS was highest in the *Pqpd-task* group at t4 (M = 10.47, SD = 12.57) and lowest in the *Quiz-task* group at t1 (M = 2.83, SD = 2.39).

A mixed ANOVA with Greenhouse-Geisser correction showed that there was a statistically significant main effect for the factor time of measurement on PERCLOS, F (2.02, 129.15) = 8.47, p < .001, *partial* η^2 = .12. There was also a significant effect of the NDRT, F (1, 64) = 7.95, p = .01, η^2 =.11. Next to the main effects, there was also a significant interaction effect between time of measurement and the NDRT, F (2.02, 129.15) = 3.15, p < .05, *partial* η^2 = .05. Thus, fatigue in the *Pqpd-task* group increased faster and more intensively compared to the *Quiz-task* group. See Figure 23 for further information.

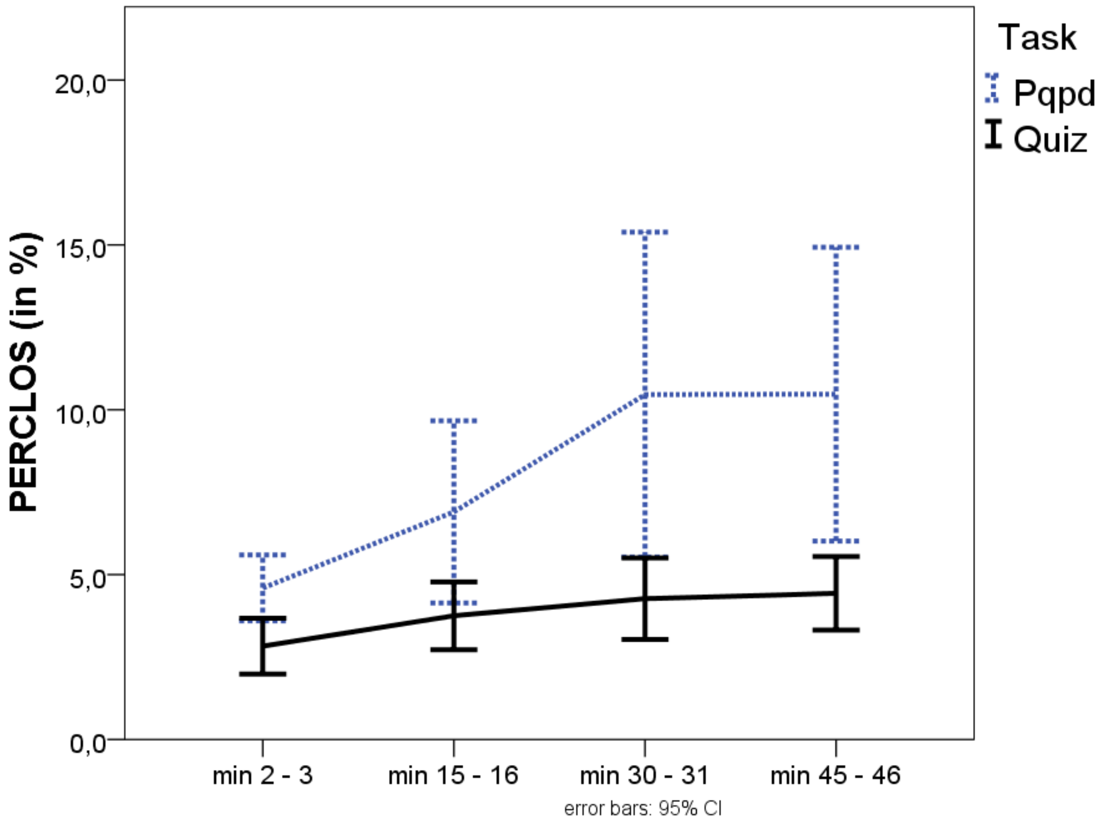


Figure 23: Experiment 2: PERCLOS over the course of the ride for the two tasks.

Effects on KSS: KSS was highest in the *Pqpd-task* group at t4 (M = 5.64, SD = 1.93) and lowest in the *Quiz-task* group at t1 (M = 3.15, SD = 1.25).

A mixed ANOVA with Greenhouse-Geisser correction showed that there was a statistically significant main effect for the factor time of measurement on KSS, F (2.53, 161.61) = 61.50, p < .001, *partial* η^2 = .49. There was no significant effect of the NDRT on the KSS, F (1, 64) = 2.14, p = .15, η^2 = .03. There was also no significant interaction effect between the time of measurement and the NDRT, F (2.53, 161.61) = 1.32, p = .27, *partial* η^2 = .02. See Figure 24 for more information to the KSS ratings.

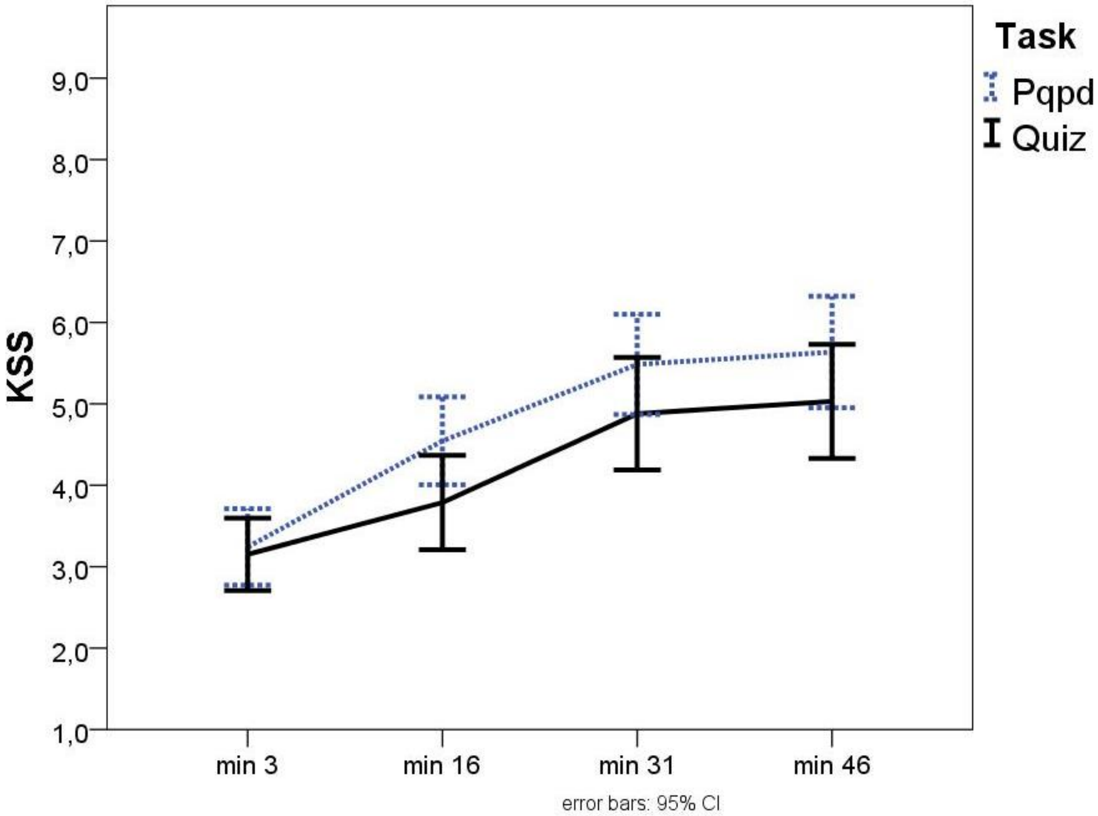


Figure 24: Experiment 2: KSS over the course of the experiment.

5.3.2. Reactions Of The Drivers After The RtI

To investigate effects of the NDRTs on the take-over performance, the reactions of the drivers upon the RtI were assessed. Due to the large variety of reactions of the human drivers, it was distinguished between three different types of possible reactions:

- *braking maneuver*: stopping behind the accident (v < 20 km/h (12.4 mph))
- *lane change maneuver*
- *accident* / *loss of control*

In Figure 25, the different driver reactions upon the RtI are presented. In both task-groups, there were $n = 33$ participants. There was no statistically significant difference in the take-over behavior between the task-groups, ($p = .20$, *Fisher's exact test*).

Accidents occurred in both NDRT groups, whereas $n = 4$ drivers from the *Pqpd-task* condition compared to $n = 2$ drivers from the *Quiz-task* condition lost control of the vehicle or could not avoid an accident.

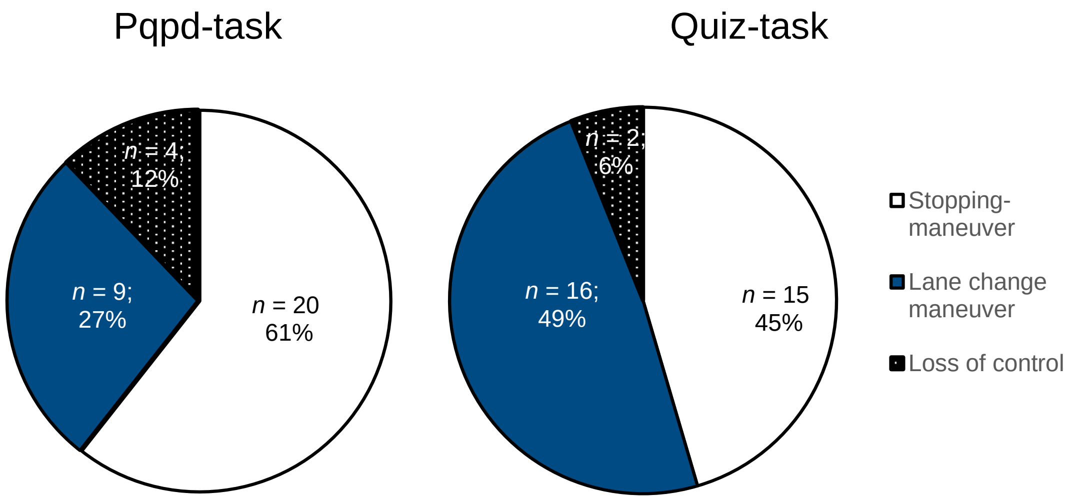


Figure 25: Experiment 2: Reactions of the drivers upon the RtI.

5.3.3. Results In Take-Over Performance

To evaluate the effects of NDRTs in prolonged CDA on take-over performance, take-over performance measures after the RtI were analyzed. Different driving performance parameters and reaction times were examined using unpaired t-tests. The dependent variable is the particular driving parameter. The two different NDRTs represent the factor of the group.

5.3.3.1. Results In The Take-Over Reaction Times

Significant differences in the reaction times according to the NDRT could be found in the t_{eyes} and the $t_{brake\_reaction}$. See Table 16 and Figure 26 for further details.

Table 16: Experiment 2: Reaction times for segments of the take-over process according to tasks.

Initial Reaction	Pqpd		Quiz		*df*	*t*	*p*
	M	*SD*	*M*	*SD*			
** t_{eyes}	1.18 s	0.39 s	0.85 s	0.31 s	63	3.84	<.001
t_{hands}	1.27 s	0.50 s	1.32 s	0.63 s	63	.36	.72
t_{steer}	3.01 s	1.92 s	3.27 s	2.24 s	64	.51	.61
* $t_{brake\_reaction}$	2.33 s	0.97 s	1.71 s	0.94 s	60	2.63	.01
$t_{brake\_maneuver}$	2.83 s	1.90 s	2.29 s	1.46 s	60	1.68	.10

Note. * $p < .050$, ** $p < .010$

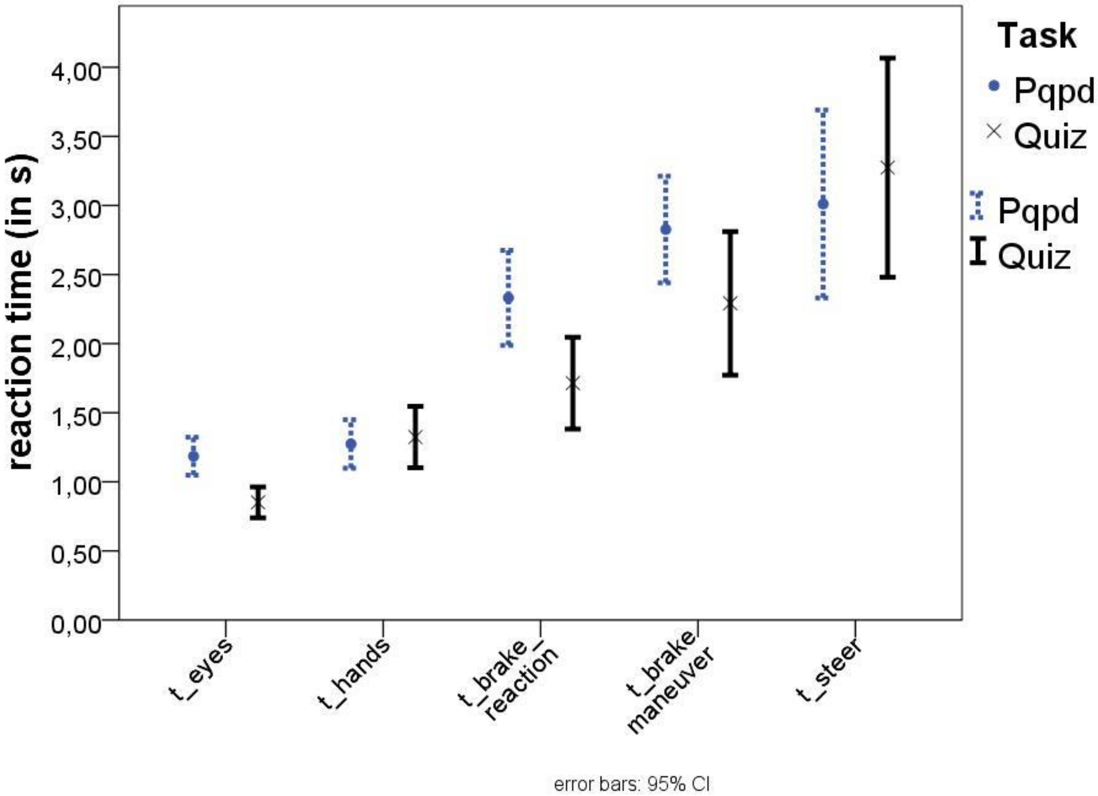


Figure 26: Experiment 2: Take-over reaction times for the two task groups.

5.3.3.2. Results In The Driving Performance Parameters In The Take-Over Situation

Acc$_{_long_max}$: Acc$_{_long_max}$ did not differ between the two NDRT-groups with the *Pqpd-task* group braking a bit stronger (*M* = -8.69 m / s², *SD* = 2.1 m / s²) compared to the *Quiz-task* group (*M* = -8.25 m / s², *SD* = 2.56 m / s²). A t-test with Welch-correction indicated that there is no statistically significant difference between the two tasks, *t* (61.70) = -.77, *p* = .44.

Acc$_{_lat_max}$: Acc$_{_lat_max}$ did not differ between the two NDRT-groups with the *Quiz-task* group steering a bit stronger (*M* = 3.56 m / s², *SD* = 2.79 m / s²) compared to the *Pqpd-task* group (*M* = 3.11 m / s², *SD* = 2.28 m / s²). There was no statistically significant difference between lateral accelerations for the *Pqpd-task* group and the quiz group, *t* (61.59) = -.72, *p* = .47.

TTC$_{_MIN}$: TTC$_{_MIN}$ did not differ between the two NDRT- groups significantly, *t* (58) = .355, *p* = .724. *TTC$_{_MIN}$* was a bit higher in the *Pqpd-task* group (*M* = 2.98 s, *SD* = 1.19 s) compared to the *Quiz-task* group (*M* = 2.87 s, *SD* = 1.19 s).

See Figure 27 for a more detailed overview of the quality metrics in the take-over situation.

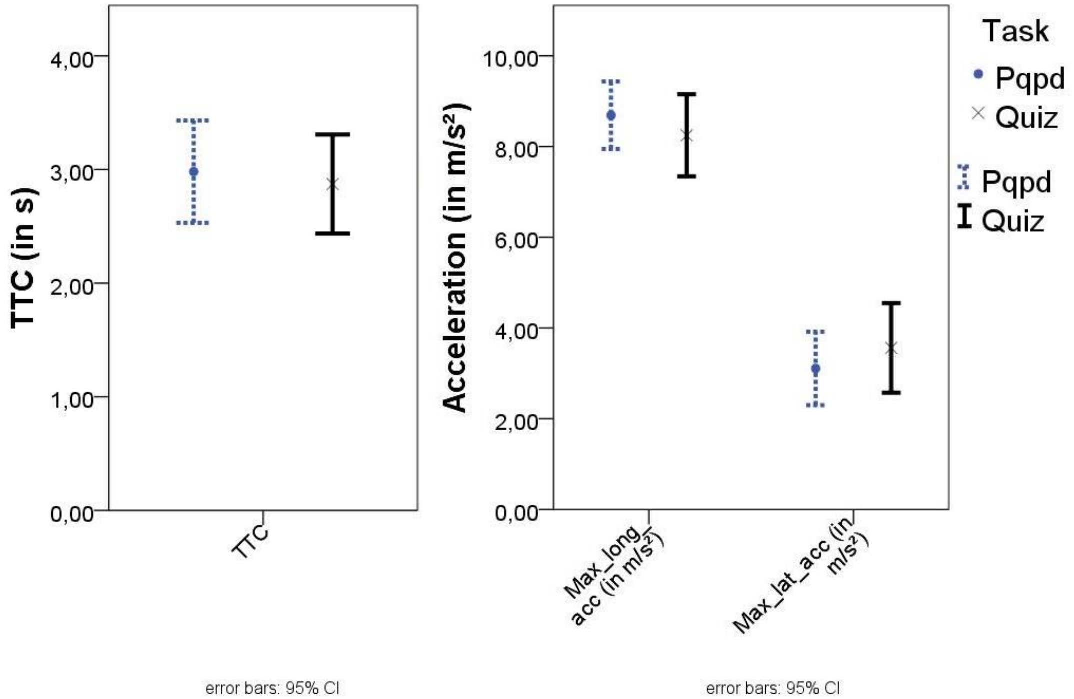


Figure 27: Experiment 2: Quality metrics in the take-over situation according to the two tasks.

5.3.4. TOC-Rating

Again, a TOC-rating was used for the assessment of the different take-over reactions upon RtI. Therefore, three trained raters independently watched the 30 s videos and rated the take-over performance using the TOC-rating sheet. See section 3.4.3. for further details about the TOC-rating.

According to the TOC-rating, the take-over performances did not differ between the two NDRT-groups significantly. However, the *Pqpd-task* group reacted a bit better (Md = 4.67, high values go along with a bad take-over performance) compared to the *Quiz-task* (Md = 5.00), exact Mann-Whitney-U-test: U = 480.00, p = .41. The effect size according to Cohen (1988) is r = .10 and corresponds to a low effect. See Figure 28 for further details.

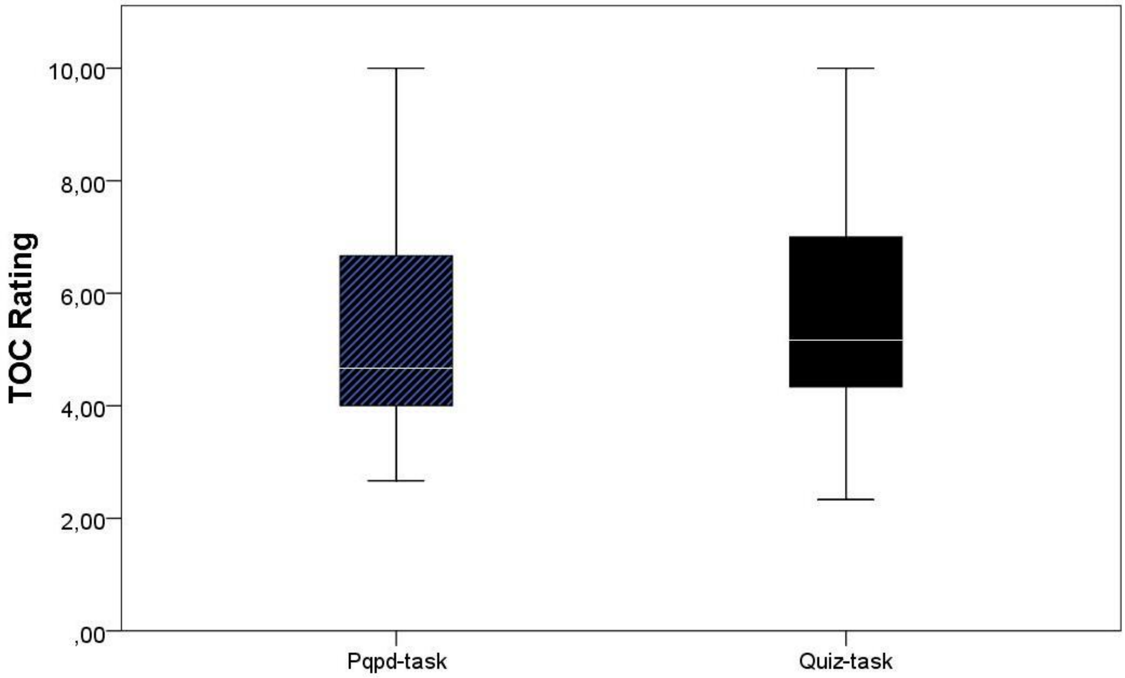


Figure 28: Experiment 2: TOC-rating according to the two task conditions.

5.4. Discussion

The main objective of the current experiment was to examine the effects of the engagement in NDRTs in prolonged CDA on the drivers' state and how this affects take-over performance.

Therefore, participants either had to deal with a fatiguing monitoring task (*Pqpd-task*), which was used to fatigue the drivers or with a *Quiz-task*, which was used to prevent drivers from getting fatigued. The monitoring task is quite similar to system supervising of the automated system whereas the *Quiz-task* is more like an activity with the infotainment system or an own device. Another focus of the experiment was, if emerging fatigue due to *passive task-related fatigue* affects take-over performance in CDA.

The manipulation of fatigue by the two NDRTs worked. Participants who had to engage in the monitoring task (*Pqpd-task*) showed higher PERCLOS measures, compared to the participants who had to engage in the activating *Quiz-task*. The results of this experiment suggest that prolonged CDA and simultaneously engaging in a monitoring task (i.e. monitoring the automated driving system) can lead to fatigue. This may lead to a driver response delay and thus negatively affect take-over performance after an RtI. This is indicated on the one hand due to the worsened reaction times and on the other hand due to the higher number of accidents and worse take-over performances of the *Pqpd-task* group. A higher mean PERCLOS in the *Pqpd-task* group was associated with slower reaction times after the RtI. The reaction time measures t_{eyes} and $t_{brake\_reaction}$ showed significant differences related to the NDRT. Participants who engaged in the *Quiz-task* focused on the road center earlier compared to the *Pqpd-task* group. In addition, the first-braking reaction was faster in the *Quiz-task* group.

A fast glance reaction on the road as well as a fast braking reaction can be advantageous for the following processes in a take-over situation: surrounding traffic can be detected earlier, situation awareness can be established faster and, with the fast braking reaction, the resulting time budget for the take-over situation increases. Thus, a lane change maneuver became more prevalent in the *Quiz-task* group compared to the *Pqpd-task* group.

Next to deteriorated reaction times, also a higher number of accidents after the RtI occurred, when participants had to engage in the *Pqpd-task* (n = 4) compared to the *Quiz-task* (n = 2). However, the majority of the participants could still avoid an accident after 50 min of automated driving.

The take-over reaction was also evaluated using the TOC-rating. Therefore three trained raters rated all take-over reactions. Although there were more accidents in the *Pqpd-task* condition, the TOC-rating indicated inferior take-over reactions for the *Quiz-task* condition. However, after looking at the different reactions the drivers fulfilled upon the RtI, it is obvious that more participants from the *Quiz-task* condition reacted via a lane-change maneuver. By doing a lane-change maneuver, suddenly more categories of the TOC-rating become relevant, which do not have to be considered in a braking maneuver. For example, there is the TOC-category *lateral vehicle control* and *securing / communication*. If one decides for a braking maneuver, as 61 % of the test persons of the *Pqpd-task* have shown (compared to 45 % in the *Quiz-task* condition), one will probably not be able to be evaluated negatively in these categories.

As a result, it must be noted that the experiment was not optimally designed for evaluation on the basis of the subjective TOC-rating. If the TOC-rating should be used, only one possible reaction in the take-over situation must be the correct one. As soon as there are two ways to react correctly (in the case of the experiment reported here a *braking* or *lane-change maneuver was possible*), this can affect the rating scores.

As the current study was similar to the study by Jarosch et al. (2017) concerning the two NDRTs, the take-over scenario as well as the driving environment and only differed with respect to the duration of automation, the measured effects may be attributed to the prolonged time of the ride. In the previous study, where the participants had to drive conditionally automated and engage in NDRTs for 25 min, no negative effects on take-over performance have been observed.

One possible explanation for the impaired take-over performance in the monitoring task group (*Pqpd-task*) can be emerged *passive task-related fatigue* due to sustained monitoring as opposed to active reasoning in the *Quiz-task* group. In the monotonous monitoring task condition, PERCLOS increased significantly over time. When people monitor the system for automated driving, similar effects could occur. According to the fatigue model of May and Baldwin (2009), all forms of fatigue in manual driving lead to a higher crash risk and impaired driving performance. Transferred to CDA, this means that all forms of fatigue lead to a deteriorated take-over performance. This becomes particularly relevant when drivers are already tired when they get into their automated vehicle and then switch on the automated driving system.

The measured PERCLOS values after 50 min were quite similar to the PERCLOS values Jarosch et al. (2017) found after 25 min of CDA and engaging in NDRTs (*Pqpd-task*: 11.5 % , *Quiz-task*: 4 %). Therefore it is suspected that the NDRTs also somehow prevented participants from experiencing extreme fatigue levels. In this experiment, no participant fell asleep compared to n = 3 participants that fell asleep after 60 min of automated driving and no task engagement (Feldhütter et al., 2018), or n = 8 sleeping participants in a study conducted by Omae et al. (2006).

When considering the KSS ratings over the course of the experiment, no significant difference between the two NDRTs could be found after 50 min. In the study conducted by Jarosch et al. (2017), significant differences in the self-reported KSS were obvious after an engagement in the same NDRTs of 25 min. Participants that had to engage in the activating *Quiz-task* reported lower levels of fatigue compared to the participants that had to engage in the monotonous monitoring *Pqpd-task*. Similar results were reported by Schömig et al. (2015), who found an activating effect for an engagement in a *Quiz-task* for 15 min. Thus, it can be concluded that any given task can be subjectively perceived as boring or fatiguing after a longer period of time as it was the case in this experiment.

In the current study it was obvious that a time budget of 7 s was unfeasible for a part of the sample after a prolonged conditional automated ride and simultaneously engaging in a NDRT, when a take-over situation suddenly occurred. All in all six drivers lost control of the vehicle due to inadequate steering maneuvers or were not able to avoid an accident. These accidents were more prevalent in the monotonous *Pqpd-task* condition.

A possible explanation for this phenomenon can be a panic reaction caused by the RtI. It is assumed that panic reactions let people respond inadequately. Due to a rush of adrenaline, perception and control of the environment is attenuated (Jamson & Smith, 2003). Such panic reactions either lead to no reaction at all (Muir, Bottomley, & Marrison,1996) or to an overcompensating reaction (Dingus, Jahns, Horowitz, & Knipling, 1998).

What is important to note is that prolonged automated driving led to a deterioration in take-over behavior. Therefore, drivers should increasingly be supported by assistance systems, especially in the case of *short-term* take-over situations after a longer automated driving period. It is also conceivable to be able to support the driver with different manifestations of the HMI in such situations.

How the HMI and driver assistance systems can support the human driver in such take-over situations should be examined in future experiments.

5.5. Limitations

This study was conducted in a motion base driving simulator and focused on measuring human performance when it comes to an RtI after prolonged automated driving. Therefore, several assistance systems like lane keep assistance or emergency braking assistance were deactivated when the RtI was presented in order to measure the reaction times and the quality of the drivers' input. This system behavior is not in accordance with the current status of the CDA system or the system information provided in the instrument cluster and / or the CID at BMW. The two NDRTs were chosen to affect driver fatigue and do not represent a natural behavior.

5.6. Précis

- In a driving simulator experiment, 73 participants experienced prolonged CDA while engaging in either an activating or a fatiguing NDRT.

- The driver state (measured with PERCLOS and KSS) changed over the course of the ride depending on the NDRT.

- After 50 min of automated driving, a take-over situation occurred, and participants had to regain control of the car.
- Results suggest that reaction times and the take-over reaction can be impaired when task-related fatigue (e.g. induced from passive system monitoring) occurs in CDA.

6. Experiment 3 – Effects Of NDRTs In Prolonged Conditional Driving Automation In Real Traffic Environment – A *Wizard-Of-Oz* Approach

Like in Experiment 1 (see section 4.), in the second experiment (see section 5.) the participants had to deal with either a monotonous monitoring task (*Pqpd-task*) or with an activating task (*Quiz-task*). Fatigue did again only emerge in the monotonous monitoring task condition (*Pqpd-task*). PERCLOS could again be identified as a valid measurement for emerging fatigue in driving simulator experiments. KSS on the other hand did not show significant effects between the two NDRT groups. Therefore it is assumed that the participants subjectively feel more fatigued when they have to perform a given task for a prolonged period of time (50 min). Another reason may be that the participants in the Experiment 2 were not able to distinguish between fatigue and boredom in the KSS assessment.

In the second experiment, an impaired take-over performance after 50 min of CDA could be observed. Some participants could not regain control of the vehicle and either crashed into the accident in front of them or lost control of the vehicle due to inadequate steering. This was not the case in the first experiment with an automated driving time of 25 min.

The first and the second experiment were both conducted in the motion base driving simulator. In this environment, the light and climate conditions can also contribute to driver fatigue. Especially when driving for longer periods of time. In addition, the route and the surrounding traffic were designed monotonously to provoke fatigue. Weather conditions have also been designed to fatigue the driver. The sky was grey and cloudy. Summarized, the conditions in the driving simulator were perfect to quickly induce fatigue. In the first two experiments, it could also be proven that PERCLOS is a valid measurement of fatigue in the driving simulator.

However this can look completely different in reality on-road. One question that came up was if people get fatigued comparably quickly in real driving environment than they do in a driving simulator? And what about the PERCLOS measurement in different lighting conditions on-road?

To investigate if the effects that have been found in the driving simulator also occur on-road in real driving environment, a *Wizard-of-Oz* experiment was conducted. The study was based on the second simulator study with the only difference that there was no take-over situation for safety reasons and that the experiment was conducted on-road in real driving environment.

The following section is based on a previous publication:

Jarosch, Paradies, Feiner, & Bengler, 2019

6.1. Introduction

Since some major parts of the introduction and the theoretical part of the original publication would overlap with those of the present thesis, only a brief summary of what was described in these two chapters and where this can be found in the current thesis is given here. All experiment-specific issues such as hypotheses, description of the participants or the exact experimental procedure are presented in detail. For the full paper please refer to the reference.

In order to provide an up-to-date link, initial cross-references were made to previous accidents involving personal injury in the context of automated driving (National Highway Safety Administration, 2017; National Transportation Safety Board, 2018a; National Transportation Safety Board, 2018b). These incidents were also used to give an introduction to the SAE taxonomy (2.2.2.) and the peculiarities of CDA (2.2.3.). In addition, the introduction also dealt with previous studies on the topic of take-over situations in CDA (2.4.2.). Additionally the take-over process model (Marberger et al. 2017, see section 2.3.1.) was explained.

6.2. Theoretical Issues

6.2.1. Wizard-Of-Oz

The method *Wizard-of-Oz* is further introduced in the methods section of this thesis. To avoid repetitions please see section 3.1.2. for detailed information about the method and the *Wizard-of-Oz* vehicle that was used in this experiment.

6.2.2. Fatigue In Manual Driving

See section 2.4.2. for effects of fatigue in manual driving.

6.2.3. Fatigue In CDA

Due to a detailed theory section on this topic, the effects of fatigue in CDA are no longer explicitly discussed here. See section 2.4.3. for detailed information.

6.2.4. Assessment And Measurement Of Fatigue

Again, subjective KSS (for further information see section 3.3.3.) and objective PERCLOS (for further information see section 3.3.2.) were used for the assessment of emerging fatigue in the CDA ride on-road.

6.2.5. Hypotheses

The aim of the study was to investigate the development of fatigue in CDA due to a monotonous monitoring task (*Pqpd-task*) compared to a *Free-choice-activity* group in real road driving conditions. The main focus here was on whether the results from the driving simulator experiments can be transferred to reality. We hypothesized that (i) an engagement in a monotonous monitoring task leads to increased subjective and objective fatigue compared to a *Free-choice-activity* and that (ii) fatigue can be prevented when participants can freely choose their activity during automated driving.

6.3. Materials And Methods

6.3.1. Participants

In the *Wizard-of-Oz* experiment N = 42 employees of the BMW group voluntarily participated. The sample consisted of 8 female (19.05 %) and 34 male (80.95 %) participants. The mean age of the participants was 34.98 yrs (SD = 10.83 yrs, minimum = 20 yrs, maximum = 63 yrs). The participants were drivers with a mean driving experience of 17.00 yrs (SD = 10.54 yrs). Most participants had already experienced at least one driver assistance system. ACC was known to n = 32 (76.19 %), lane-keeping assistant to n = 30 (71.43 %) and traffic jam assistant to n = 23 (54.76 %) of the 42 participants.

6.3.2. Apparatus

The study was conducted in a *Wizard-of-Oz* vehicle. Therefore, a BMW X5 was modified and a second driver's workplace was installed in the rear of the vehicle. The second driver's seat for the *Wizard* was placed in the center of the rear of the vehicle and was equipped with a second steering-wheel, pedals and all other instruments that are necessary for safe driving. The *Wizard* was able to secure the driving environment via a camera-monitor system (for further details about the *Wizard-of-Oz* vehicle see section 3.1.2.).

The participant who is acoustically and visually shielded from the *Wizard* through a privacy glass can activate and deactivate the automated system (= the *Wizard*) via two buttons on the steering-wheel. A Dikablis 3.0 head-mounted eye-tracker was used for measuring PERCLOS. The NDRTs were presented on a Windows Surface tablet (10.8 in) mounted in front of the CID.

6.3.3. Experimental Setup

The basic conditions of the experiment were preset due to exemption approvals of the legislator. Experiments with the *Wizard-of-Oz* vehicle were only legally permitted on the Autobahn section A92 between Munich Airport and Plattling. Also the time periods for the experimental rides were regulated legally. From Monday till Wednesday 9:00 am – 5:00 pm, Thursday 9:00 am – 3:00 pm and Friday 9:00 am – 12:00 am on-road testing on the Autobahn was possible.

Due to these limitations, two participants per day from Monday to Thursday (starting from 8:30 am and 12:30 pm) and one participant on Friday (8:30 am) could participate in the experiment. Each trial took about 2:30 hrs.

To make the whole experiment as realistic as possible the participants were not informed about the *fake*-CDA system and the *Wizard*-driver in the rear of the vehicle until the end of the experimental ride.

At first, when the participants arrived, they were instructed as if they were driving in an automated vehicle for the first time. On arrival, the participants received a written description of the CDA system (the original text was written in German):

The system can be activated when you drive with a maximum of 110 km / h (68 mph) on the right lane on the highway. After activation, the system executes longitudinal and lateral control of the vehicle. This means you do not have to accelerate, brake or steer. The system perceives surrounding traffic and can perform lane-changes or braking maneuvers if required. As this still is a prototype system, the speed is preset to 110 km / h (68 mph). When the system detects safety critical or unknown situations (e.g., exit from the highway, passage of certain construction sites), a request to intervene (RtI) is issued and you have to take over control of the vehicle. These RtIs are issued with a certain amount *of time in advance before you actually have to regain control, giving you enough time to be prepared.*

Participants were also told, that in CDA, the engagement in NDRTs is possible and that the engagement in NDRTs was part of the experiment.

After the introduction participants had to fill out different documents including a confidentiality statement and a demographic form including questions regarding their age, gender, driving experience and their trust in automated systems.

In the meantime, the *Wizard*-driver prepared the vehicle. He started the engine and the camera-monitor system in the rear. The vehicle was parked backwards in an underground car park of the BMW research department.

The examiner brought the participant to the vehicle and seated them on the driver's seat. The backdoors were locked at this moment so it was impossible for the participant to detect the *Wizard*.

When the participant sat in the vehicle, the Dikablis Professional head-mounted eye-tracker was calibrated. Another introduction to the function of the system followed. It was explained, how the *automated driving function* can be activated and how an RtI looks like. Still standing, the participant then had to activate the CDA system for two times and also had to regain control for two times after a RtI was issued. After that part, the NDRT was explained. Depending on which group the participant was assigned to, either the *Pqpd-task* or the possibility of a freely selectable activity (*Free-choice-activity*) was explained.

After the introduction, the participant had to drive the vehicle out of the garage. On a rural road with little traffic, the control-transition process to the system and back to the participant was again practiced for two times.

After this part was completed and there were no further questions of the participant, the participant had to drive the vehicle in the direction to the Autobahn section approved for the automated ride.

Altogether after about 1 hr, the participant reached the section approved for the automated ride on the Autobahn. When the section was reached the examiner told the participant to turn on the system for automated driving. When the system (= the *Wizard*) then controlled the vehicle for the first time, a familiarization period followed for about 10 min. During this phase of the experiment, the participant had again time for asking questions about the system to the examiner and to get familiar with the automated vehicle. After the 10 min, the participants again had to regain control of the vehicle.

Now the real experimental ride started. The participants were told to directly switch on the system for automated driving and then engage into the NDRT *Pqpd-task* or the *Free-choice-activity* for the entire ride. In the case of a RtI, issued either by the examiner or the *Wizard* driver, participants had to regain control of the vehicle again.

During the automated ride, fatigue was assessed for defined points of time. Subjective fatigue of the participants was assessed by the KSS and objective fatigue was assessed by measuring PERCLOS. After about half an hr, the participants had to regain control of the vehicle as they had to leave the Autobahn for a U-turn to start the return trip. The system (= the *Wizard*) could then be reactivated immediately. After about 1 hr of automated driving in total, the experimental ride was over. Participants then again had to regain control and drive the vehicle back to the BMW research department in Garching.

6.3.4. NDRTs

NDRTs were used to affect the fatigue state of the drivers. To investigate effects of different NDRTs on the drivers' fatigue state, participants were assigned to two different NDRT groups: in one group (*Pqpd-task*, see section 3.2.1.), participants had to deal with a monotonous monitoring task, which should induce *passive task-related fatigue* (May & Baldwin, 2009). This task was used in experiments before by Jarosch et al. (2017) and Jarosch et al. (2019). In these experiments, fatigue was significantly higher when the participants were assigned to the *Pqpd-task* compared to a *Quiz-task* group.

In the other NDRT group, participants were not told to engage in a preset NDRT. They could perform any activity they wanted during the automated ride. Various games were available on the tablet PC mounted in front of the CID, and in their invitation to the experiment, these participants were told to bring along all the activities they would like to do while driving in a train or flying in an airplane. In this group, monotony while driving automated should be prevented.

Thus, in one group monotony while driving should be provoked (*Pqpd-task*) and in the other group monotony should be avoided (*Free-choice-activity*) to see if and how fast *passive task-related fatigue* also occurs on-road in real traffic environment. Another focus was to investigate if fatigue-detection methods are reliable under real driving conditions and on-road on the Autobahn, too.

6.3.5. Specifics Of CDA / The *Wizard*

The system for automated driving (i.e. the *Wizard*) could be activated by pressing a button on the steering-wheel. By pressing the button, the *Wizard* in the back was signaled to take over control of the vehicle. After the *Wizard* confirmed that he took over vehicle control, the system status displayed in the main instrument cluster changed to *automated driving* mode. In this moment, the participants had to take their hands off the steering-wheel and their feet from the pedals. For safety reasons, in this driving mode steering from the driver's seat was only possible with a very high effort due to a safety system in order to prevent operating errors of the participants.

The *Wizard*-driver drove with a passive driving style at a speed of about 110 km / h (68 mph). He took over longitudinal and lateral control and could overtake slower driving vehicles.

The system (the *Wizard*) could be deactivated by pressing the button on the steering-wheel again. In this case the *Wizard* was displayed a light signal and then could release the system-control to the human driver in front. In safety-critical situations, the *Wizard* could release an RtI and the human driver then had to regain control within a 45 s cascade.

6.3.6. Human Machine Interface

Due to technical difficulties, the HMI design differed from the driving simulator experiments. The state of the CDA system was displayed in the vehicle's main instrument cluster. Three different states could be displayed: when conditional automation was available, a text notice *Autobahnassistent available* was shown. With activation of the system by the button on the steering-wheel the *Wizard* took over control and the screen changed to a reduced mode, showing the ego-car in a lane, the current speed and an icon indicating that automated driving is active. In case of an RtI, *Autobahnassistent ends. Please drive yourself.* and an expiring timer for 45 s was displayed. Simultaneously an auditory signal was presented. Also the experimenter could trigger RtIs and give instructions to the *Wizard* (continuous speech connection from the front to the rear) or the participant.

6.3.7. Scenario

The experimental rides were conducted on a two-lane highway with a hard shoulder in southern Germany from 20th April till 30th May 2018. With a special permit, the rides could be conducted on the Autobahn A92 from Munich airport to Pilsting and back to Munich airport. Each participant drove automated (with the *Wizard*) for about 1 hr. The rides were conducted outside rush hours to avoid traffic jams and high traffic densities.

After about 30 min of automated driving, participants had to regain control of the vehicle to leave the Autobahn and to do a U-turn. Another 30 min period of automated driving followed. Again the participants had to engage in the assigned NDRT (*Pqpd-task*) or the *Free-choice-activity*. In traffic jam situations and construction zones, the Wizard was not permitted to drive. Thus, in such situations, the participants had to regain control for safety reasons. In Figure 29, the section of the Autobahn A92 that was used for the experimental ride can be seen.

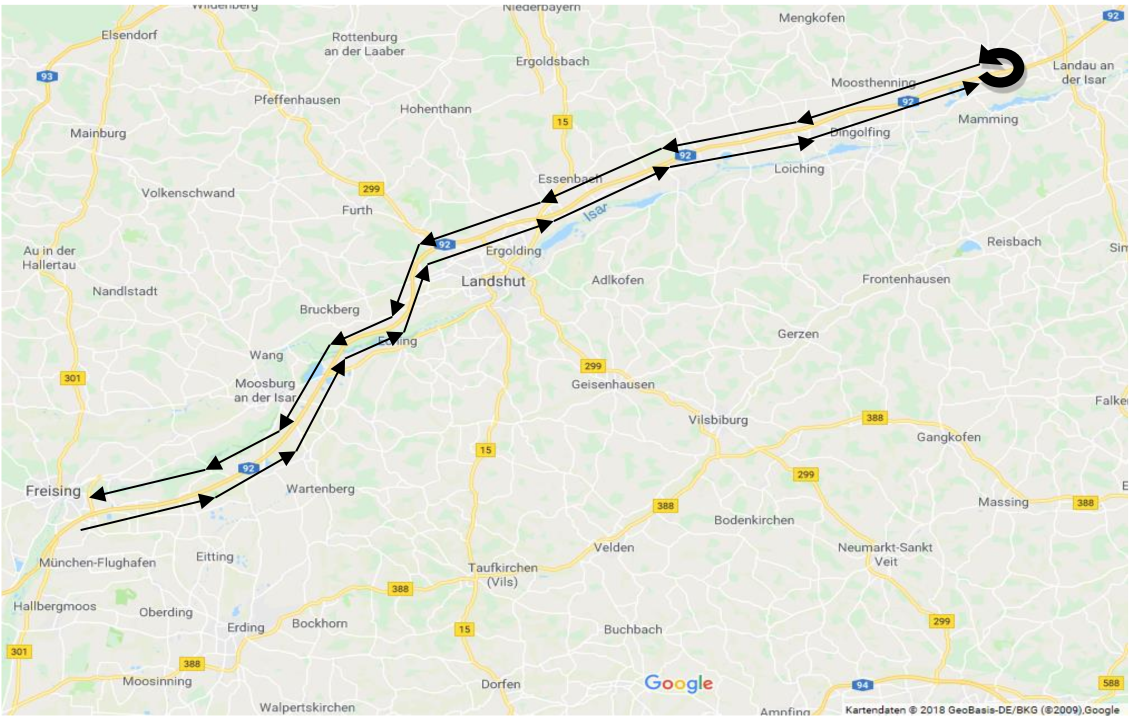


Figure 29: Experiment 3: Section on the Autobahn A92 used for the experiment. Mapdata © 2018 GeoBasis-DE/BKG (©2009), Google.

6.3.8. Dependent Variables

In this experiment the influence of a monotonous monitoring task (*Pqpd-task*) compared to a *Free-choice-activity* on the fatigue state of the driver whilst automated driving should be assessed. Additionally, measurements for fatigue-detection should be examined regarding their suitability in real traffic environment. After the rides, the participants had to answer further questionnaires about their subjective sensations regarding the CDA ride and the NDRT-engagement.

6.3.8.1. Effects of NDRTs on the Drivers' Fatigue State

For assessment of emerging fatigue, subjective KSS (Åkerstedt & Gillberg, 1990) and objective PERCLOS (Wierwille, 1994) were recorded over the course of the experiment for five defined points of time during the experimental ride. To not affect the 1 min interval PERCLOS measurement, KSS was verbally assessed after the PERCLOS measurements. In an experiment conducted by Schmidt et al. (2011), a 1 min verbal communication did not affect objective fatigue in the long term. In the experiment by Schmidt et al. (2011) measurements of fatigue returned to their pre-communication level after 2 min. See Table 17 for the times of measurement in detail.

Table 17: Experiment 3: Procedure of the experimental ride (� PERCLOS (1 min), afterwards KSS).

t1	t2	t3	U-turn	t4	t5 (end)
Min 2-3	15 – 16	29 – 30	31 - 32	45 – 46	59 - 60
�	�	�		�	�

6.3.8.2. Effects of NDRT Engagement in CDA on Subjective Sensation

Participants were asked if and to what extent they would trust the system and deal with NDRTs in CDA. This question was asked for two times, before and after the CDA ride. A final questionnaire dealt with questions about the subjective effects of NDRTs in automated driving on the *perceived ability to drive*, *perceived monotony*, *perceived exhaustion, perceived fatigue* and *the incentive to deal with NDRTs.* All these questions could be answered on a seven-point Likert scale.

6.4. Results

The main objective of the study was to examine how engagement in a monotonous monitoring task (*Pqpd-task*) affects the fatigue state of the driver compared to a *Free-choice-activity* in CDA. To investigate whether the results found in the driving simulator experiments, where fatigue immediately emerged when driving automated, also occur on-road in real traffic environment, this experiment was conducted in real traffic environment on the Autobahn.

6.4.1. Effects Of NDRT Engagement On The Fatigue State Of The Drivers

To examine the effects of CDA and simultaneously engaging in either a monotonous monitoring task or the ability to freely choose the current activity while driving automated, objective PERCLOS and subjective KSS were evaluated using mixed ANOVAs. The factor time of measurement was used as the within-subjects factor and the NDRT / activity was used as the between-subjects factor. For violations of sphericity, Greenhouse-Geiser correction was used. The drivers' fatigue state dependent on the NDRTs was evaluated for five times of measurement (t1, t2, t3, t4 and t5).

6.4.1.1. Effects on PERCLOS

PERCLOS was highest in the *Pqpd-task* group at t5 (M = 20.53, SD = 21.11) and lowest in the *Free-choice-activity* group at t2 (M = 5.47, SD = 3.55).

A mixed ANOVA showed that there was a statistically significant main effect for the factor time of measurement on PERCLOS, F (1.47, 57.49) = 5.54, p = .012, *partial* η^2 = .12. There was also a significant effect of the NDRT on PERCLOS, (F (1, 39) = 15.55, p < .001, η^2 = .29. Next to the main effects of the task and the time of the measurement, there was also a significant interaction effect between these two factors, F (1.47, 57.49) = 5.57, p = .012, *partial* η^2 = .13). See Figure 30 for further information.

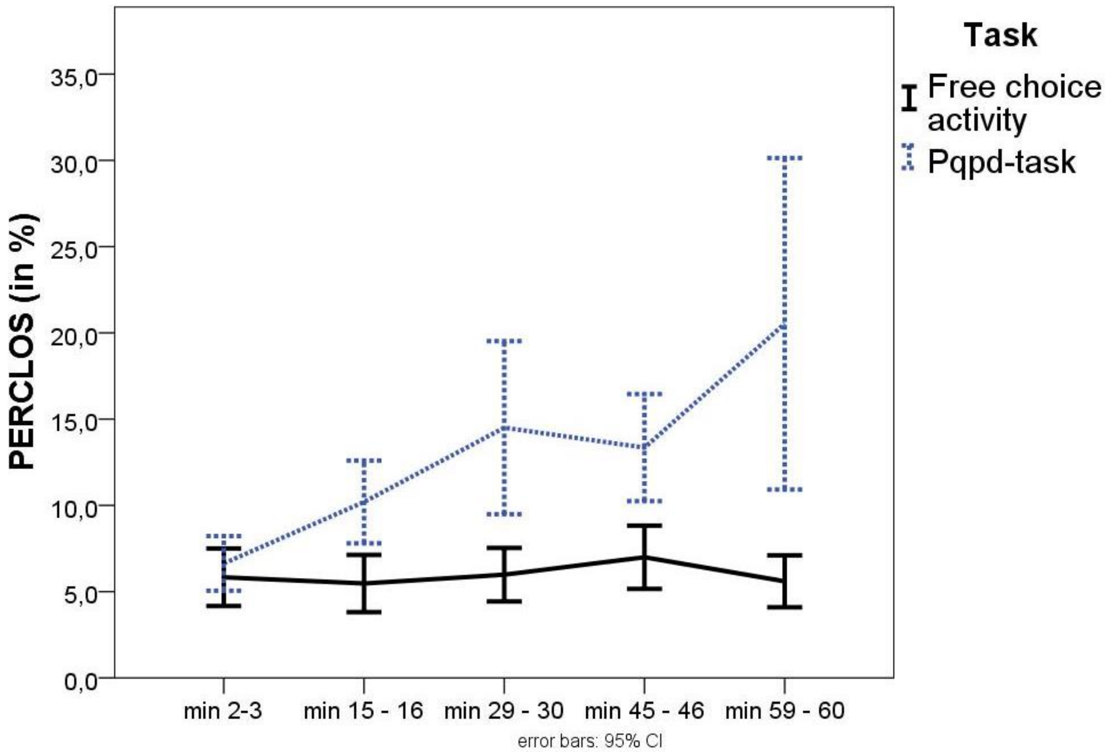


Figure 30: Experiment 3: Effects of NDRTs on PERCLOS over the course of the ride.

6.4.1.2. Effects on KSS

KSS was highest in the *Pqpd-task* group at t5 (M = 6.86, SD = 1.36) and lowest in the *Free-choice-activity* group at t1 (M = 3.45, SD = 1.4).

A mixed ANOVA showed that there was a statistically significant main effect for the factor time of measurement on KSS, (F (2.32, 92.71) = 30.42, p < .001, *partial* η^2 = .43). There was also a significant main effect of the factor NDRT on KSS, (F (1, 40) = 21.97, p < .001, η^2 = .93. Next to the main effects, there was also a significant interaction effect between the time of measurement and the NDRTs, F (2.32, 92.71) = 11.36, p < .001, *partial* η^2 = .22). See Figure 31 for further information concerning the KSS in the *Wizard-of Oz* experiment.

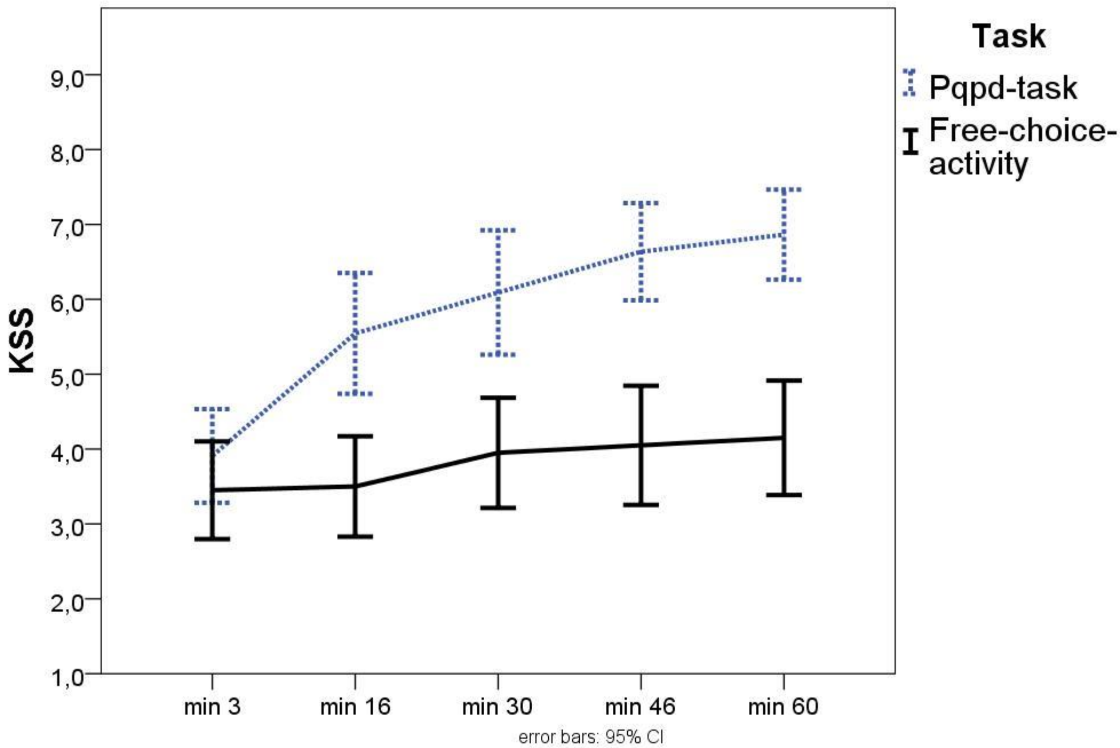


Figure 31: Experiment 3: Effects of NDRTs on KSS.

6.4.2. Subjective Assessments Of The Participants Regarding NDRTs And CDA

To investigate how task engagement in either a monotonous monitoring task (*Pqpd-task*) or a *Free-choice-activity* while driving conditional automated affects trust, subjective fatigue and monotony, a questionnaire was handed out before and after the ride.

6.4.2.1. Effects On Trust In Automation

Both before and after the ride, subjective trust ratings did not significantly differ amongst the two task groups, monotonous monitoring task (*Pqpd-task*) and *Free-choice-activity*. Before the ride, participants from the *Pqpd-task* rated trust a bit higher (*Md* = 6.00, on a 7-point Likert scale, high values indicate high trust) compared to the *Free-choice-activity* group (*Md* = 5.00), Mann –Whitney-*U*-Test: *U* = 220.00, *p* = .99. Also after the ride, participants from the *Pqpd-task* rated trust a bit higher (*Md* = 7.00) compared to the *Free-choice-activity* group (*Md* = 6.00), Mann –Whitney-*U*-Test: *U* = 175.50, *p* = .21.

6.4.2.2. Effects On Likelihood To Engage In NDRTs During CDA

The simple question *When CDA is active I would like to deal with NDRTs* was assessed before and after the ride. Before the ride, both groups indicated to engage in NDRTs in a similar way. The *Pqpd-task* group indicated a bit higher likelihood for task-engagement (*Md* = 5.00) compared to the *Free-choice-activity* group (*Md* = 4.00), Mann –Whitney-*U*-Test: *U* = 212.50, *p* = .84. Also after the ride, the *Pqpd-task* group (*Md* = 6.00) rated their willingness to engage in NDRTs a bit higher compared to the *Free-choice-activity* group (*Md* = 6.00), Mann–Whitney-*U*-Test: *U* = 166.00, *p* = .141.

6.4.2.3. Subjective Assessment Of Effects Of NDRTs During CDA On The Driver

After the ride, another questionnaire was handed out to the participants. The questionnaire was used to investigate how the drivers subjectively perceived how the NDRTs affected their state while driving automated. The results can be seen in Table 18.

Table 18: Experiment 3: Effects of NDRTs on subjective driver state.

	Question (1 = strongly disagree; 7 = strongly agree)	Median Pqpd	Median Free-choice-activity	Mann-Whitney-*U*	*p*
**	The automated ride was so monotonous that it was hard for me to stay awake.	6.00	2.00	26.50	< .001
**	The automated ride was exhausting in the long run.	5.00	3.00	94.50	= .001
	In the long term, the NDRT would restrict my ability to drive.	5.00	3.00	147.50	= .062
**	The NDRT I have worked on during the ride kept me awake.	4.00	1.00	42.50	< .001

Note. * *p* < .050, ** *p* < .010

6.5. Discussion

The main objective of this *Wizard-of-Oz* on-road experiment was to investigate if fatigue due to passive task engagement emerges in CDA. Therefore one group of participants had to engage in a monotonous monitoring task (*Pqpd-task*) and a second group of participants could freely choose their non-driving-related activity (*Free-choice-activity*) whilst driving conditionally automated. Another focus of the experiment was to examine if the objective measurement for fatigue, PERCLOS, also can be used in real driving environment. So far, studies that used PERCLOS for measurements of the drivers' fatigue state were only conducted in driving simulator environment with standardized light conditions. As the results from the driving simulators indicate that *passive task-related fatigue* emerges within short time periods of about 20 – 25 min, it should now be investigated if similar findings occur in real traffic environment on-road.

The results of this experiment suggest that fatigue in CDA on-road can emerge as fast as it did in the before mentioned driving simulator experiments. This is especially the case, when the participants were told to engage in a monotonous monitoring task like the *Pqpd-task*, they had to work on while driving automated. On the other hand, a *Free-choice-activity* during CDA seems to have the potential to keep *task-related fatigue* on a relatively low level. Especially when participants can do what they want to do and do not have to monitor the automated vehicle or the traffic environment around them.

That fatigue emerged in the monotonous task condition, in the *Pqpd-task* group, further and faster compared to the *Free-choice-activity*, can be concluded when regarding the course of the PERCLOS measurement over the course of the automated ride. The kind of the NDRT, *Free-choice-activity* vs. a preset monotonous monitoring task (*Pqpd-task*), as well as the time of the measurement significantly affected this objective measurement for fatigue. Also the interaction effect between the NDRT and the time of the measurement significantly affected PERCLOS. This indicates, that especially prolonged engaging in a monotonous monitoring task leads to emerging fatigue. Such a situation may also occur, when driving in a CDA vehicle, observing the system behavior and not engaging in a NDRT at all.

Also the subjective measurement KSS indicates that fatigue can emerge in CDA. In the monotonous monitoring task group, KSS ratings were significantly higher, compared to the *Free-choice-activity* group. Higher values of the KSS are associated with increasing fatigue. With increasing journey time, the subjective fatigue further increased, when the participants had to engage in the monotonous monitoring *Pqpd-task*.

Other than previously expected, fatigue in the on-road driving environment experiment was even higher compared to the results from the driving simulator. PERCLOS in the driving simulator did not exceed 13 % in mean in the driving simulator (Jarosch et al., 2017; Jarosch et al., 2019). In the on-road experiment, a mean value of 20.53 % was measured, when participants had to engage in the monotonous monitoring task for about 1 hr. One possible explanation for this further arise in PERCLOS compared to the driving simulator studies can be the not controllable light conditions on-road. A direct sunlight from the front could lead to a higher blink-frequency of the participants. One indicator for this assumption is that the participants from the *Free-choice-activity* group also had higher PERCLOS values than those who have worked on the activating *Quiz-task* in the driving simulator experiments (Jarosch et al., 2017; Jarosch et al., 2019).

After about 30 min of automated driving, participants had to take over control of the vehicle and execute a U-turn maneuver for heading back. Therefore, they had to leave the Autobahn, turn around and then enter the Autobahn in the opposite direction again. All in all this maneuver took about 1 min. After the self-driving maneuver the participants again had to engage in their assigned NDRT (*Free-choice-activity* or *Pqpd-task*). This control transition maneuver can also be seen in the PERCLOS measurement. For t4 (45 – 46 min), the time of measurement after the U-turn, the mean PERCLOS value in the monotonous monitoring task group declined compared to t3 (29 – 30 min), the time of measurement directly before the U-turn. This effect can only be seen when participants had to engage in the monotonous monitoring task. Thus, it can be expected that a short period of manual self-driving can counteract emerging fatigue when confronted with *passive task-related fatigue*. This decline can only be seen in the PERCLOS and not in the subjective KSS. Since the KSS may also measure subjectively perceived boredom during the automated ride, one can assume that PERCLOS reacts more sensitively to such maneuvers.

Other than expected, the measured PERCLOS and KSS values in the monotonous monitoring task group were even higher in the real driving environment compared to the driving simulator experiments. Thus, it could be shown that PERCLOS measurements are sensitive for fatigue detection also in real driving environment on-road.

In further research it now has to be investigated, if the emerging fatigue in on-road driving environment negatively affects take-over performance when it comes to a *short-term* take-over situation. In driving simulator studies such negative effects were obvious, when participants had to engage in monotonous tasks while driving automated for a longer period of time (Jarosch et al., 2019). It then also has to be investigated, how the human driver can best be supported in such take-over situations.

6.6. Limitations

This study was conducted in a *Wizard-of-Oz* vehicle and focused on measuring how the drivers' fatigue state changes in CDA in real driving environment on the Autobahn. To affect the drivers' state, passengers had to deal with a monotonous monitoring task (*Pqpd-task*) or had free choice of their activity (*Free-choice-activity*) whilst driving automated. Especially the monotonous monitoring task represents an activity which drivers will not do under normal circumstances. However, it can also be expected, that a driver who has been fatigued before, drives an automated vehicle. Especially this driver will want to use the automated driving system.

6.7. Précis

- In a *Wizard-of-Oz* experiment, N = 42 participants experienced prolonged CDA while engaging in either an activating or a fatiguing NDRT on-road in real traffic environment.

- The drivers' fatigue state (measured with PERCLOS and KSS) changed over the course of the ride depending on the NDRT.

- Participants that had to deal with a monotonous monitoring task were significant more fatigued compared to participants that could freely choose their activity after 60 min.

- The *Wizard-of-Oz* approach worked well and not one participant noticed the human *Wizard* driver.

7. Experiment 4 – How To Support The Driver In Short-Term Take-Over Situations?

After the third experiment (see section 6. for further details), which was conducted in the *Wizard-of-Oz* vehicle in real traffic environment, it is obvious that *passive task-related fatigue* due to monotonous monitoring tasks (*Pqpd-task*) while driving conditionally automated cannot only occur in the driving simulator but as well in on-road environment.

In the second experiment (see section 5.), an impaired take-over performance could be observed after 50 min of automated driving and simultaneously engaging in a NDRT. These effects were most notably, when participants had to engage in a monotonous monitoring task (*Pqpd-task*). Significant differences in their take-over performance compared to participants that had to engage in an activating task (*Quiz-task*) were a reduced eyes-on-road time, t_{eyes}, as well as a slower first braking reaction, $t_{brake\_reaction}$. Next to these impaired reaction times, also the take-over behavior was impaired. Six participants could not control the vehicle in the take-over situation or crashed into the accident. Four of them had to engage in the passive and monotonous monitoring *Pqpd-task*.

When considering the results of Experiment 2 and those of the Experiment 3 (in the *Wizard-of-Oz* vehicle), it can be concluded that the fatigue that arose in both experiments in the passive task condition also leads to similarly deteriorated take-over performance in real traffic environment on-road. Deteriorated take-over performances were measurable in the Experiment 2 in the driving-simulator. In Experiment 3, which was conducted in real on-road environment, take-over situations were avoided in order to not endanger the participants.

In order to improve the take-over performance in *short-term* take-over situations, in the fourth experiment it should be investigated how the human driver can be best supported in such *short-term* take-over situations with the help of different HMI designs.

7.1. Theoretical Issues

This experiment should investigate whether certain HMI components can be used to support the driver in a *short-term* take-over situation. A HMI concept in CDA has to fulfill several tasks:

- Display the current system status to the human driver.
- Warn the driver in case of a system initiated take-over situation (RtI).
- Support the driver in a specific take-over situation.

However, in previous experiments it could be seen, that especially fatigued drivers had problems to fully regain control of the vehicle when they drove conditionally automated for a prolonged time and simultaneously had to fulfill a passive NDRT. A significant impaired eyes-on-road reaction time as well as a significant impaired first braking reaction time was measured. Additionally, it seemed like some of the participants reacted quickly at first (e.g. hands-on reaction or eyes-on-road reaction), but then went into a kind of state of shock. Some of the participants seemed to be paralyzed and showed no further reactions. Others overreacted and the steering-wheel was torn by over 90°, losing control of the vehicle.

The focus of this experiment now was to investigate how to support the driver in such *short-term* take-over situations. For this purpose, two different HMI concepts were developed in expert workshops. On the one hand a concept was developed, which should support the driver regarding the eyes-on-road reaction (*LED-group*), and one, which should warn the driver by a speech output of the forthcoming situation (*SPEECH-group*). These two concepts were tested against a baseline concept, a signal tone accompanied by a visual representation in the instrument cluster (*BASELINE-group*). In the between-subjects experiment, each driver experienced one of the three HMI concepts in two different take-over situations. As especially fatigued drivers seemed having problems to react adequately upon an RtI, in this experiment all participants had to engage in the *Pqpd-task* which was successfully used to provoke fatigue in the experiments before.

7.1.1. HMI Concepts In CDA

In previous experiments a number of different HMI concepts have already been tested. Especially in experiments in which the take-over performance was examined in a critical situation, a standard variant of this concept has become established which warns the driver redundantly both acoustically and visually. However, also further attempts have been tested so far. These include, among other concepts, the promising possibilities for speech outputs to warn the driver (see section 7.1.2.) and warning the driver with peripheral light signals (see section 7.1.3.).

7.1.2. Speech Output In CDA

When driving manually, speech-based systems have been shown to be advantageous for driving performance. These include for example lower lane variation and a more consistent speed (Barón & Green, 2006; Stanton & Edworthy, 1999). Also in CDA, experiments including speech-outputs have been conducted before. In an experiment conducted by Naujoks, Forster, Wiedemann, and Neukum (2016) it was investigated whether a communication of upcoming manoeuvers of the automated system positively affects human-machine collaboration. As a results it can be said that a speech output lead to a decreased visual workload and reduced interference with the task the driver is dealing with whilst driving automated. In this experiment participants assessed the speech-output as useful. In another experiment Walch, Lange, Baumann, and Weber (2015) investigated the efficiency and subjective sensation of three different HMI designs, one including a speech-output. In this experiment participants favored the combination of a speech-output which gave the reason for the take-over (e.g. *caution fog*) in combination with the RtI. Whether speech-output has advantages over other HMI designs has also already been investigated in take-over situations. In an experiment conducted by Forster, Naujoks, Neukum, and Huestegge (2017), the test participants had a time budget of 20 s to take over control of the vehicle. In this experiment, faster reaction times could be achieved with the speech-output HMI compared to a generic HMI design.

However, amongst others Forster et al. (2017) suggest, that a speech output can be beneficial when a larger temporal window is left to react upon the RtI but whether a speech-output based RtI is still advantageous in *short-term* take-over situations has not been examined yet.

7.1.3. RtIs Including Peripheral Visual Components In CDA

Next to the before described speech-output based HMIs, also light signals that support the driver in the specific take-over situation have been investigated before. A standard RtI, like used in most experiments including take-over situations in CDA before, consists of an auditory warning signal in combination with a visual notification in the instrument cluster. Such a RtI informs the driver that he has to take over control but does not support him in the following take-over process (e.g. decision making). Thus, in case of a RtI, the driver first has to scan the scenery to detect the obstacle / the situation that led to the RtI. Therefore, HMI concepts that draw the driver’s attention towards the relevant elements that caused the RtI can be advantageous in such a situation. Similar approaches were already investigated in the context of research on ACC systems. In an experiment conducted by Stanton, Dunoyer, and Leatherland (2011) different HMI concepts including different amount of information about the surrounding traffic situation have been investigated. In this experiment, the display with the most information helped to understand the system. However, it was also experienced to be the most demanding one compared to the other concepts. In another experiment, a HMI concept including augmented reality components (e.g. arrows, carpets and circles marking an object on the road) was used to suggest the driver how to react in a specific driving situation in which cooperation with other road users was necessary, appropriately. Beneficial results of this AR concept could be shown for cooperative behavior of the participants and increased safety (Zimmermann, Bauer, Lutteken, Rothkirch, & Bengler, 2014).

The efficiency of different HMI designs in take-over situations (i.e. vibrotactile warning / a vibrotactile warning + augmented sphere highlighting an obstacle / vibrotactile warning + augmented-reality overlay / vibrotactile warning + augmented reality arrows) was investigated by Eriksson et al. (2019). Results of the experiment suggest that the different HMI designs did not significantly affect the initial reactions of the drivers upon the RtI. In the specific scenarios the drivers had 12 s to react appropriately upon the RtI.

However, in the experiments conducted as part of this thesis, an impaired reaction time, especially for the first eyes-on-road reaction, was obvious when the drivers were fatigued. Therefore, in this last experiment, the research focus has been to investigate how the take-over reaction of the human drivers in CDA can be improved and how to best support the human driver in such situations with the help of different HMI components.

7.2. Materials And Methods

7.2.1. Participants

N = 64 employees of the BMW group voluntarily participated in the third driving-simulator experiment. The sample consisted of 14 female (21.88 %) and 50 male (78.12 %) participants. The mean age of the participants was 33.4 yrs (SD = 12.00 yrs, minimum = 20 yrs, maximum = 65 yrs). The subjects were drivers with a mean driving experience of 14.94 yrs (SD = 10.95 yrs). Most participants had already experienced at least one driver assistance system. ACC was known to *n* = 40 (62.50 %), lane-keeping assistant to *n* = 33 (51.56 %) and traffic jam assistant to *n* = 18 (28.13 %) of the 64 participants.

7.2.2. Apparatus

As well as Experiment 1 and Experiment 2, the fourth experiment was again conducted in the motion base driving simulator at BMW facilities in Munich. The same mock-up was used like in the other two driving-simulator experiments. See section 3.1.1. for further information regarding the driving simulator.

Again a Dikablis Professional head-mounted eye-tracker was used for PERCLOS measurement for fatigue assessment. The NDRT that was used to provoke fatigue (*Pqpd-task*) was presented on a Windows Surface tablet (10.8 in) mounted in front of the CID.

Additionally a LED-light band was installed in the exterior of the vehicle in front of the root of the windshield to display the light signal in the *LED-group* (see Figure 32).

The speech-output, which was played in the *SPEECH-group* when the RtI was triggered, was previously recorded by a professional speaker and played in the interior of the vehicle via the sound-system.

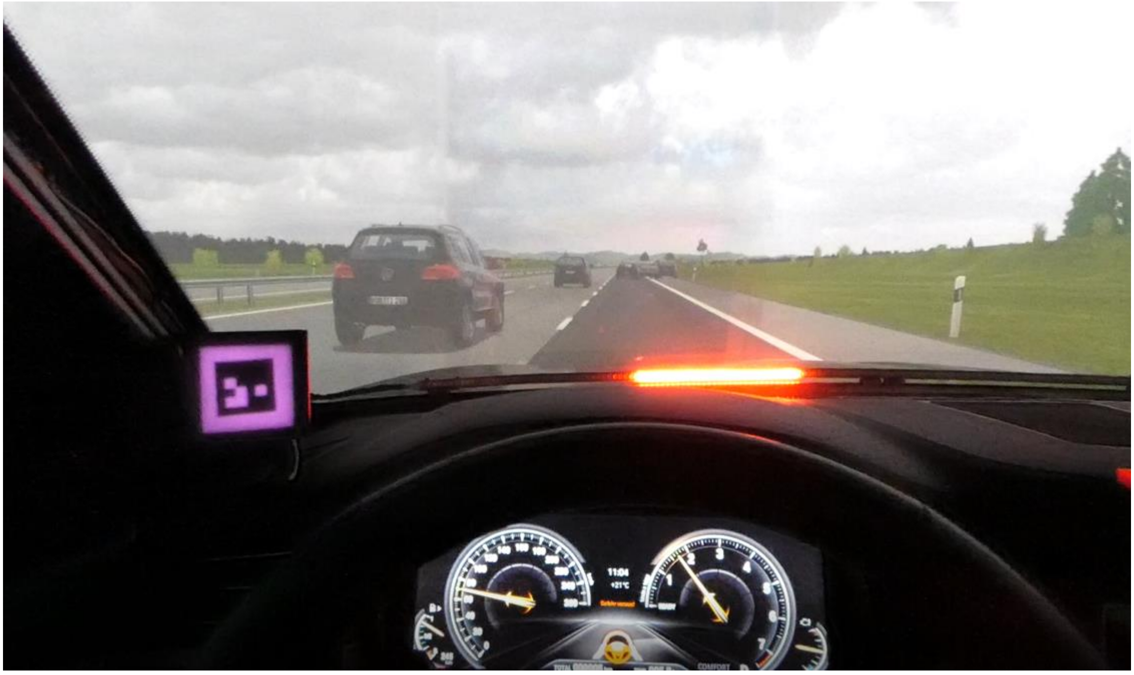


Figure 32: Experiment 4: LED light signal in the *LED-concept.*

7.2.3. Experimental Setup

During the experiment five participants were tested per day (8:15 am, 10:15 am, 1:00 pm, 3:00 pm and 5:00 pm). Each trial took about 1:45 hrs. Upon arrival, participants were handed out a description of the CDA system that was used in the driving simulation. They were informed about the system boundaries and the possibility of RtIs in CDA as well as the characteristics of the system. After that, each participant had to sign a confidentiality statement and had to fill out a demographic form with questions about driving experience, age, gender and experience with driver assistance systems experienced before. Before the participants could test drive the simulator in a familiarization ride, the head-mounted eye-tracker was calibrated.

A familiarization ride to accustom all participants with the driving simulator followed (Hergeth et al., 2017). This was necessary to achieve an equal level of experience with driving-simulators as some participants had experienced driving-simulator experiments before. In this part of the experiment, the test persons first experienced a manual driving section in order to get used to the driving characteristics of the simulator. After a few min of manually driving the test persons were asked to activate the system for automated driving by pushing a yellow marked button on the steering-wheel. The examiner explained that after activation of the system, the automated driving system executes longitudinal and lateral guidance, can overtake slower vehicles and regulates the speed on vehicles driving in front. In this test session, the examiner also explained that RtIs and take-over situations can occur. After this explanation, an RtI was issued by the examiner and participants had to regain control of the automated driving vehicle. The RtI the test persons experienced in the test session was already adapted to their later experimental group (*BASELINE-group*, *LED-group* or *SPEECH-group*; see section 7.2.6. for further details). When the RtI was issued, participants had to deactivate the CDA system either by braking, by steering or by pushing the button on the steering-wheel. In the familiarization session, the participants had to deactivate the system with each of the three possibilities once.

After the training session the experimental ride followed. Since in the previous experiments especially participants who had to deal with a monotonous monitoring task (*Pqpd-task*) during the conditional automated ride had problems with the *short-term* take-over situation (loss of control of the vehicle, crashes and impaired reaction times), the monotonous monitoring task (*Pqpd-task*, see section 7.2.4.) was chosen for all participants. In order to control if fatigue increased during the experimental rides and independently from the different HMI groups, objective PERCLOS (see section 3.3.2.) and subjective KSS (see section 3.3.3.) were assessed for several times (see Table 19) over the course of the experiment. To not affect the PERCLOS measurement, KSS was surveyed after the PERCLOS measurement.

Table 19: Experiment 4: Procedure of the experimental ride (� PERCLOS measurement, afterwards KSS).

t1	t2	t3	t4	t5
Min 2-3	34 – 35	40	52 – 53	58
�	�	***RtI***	�	***RtI***

The focus of this experiment was to examine different HMI-concepts to support the human driver in *short-term* take-over situations (see section 7.2.6. for detailed information about the HMI concepts). In order to investigate the effectiveness of the different concepts, two *short-term* take-over situations in which a fast reaction of the human driver was required, were part of the experiment. The first take-over situation occurred after 40 min and the second after 58 min of the ride. To avoid training effects both take-over situations differed strongly from each other (see section 7.2.7. for details about the two situations). However, in both take-over situations, different reaction times of the drivers upon the RtI as well as metrics indicating the quality of the take-over reactions of the human drivers were recorded. See section 3.4. for further details about the assessment of the take-over reaction.

After the second take-over situation, the participants had to stop on the hard shoulder and the experimental ride was over. After the motion base driving simulator was brought back to its initial position, the examiner picked up the participants and took off the eye-tracker. This was followed by a last survey asking specific questions about the respective HMI concept, the test person experienced in the ride.

7.2.4. NDRT

In this experiment, in contrast to the experiments conducted before, the NDRT was not used as one independent variable. Since it could already be shown in the previous experiments (Experiment 1 and Experiment 2) that especially tired drivers (i.e. due to *passive task-related fatigue*) have problems when taking over vehicle control, all drivers in this experiment should be confronted with the take-over situation in a fatigued state. Therefore, in this experiment, the monotonous monitoring task (*Pqpd-task*) was used to cause *passive task-related fatigue* for all participants. In previous experiments, the task was successfully used to induce fatigue and it could also be shown that the generated fatigue was related to a deterioration in the take-over performance. See section 3.2.1. for further details about the NDRT.

When the RtI was issued, the NDRT stopped and a black-screen was displayed on the Windows Surface PC which was installed in front of the CID. With each activation of the CDA system, the NDRT started again.

7.2.5. Specifics Of CDA

As in Experiment 1 and Experiment 3, the Mock-Up was equipped with a CDA system. Therefore the same system was used as in the driving simulator experiments before. For further details about the CDA system that was used see section 3.1.1.

7.2.6. Human Machine Interfaces

In previous driving simulator experiments, an impaired take-over performance in *short-term* take-over situations after engaging in a NDRT while driving automated was obvious. Significant differences in the reaction times of the drivers could be found at the first glance on the road reaction time (t_{eyes}) and at the first braking reaction time ($t_{brake\_reaction}$) of the human drivers. It could also be shown that especially fatigued drivers reacted impaired compared to less fatigued drivers.

In order to support the (fatigued) drivers in *short-term* take-over situations, different concepts were designed with HMI-experts, which seemed to be promising to improve the take-over performance of the human drivers.

Two concepts were developed which had to compete against the previously used *BASELINE-concept*, in the experiment. A between-subjects design was used for this purpose. In the following section, the HMI concepts are explained in detail.

LED-concept: The main feature of this RtI is a red LED light signal, which should quickly direct the gaze of the driver to the location of the danger as soon as it can be perceived in the periphery. The LED bar lights up red at the moment of the RtI. In order to direct the attention of the driver to the emergency situation, the LED light band flashes in the direction of the accident ahead. Simultaneously an earcon was presented and a red text *Gefahr voraus!* (Eng.: *Danger ahead!*) as well as a red steering-wheel with two hands grabbing it was displayed in the central information cluster. The concept can be seen in Figure 33, left.

SPEECH-concept: The main differentiator of this RtI was a speech-output that was intended to alert the human driver to the take-over situation. In the moment of the RtI, *Selbst fahren! Gefahr voraus!* (Eng.: *Drive yourself! Danger ahead!*) was played via the integrated sound-system of the vehicle, proceeded by a short earcon. The speech-output was recorded by a professional narrator before and pretested against other possible speech-outputs in a small concept experiment. Additionally a red text *Gefahr voraus!* (Eng.: Danger ahead!) and a red steering-wheel icon with two hands grabbing it was displayed in the central information cluster. See Figure 33, middle for this concept.

Baseline – RtI. In the baseline condition, an earcon was presented when the RtI was issued. Simultaneously, a red text *Selbst fahren! Gefahr voraus!* (Eng.: *Drive yourself! Danger ahead!*) and a red steering-wheel icon with two hands grabbing it was displayed in the central information cluster. The Baseline-concept is displayed in Figure 33, right.

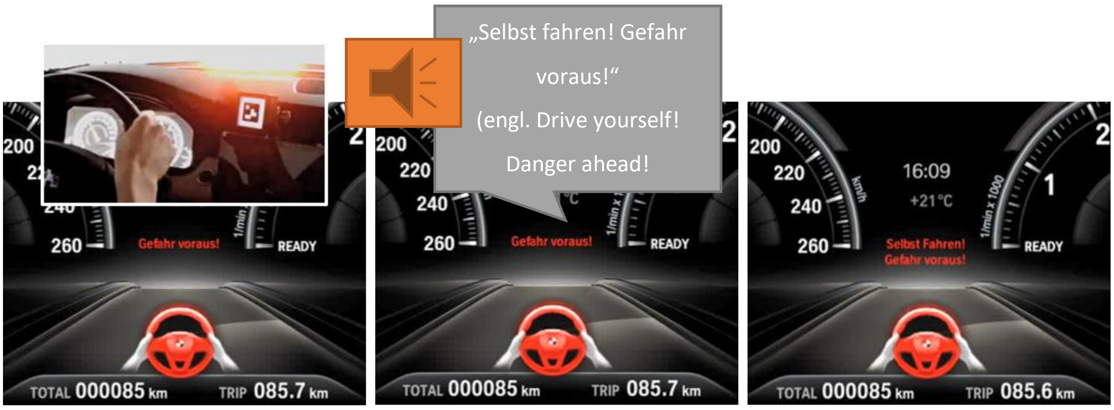


Figure 33: Experiment 4: HMI concepts (left: *LED-concept*, middle: *SPEECH-concept* & right: *Baseline*).

7.2.7. Scenarios

To examine the three HMI concepts regarding their efficiency in take-over situations, each participant experienced two different take-over situations. The first situation appeared after 40 min of the automated ride and the second one after 58 min. To achieve comparable results, each participant should experience the two take-over situations in the same order. Thus, a permutation of the two situations was rejected. Another reason for a no-permutation design of the two situations was that it cannot be assumed that the level of fatigue is the same in both situations, as one occurred after 40 min and the second one after 58 min. This would have meant that the concept-groups could only be viewed separately from each other. In the following, the two scenarios are further explained.

Take-over situation avoidance maneuver: In the first take-over situation that appeared after 40 min of the automated ride the right lane (lane of the ego-vehicle) was blocked by two breakdown vehicles. In the same moment, a bus, which could have been seen in the mirror, tailgated the own vehicle. In the two lanes left of the ego-vehicle no other road users were present. The TTC in the moment of the RtI was 7 s. Due to these circumstances, this situation could in best-case be solved by an avoidance maneuver / lane change maneuver to the lane left, the middle lane. Due to the close approaching bus, which should have been seen when securing by looking into the mirror, a braking maneuver is not suitable in this situation. See Figure 34 for the avoidance maneuver situation.

Figure 34: Experiment 4: Avoidance maneuver.

Take-over situation emergency stop: The second take-over situation happened after 58 min of the start of the automated ride. In this situation, three breakdown vehicles were blocking the lane of the ego-vehicle (right lane). In the moment of the RtI, no other road user is driving in the right lane behind the ego-vehicle. The lane left of the ego-vehicle (middle-lane) is blocked by a convoy of vehicles that drive close to each other (distance between the vehicles was about 15 m). The TTC to the accident in front in the moment of the RtI was 7 s again. Thus, in this situation, a stopping maneuver was the dedicate reaction of the drivers. Due to the convoy of vehicles in the middle lane, a lane change maneuver was not possible without endangering oneself or other road users. See Figure 35 for the braking maneuver situation.

Figure 35: Experiment 4: Braking maneuver.

7.2.8. Dependent Variables

Effects on the drivers' state: In order to control the drivers' fatigue states PERCLOS was recorded with the Dikablis Professional head-mounted eye-tracker. Additionally the subjective KSS was examined. See section 3.3. for information about these two fatigue measurements.

Take-over reactions. As the main focus of this experiment was to examine the effectiveness of different HMI concepts, take-over reaction measures were recorded in both take-over situations. Like in the first two driving simulator experiments, quality-based metrics as well as reaction times upon the RtI were recorded. See section 3.4. for details about the different metrics concerning the quality of the take-over reaction.

7.3. Results

7.3.1. Results On The Drivers' State

In order to assess if the monotonous monitoring *Pqpd-task*, emerged the fatigue state of the participants, PERCLOS and KSS were assessed for three times during the experimental ride. The first baseline-measurement was conducted in the beginning of the ride. The following measurements each took place 5 min before the take-over situations. It was also investigated whether the fatigue levels in the first and second take-over situation were similar, and whether the fatigue levels differed according to the three HMI groups.

7.3.1.1. Subjective Fatigue Assessment - KSS

KSS was highest in the *SPEECH-concept* group at t2 (M = 6.96, SD = 1.73) and lowest in the *LED-concept* group at t1 (M = 3.62, SD = 1.47).

A mixed ANOVA with Greenhouse-Geisser correction revealed that there was a statistically significant main effect for the factor time of measurement on KSS, F (1.97, 118.11) = 103.21, p < .001, *partial* η^2 = .63. As expected, there was no significant effect of the HMI concept on the KSS, F (2, 60) = 1.13, p = .33, η^2 = .04. There was also no significant interaction effect between the time of measurement and the HMI concept, F (3.94, 118.11) = .51, p = .73, *partial* η^2 = .02. See Figure 36 for the course of the KSS for the three different HMI concepts.

Bonferroni-adjusted post-hoc analysis revealed a significant difference of the KSS between t1 and t2 (p < .001; -2.52, 95 %-CI [-2.98, -2.05]) as well as between t1 and t3 (p < .001; -2.13, 95 %-CI [-2.62, -1.64]). There was no significant difference of the KSS between t2 and t3. (p = .099; 0.39, 95 %-CI [-0.83, 0.05]).

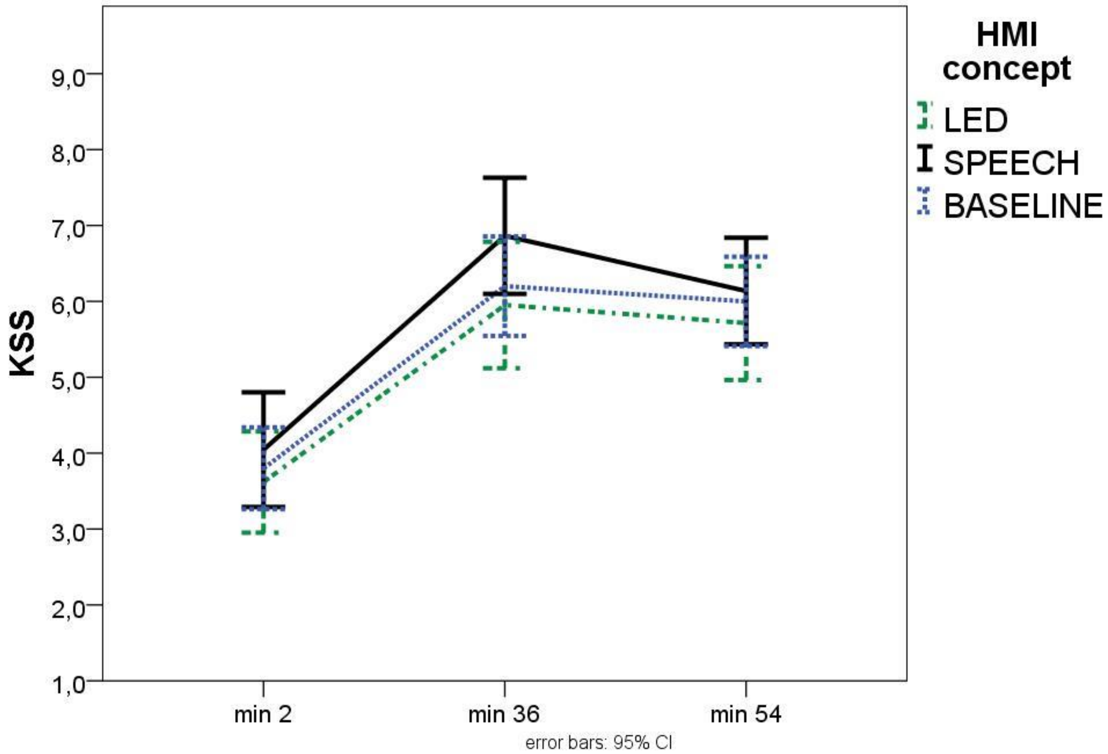


Figure 36: Experiment 4: KSS values depending on the HMI concept.

7.3.1.2. Objective Fatigue Assessment - PERCLOS

PERCLOS was highest in the *SPEECH-concept* group at t3 (M = 14.64, SD = 14.03) and lowest in the *BASELINE-concept* group at t1 (M = 5.17, SD = 2.56).

A mixed ANOVA with Greenhouse-Geisser correction showed that there was a statistically significant main effect for the factor time of measurement on PERCLOS, F (1.24, 73.08) = 14.92, p < .001, *partial* η^2 = .20. As in the KSS, there was no significant effect of the HMI concept on PERCLOS, F (2, 59) = .62, p = .54, η^2 = .021. There was also no significant interaction effect between the time of measurement and the HMI concept on PERCLOS, F (2.48, 73.08) = .38, p = .73, *partial* η^2 = .01. In Figure 37 the course of PERCLOS in dependency of the three HMI concepts is displayed.

Bonferroni-adjusted post-hoc analysis revealed a significant difference of the PERCLOS between t1 and t2 (p < .001; -4.8, 95 %-CI [-7.46, -2.14]) as well as between t1 and t3 (p = .001; -6.88, 95 %-CI [-11.12, -2.64]). There was no significant difference of the PERCLOS between t2 and t3. (p = .089; -2.08, 95 %-CI [-4.38, .22]).

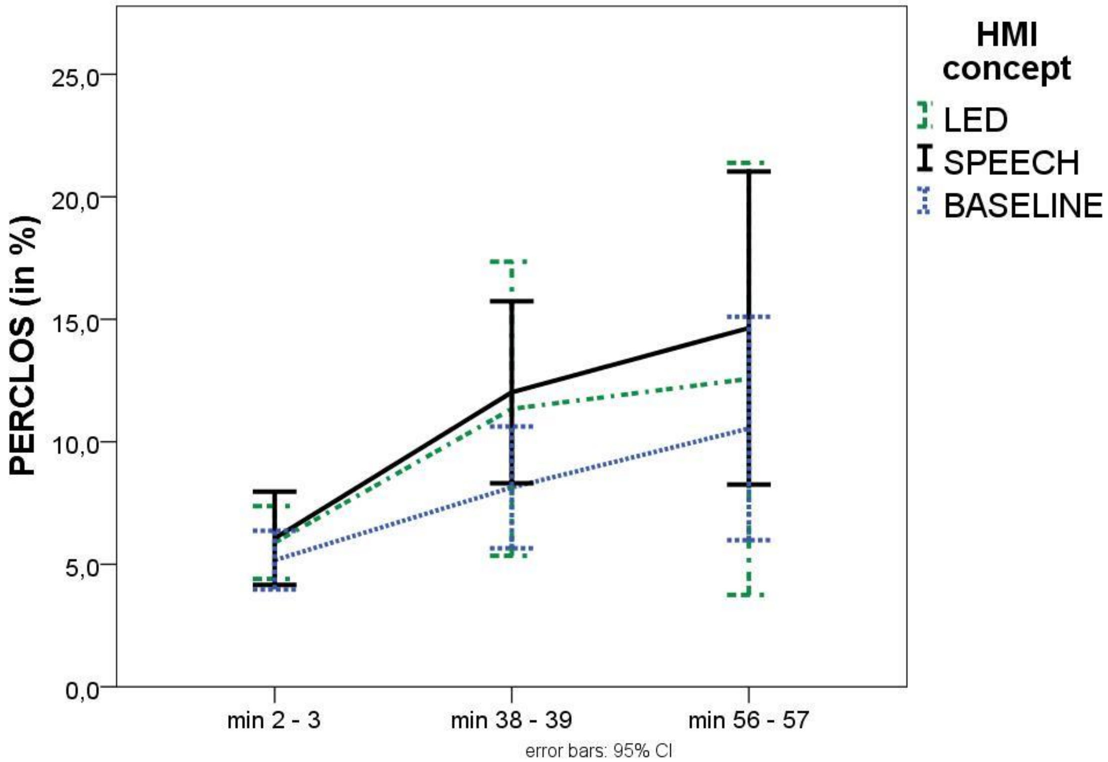


Figure 37: Experiment 4: Course of PERCLOS referring to the different HMI concepts.

7.3.2. Reaction Upon RtI Depending On The Situation

Each participant experienced two different take-over situations. In the first take-over situation, the left lane was completely free of traffic, whereby an avoidance-maneuver on the left lane would have been the appropriate reaction. In the second take-over situation, the lanes left of the ego-vehicle were blocked by other road users. In this situation a braking / stopping maneuver would have been the appropriate reaction.

7.3.2.1. Avoidance Maneuver – First Take-Over Situation

In the first take-over situation, namely the avoidance maneuver, most of the participants reacted in a situation-adapted manner. In order to investigate the influence of the different HMI concepts on take-over performance of the participants a chi-square test was used. As 3 cell frequencies were below 5, *Fishers exact test* was used. Results showed no significant effect for the HMI concept on the take-over reaction in the avoidance-situation, Fishers exact χ^2 (2) = .36, p = .87, φ = 0.07. See Figure 38 for details.

One participant of the *BASELINE-condition* could not control the vehicle and skidded on the road.

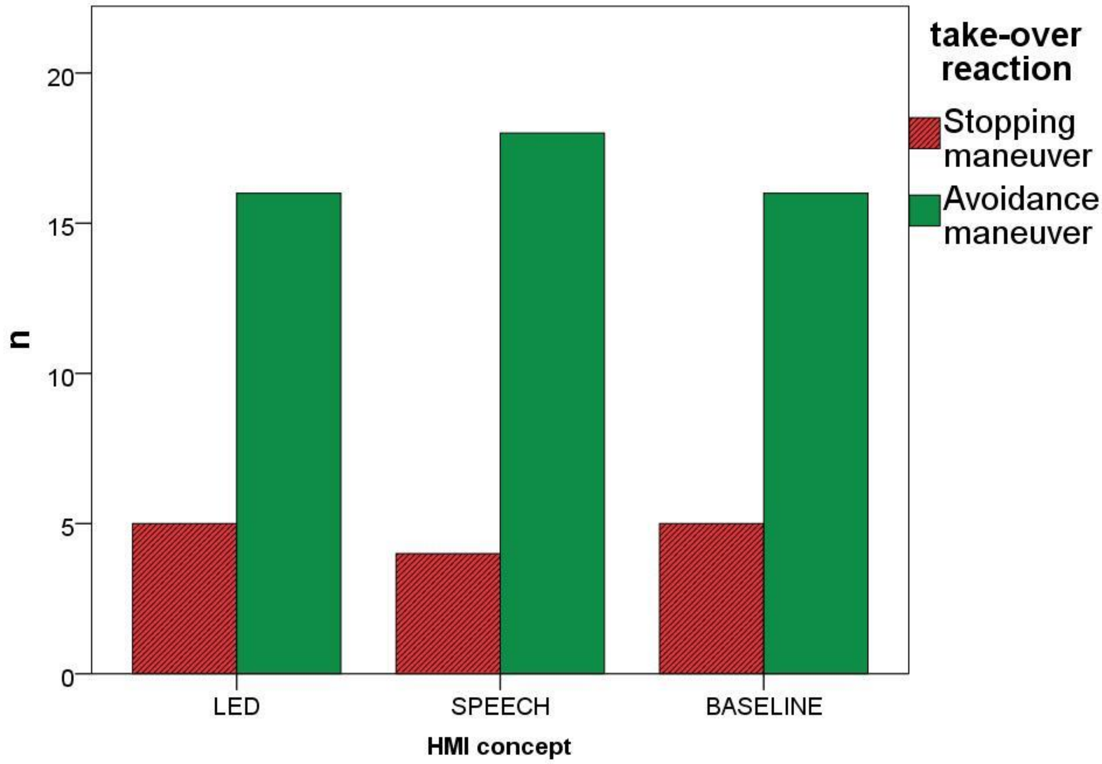


Figure 38: Experiment 4: Take-over reaction depending on the HMI concept (green = appropriate reaction).

7.3.2.2. Take-Over Situation – Stopping Maneuver

In the second take-over situation, again most of the participants reacted appropriate to the situation. However, especially participants from the *LED-concept* group solved the situation appropriately. As 3 cell frequencies were below 5, *Fishers exact test* was used for statistical analysis. Results show no significant effect of the HMI concept on the take-over reaction, Fishers exact χ^2 (2) = 5.3, p = .06, φ = .287. See Figure 39 for further details.

In this take-over situation, each one participant from the *LED-concept*, and the *SPEECH-concept* skidded due to inadequate steering. One participant from the *BASELINE-condition* collided with a vehicle in the middle lane.

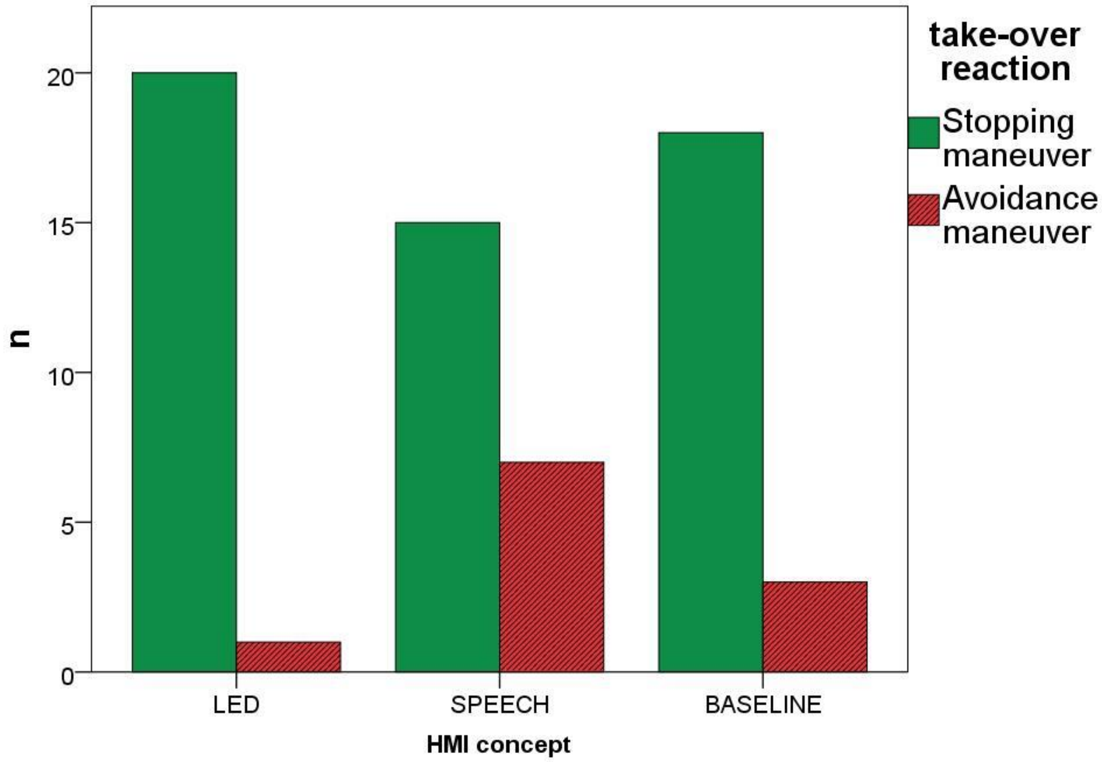


Figure 39: Experiment 4: Take-over reaction in the second situation depending on the HMI concept (green = appropriate reaction).

7.3.3. Results In Take-Over Performance

To evaluate the effects of the different HMI concepts on take-over performance in prolonged CDA, take-over performance measures after the RtI were analyzed. Different driving performance parameters and reaction times depending on the HMI concepts were examined using one-way ANOVAs. Each dependent variable was one particular driving parameter. The three different HMI concepts represent the factor of the group.

7.3.3.1. Driving Performance Parameters

7.3.3.1.1. Maximum Longitudinal Acceleration (in m / s^2)

Avoidance maneuver – first take-over situation. In the first take-over situation, an avoidance maneuver was the appropriate reaction upon the RtI. The majority of the participants reacted appropriately and did the lane-change maneuver. However, even the test persons who reacted correctly by a lane-change, braked. With M = 3.86 m / s^2 (SD = 4.39 m / s^2) for the *SPEECH-concept* (n = 18), M = 4.6 m / s^2 (SD = 3.54 m / s^2) for the *LED-concept* (n = 16) and M = 5.91 m / s^2 (SD = 3.87 m / s^2) for the *BASELINE-concept* (n = 16) the means for longitudinal acceleration did not differ significantly between the three concepts, F (2, 49) = 1.147, p = .33, *partial* η^2= .047.

However, in the *SPEECH-concept* (n = 4), in the *LED-concept* (n = 5) and in the *BASELINE-concept* (n = 5), some participants reacted by a stopping maneuver. The strongest longitudinal acceleration of participants that reacted by a stopping maneuver could be observed in the group *SPEECH-concept* (M = 10.03 m / s^2, SD = 0.23 m / s^2) whereas the lowest longitudinal acceleration could be observed in the *BASELINE-concept* (M = 9.65 m / s^2, SD = 0.82 m / s^2). In the condition *LED-concept*, a mean maximum longitudinal acceleration of M = 9.94 m / s^2 (SD = 0.23 m / s^2) was recorded. Due to the low number of cases, no further statistics were calculated.

Stopping maneuver – second take-over situation. In the second take-over situation, a braking maneuver was the appropriate reaction upon the RtI. The majority of the participants reacted the appropriate way and stopped in front of the accident. Only participants that reacted with a braking maneuver are considered in this evaluation. The strongest longitudinal acceleration could be observed in the *SPEECH-concept* group (M = 10.32 m / s^2, SD = 0.58 m / s^2) whereas the lowest longitudinal acceleration could be observed in the *LED-concept* group (M = 10.11 m / s^2, SD = 0.38 m / s^2). In the *BASELINE-concept* group, a mean maximum longitudinal acceleration of M = 10.14 (SD = 0.38) was recorded. However, the results differ only very slightly and do not differ significantly from each other F (2, 51) = .997, p = .37, *partial* η^2 = .04.

In the second take-over situation, some participants reacted by a lane-change maneuver instead of a stopping maneuver. In the group *SPEECH-concept* (n = 6) a longitudinal acceleration of M = 4.27 m / s^2 (SD = 3.35 m / s^2) compared to M = 4.92 m / s^2 (SD = 2.81 m / s^2) in the BASELINE-condition was recorded. Due to the low number of cases, no further statistics were calculated.

7.3.3.1.2. Maximum Lateral Acceleration (In m / s²)

Avoidance maneuver – first take-over situation. In the first take-over situation, an avoidance maneuver was the appropriate reaction upon RtI. All participants who reacted by a lane-change maneuver were included in this analysis. The strongest lateral acceleration occurred in *SPEECH-concept* group (M = 3.34 m / s^2, SD = 1.13 m / s^2) whereas the lowest lateral acceleration occurred in *LED-concept* group (M = 2.49 m / s^2, SD = 0.93 m / s^2). In the *BASELINE-condition*, a lateral acceleration of M = 2.88 m / s^2 (SD = 1.43 m / s^2) was measured. These measured accelerations did not differ significantly depending on the different HMI concepts, F (2, 49) = 2.2, p = .122, *partial* η^2 = .086.

The test persons who reacted by a braking maneuver are not reported here, as they did not achieve any significant lateral accelerations

Stopping maneuver – second take-over situation. In the second situation a braking reaction and stopping in front of the accident was the appropriate reaction. Only in the *SPEECH-concept* group (*n* = 6) and in the *BASELINE-condition* (*n* = 2) some participants reacted by a lane change maneuver (although there was a convoy on the lane left lane). The lateral acceleration was higher in the *SPEECH-concept* group ($M = 3.12$ m / s^2, $SD = 1.03$ m / s^2) compared to the Baseline group ($M = 2.23$ m / s^2, $SD = 0.04$ m / s^2). Due to the low number of cases, no further statistics were calculated.

The participants who reacted by a braking maneuver and stopped in front of the accident are not reported here as they did not achieve any significant lateral accelerations.

7.3.3.1.3. $TTC_{\_MIN}$

$TTC_{\_MIN}$ was calculated for the two take-over situations.

Avoidance maneuver – first take-over situation. In the first take-over situation, with $M = 2.08$ s ($SD = 0.69$ s) the *SPEECH-concept* group had the smallest $TTC_{\_MIN}$. The *BASELINE-concept* had a $TTC_{\_MIN}$ of $M = 2.37$ s ($SD = 0.77$ s) and the *LED-concept* group had the highest $TTC_{\_MIN}$ with $M = 2.48$ ($SD = 0.74$ s). These differences were not significant, $F = (2, 63) = 1.68$, $p = .20$, *partial* $\eta^2 = .05$.

Stopping Maneuver – second take-over situation. In the second situation, with $M = 2.36$ s ($SD = 0.91$ s) the smallest $TTC_{\_MIN}$ was recorded in the *BASELINE-concept* group. The *SPEECH-concept* had a mean $TTC_{\_MIN}$ of 2.49 s ($SD = 1.12$ s). The biggest $TTC_{\_MIN}$ was recorded in the LED-concept group with a $TTC_{\_MIN}$ of $M = 3.43$ s ($SD = 0.94$ s). These differences in the minimal $TTC_{\_MIN}$ were statistically significant, $F(2, 59) = 6.98$, $p = .002$, *partial* $\eta^2 = .2$. A Bonferroni-adjusted post-hoc analysis revealed a significant difference ($p = .003$) in $TTC_{\_MIN}$ between the *LED-concept* and the *BASELINE-concept* (1.08, 95 %-CI [0.30, 1.85]) as well as a significant difference between the *LED-concept* and the *SPEECH-concept* (.94, 95 %-CI [0.17, 1.72]). Thus, the *LED-concept* improved $TTC_{\_MIN}$ compared to the *SPEECH-concept* and the *BASELINE-concept*.

7.3.3.2. Take-Over Reaction Times

Five different reaction times upon the RtI were evaluated to determine whether there are differences between the HMI concepts.

Avoidance maneuver – first take-over situation. In the first take-over situation, there were significant differences in t_{eyes}, $t_{brake\_maneuver}$ and t_{steer}. There were no significant differences for $t_{brake\_reaction}$ and t_{hands}. For further information see Figure 40 and Table 20.

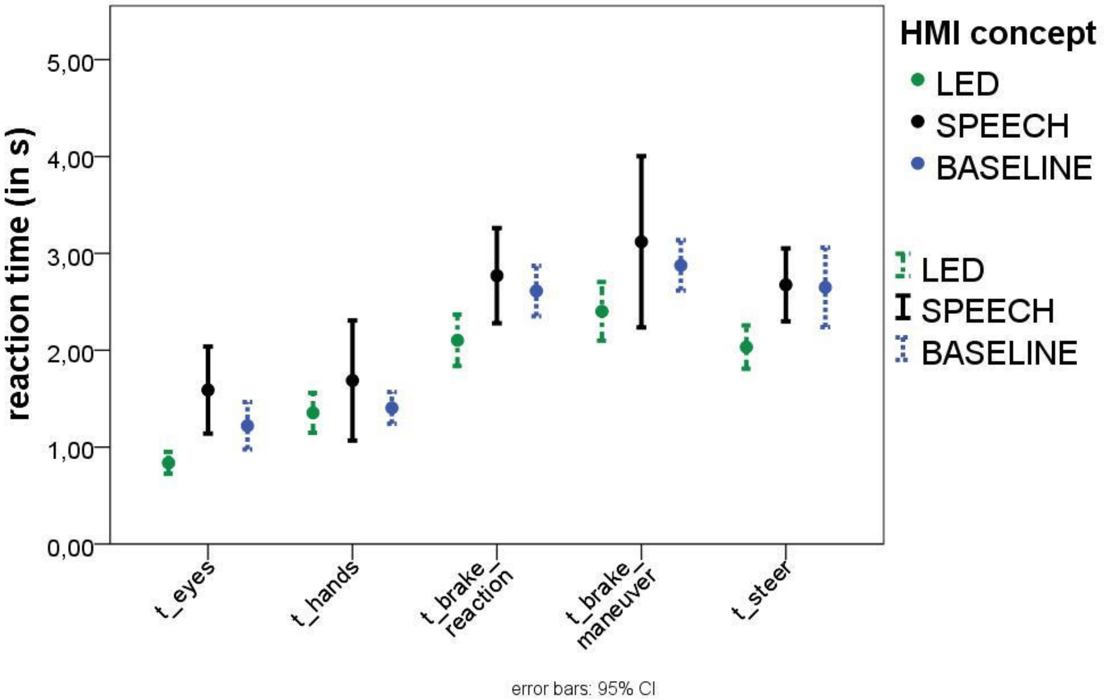


Figure 40: Experiment 4: Reaction times in the first take-over situation.

Table 20: Experiment 4: Reaction times (in s) depending on the different HMI concepts.

	Reaction-time (in s)	BASE-LINE		SPEECH		LED		*F*	*p*	*partial η^2*
		M	*SD*	*M*	*SD*	*M*	*SD*			
**	t_{eyes}	1.23	0.46	1.57	0.60	0.82	0.25	(Welch) 17.97	< .001	.32
	t_{hands}	1.47	0.39	1.73	0.63	1.33	0.33	(Welch) 2.89	.071	.12
**	t_{steer}	2.60	0.73	2.64	0.53	2.00	0.39	5.47	< .001	.22
	$t_{brake\_reaction}$	2.70	0.58	2.74	0.48	2.39	0.66	1.97	.15	.07
*	$t_{brake\_maneuver}$	2.85	0.50	3.04	0.89	2.41	0.51	3.62	.035	.12

*Note. * p < .050, ** p < .010*

Stopping maneuver – second take-over situation. In the second take-over situation, there were significant differences in t_{eyes}, t_{hands}, $t_{brake\_reaction}$, and $t_{brake\_maneuver}$. There were no significant differences for the t_{steer}. For further information see Figure 41 and Table 21.

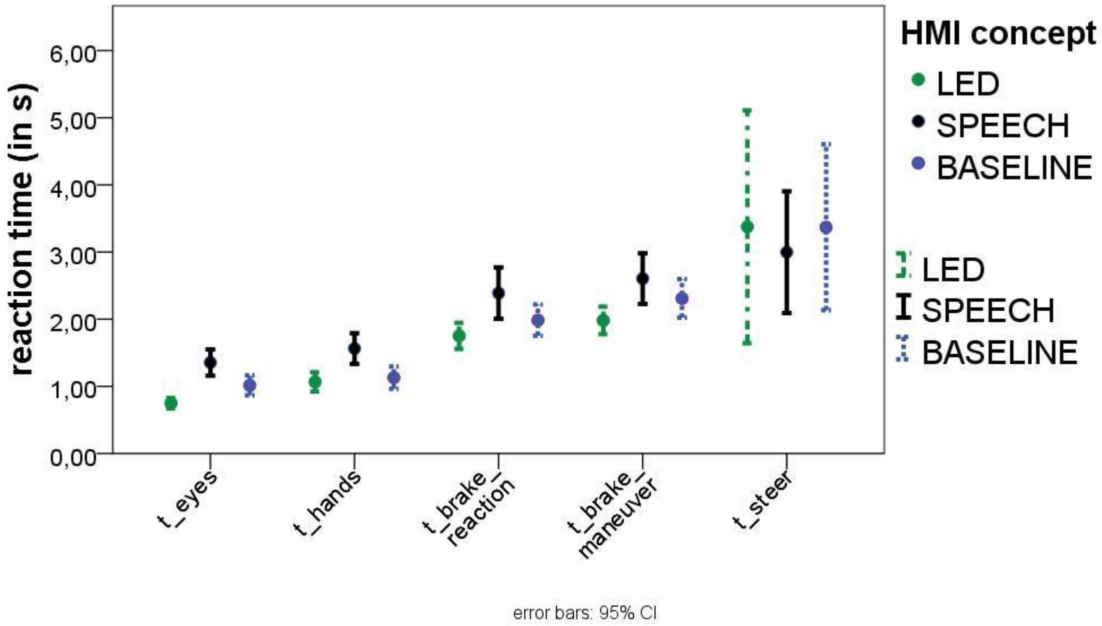


Figure 41: Experiment 4: Reaction times in the second take-over situation.

Table 21: Experiment 4: Reaction times (in s) depending on the different HMI concepts.

	Reaction time (in s)	BASELINE		SPEECH		LED		*F*	*p*	partial η^2
		M	*SD*	*M*	*SD*	*M*	*SD*			
**	t_{eyes}	1.02	0.30	1.35	0.41	0.75	0.17	19.296	< .001	.42
**	t_{hands}	1.13	0.31	1.56	0.41	1.07	0.25	9.301	< .001	.32
	t_{steer}	3.37	2.64	3.00	1.94	3.38	3.70	.115	.892	.00
*	$t_{brake\_reaction}$	1.96	0.50	2.39	0.82	1.75	0.42	(Welch) 4.990	.012	.17
*	$t_{brake\_maneu\text{-}ver}$	2.31	0.61	2.60	0.77	1.98	0.44	4.883	.011	.15

Note. * $p < .050$, ** $p < .010$

7.3.3.3. TOC-Rating

Additionally, the take-over performances the participants showed in the two situations were evaluated using the TOC-rating tool. It was examined whether the different HMI concepts lead to a difference in the evaluation for both situations.

Avoidance maneuver – first take-over situation. A Kruskal-Wallis test showed that the take-over performance evaluated with the TOC-rating was not influenced by the different HMI concepts in the first situation (Chi-square = .522, *p* = .77). See Figure 42 for details.

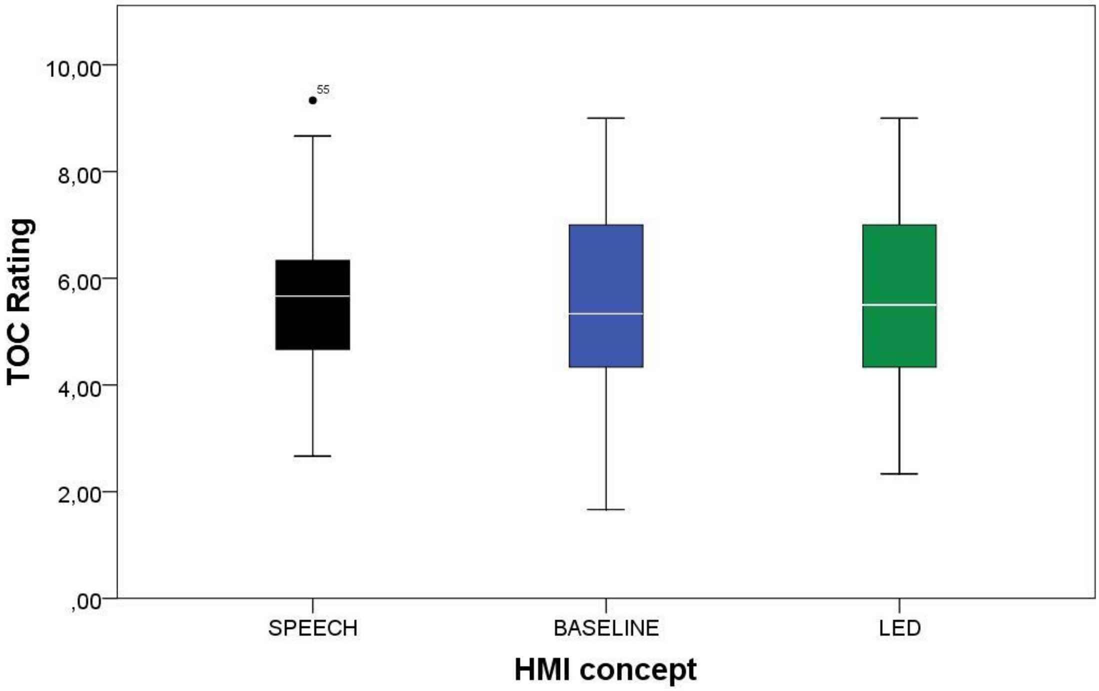


Figure 42: Experiment 4: TOC-rating in the first take-over situation according to the different HMI concepts.

Stopping maneuver – second situation. A Kruskal-Wallis test showed that the take-over performance evaluated with the TOC-rating was not influenced by the different HMI concepts in the second situation (Chi-square = 3.575, p = .17). Details can be seen in Figure 43.

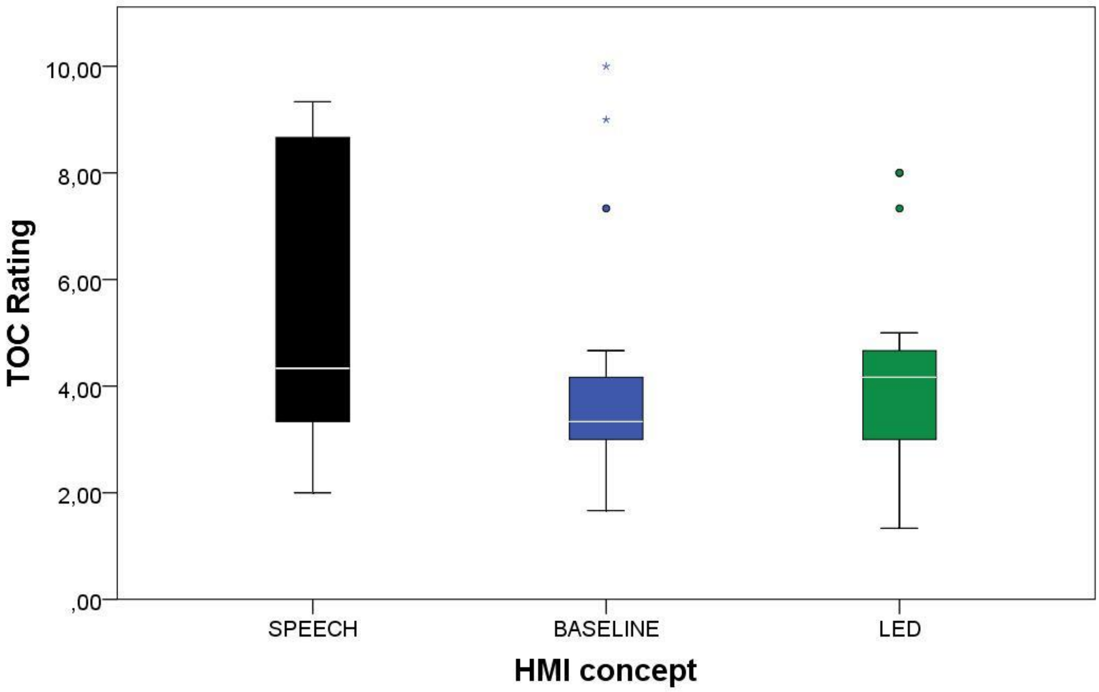


Figure 43: Experiment 4: TOC-rating in the second take-over situation.

7.4. Discussion

This experiment was conducted to investigate how the reaction of passive task-related fatigued drivers in *short-term* take-over situations after prolonged CDA can be improved with different HMI components. An impaired take-over performance of fatigued drivers after prolonged automated driving was obvious in experiments conducted before (e.g. see Experiment 2). In this experiment (Experiment 2) for example t_{eyes} as well as the $t_{brake\_reaction}$ were significantly delayed when the drivers were fatigued due to a monotonous monitoring task (*Pqpd-task*) that was used to induce task-related fatigue.

In the current experiment (Experiment 4) it should therefore be investigated how the driver can be supported in such take-over situations and how reaction times upon a RtI can be improved with the help of different HMI concepts. Therefore, three different HMI concepts with different main components, were compared and investigated: *LED-concept* (LED light signal to direct the drivers' gaze) vs. *SPEECH-concept* (speech output that warns the driver) vs. *BASELINE-concept* (text in instrument cluster).

In two different take-over situations, the efficiency of these HMI concepts was investigated. The first take-over situation could best be solved by an avoidance maneuver. In this situation, the own lane was blocked by two accident-damaged cars (TTC = 7 s). Simultaneously, the middle lane (lane left of the ego vehicle) was completely free from other road users. In the second take-over situation, on the other hand, a braking maneuver would have been the most adequate reaction of the human drivers. Again, the own lane was blocked due to an accident in front of the ego vehicle. However, in the second situation the middle lane was occupied by a convoy of other road users whereby a safe lane-change maneuver got impossible. TTC in this second take-over situation again was 7 s.

In the previous experiments especially fatigued drivers had problems in the take-over situation. Therefore, in this experiment all drivers had to deal with the monotonous monitoring *Pqpd-task,* which was successfully used to induce *passive take-related* fatigue in experiments before. The task was presented on a Windows Surface tablet PC that was installed in the front of the CID.

The experiment had two main focuses: on the one hand, the course of the drivers' fatigue was recorded to verify previous results on emerging fatigue in automated driving, and on the other hand, it was of great interest how the different HMI concepts can support the driver in the take-over situations.

Results support the previous findings that fatigue in CDA can emerge quickly, when the human driver just has to monitor a system (in the case of the experiment this was represented by the *Pqpd-task*). Subjective KSS as well as objective PERCLOS significantly increased with increasing time of automated driving and simultaneously engaging in the monitoring task. The measured fatigue level between the baseline measurement in the beginning of the ride significantly differed compared to the fatigue-measurements before the two take-over situations. Therefore, PERCLOS could again be verified as one sensitive objective fatigue measurement. At some participants PERCLOS rose up to 15 % during the CDA ride, which is defined as a state of *fatigue* (Wierwille et al., 1994).

The other main focus of this experiment was to investigate, which HMI concept can best support the human driver in *short-term* take-over situations. Here, the results of the present experiment clearly indicate that the different HMI concepts can significantly affect the performance of the human driver in *short-term* take-over situation.

Reaction times of the drivers could be shortened, with the *LED-concept*. This concept existed of a red flashing LED stripe that was mounted on the engine hood directly in front of the windshield. As a result, t_{eyes} could be significantly shortened and also other reaction times (e.g. $t_{brake\_reaction}$, t_{hands}) showed significant improvements compared to the *BASELINE-* and the *SPEECH-concept*. These result indicate that peripheral light signals can help to direct the gaze of the drivers to the hazard on the one hand, and that hereby also other reactions that have to be fulfilled by the driver in a take-over situation can be shortened.

Another result of this experiment was that the reaction times measured in the *SPEECH-concept* group were significantly slower compared to the *LED-* and also to the *BASELINE-concept*. However, this can be explained by the fact that it took a longer time (2.5 s) until the complete speech-output was played. This can have led to the fact that the participants reacted first after they had heard the speech output completely.

Also in other evaluations, the *LED-concept* could show advantages over the other concepts. Considering the take-over reactions of the drivers upon the RtI, it appears like the *LED-concept* helped the participants to react adequately to the specific take-over situation. This was especially the case in the second take-over situation where the left lane was blocked by other road users and the participants best would have reacted by a braking maneuver. One possible explanation for the good results achieved in the *LED-group* could be that the red flashing of the LED was interpreted as braking lights of a vehicle in front, whereby a braking reaction was triggered. Only one participant of the *LED-concept* group reacted inadequately by doing a lane-change maneuver. The HMI-concept with which the drivers reacted the least adapted to the situation was the *SPEECH-concept*. In the second take-over situation, where a braking reaction was the adequate reaction, *n* = 7 (32 %) participants of the *SPEECH-concept* group opted for the wrong reaction and performed a lane-change maneuver. A lane-change maneuver in this specific situation was always associated with a danger to oneself or other road users due to small safety distances. In the first take-over situation, in which a lane-change maneuver on the middle lane was the appropriate reaction, no HMI concept had benefits compared to the other concepts regarding the reactions of the drivers. In all HMI concepts, a rather similar amount of the participants (*LED-* & *BASELINE-concept* 31 %, *SPEECH-concept* 22 %) performed a braking maneuver instead of the preferable lane-change maneuver.

Considering the measurements that provide information about the quality of the human drivers' intervention in the take-over situations, no differences could be found for longitudinal and lateral accelerations with respect to the different HMI concepts. However, when considering the $TTC_{\_MIN}$, again the *LED-concept* has achieved the best results compared to the other concepts. In the second take-over situation, these differences were significant.

In summary, it can be said that the drivers of the *LED-concept* group in particular achieved the best results in the *short-term* take-over situations. Especially in take-over situations, in which a braking maneuver is required, such a visual concept can contribute to the safety in CDA. On the other hand, the *SPEECH-output* did not show any advantages in *short-term* take-over situations.

Future research should now investigate how such a light signal concept can best be implemented in a production vehicle, where the light signal should be displayed and with which source such a light signal can be generated. It must also be investigated whether such a warning signal can withstand different lighting conditions on-road.

7.5. Limitations

The experiment was conducted in the motion base driving simulator of the BMW Group. All participants were employed at BMW, so it can be assumed that the results are not unrestrictedly transferable to the overall population. A further limitation of the interpretability of the results is that the lighting conditions in the simulator do not correspond to those in real daylight traffic. This has probably increased the effect of the light stimulus in the *LED-concept* group. It is also unclear to what extent the take-over performance parameters (take-over reaction times and quality parameters) can be transferred to real driving performance parameters.

7.6. Précis

- In a motion base driving simulator experiment three different HMI concepts for *short-term* take-over situations were evaluated.

- The following concepts were considered: *LED-concept* (red flashing LED light that directs the gaze of the driver towards the hazard) vs. *SPEECH-concept* (speech output that warns the driver) vs. *BASELINE-concept* (text in instrument cluster).

- Each participant experienced two different *short-term* take-over situations whereas once a braking response and once a lane-change maneuver would have been the appropriate reactions.

- Results suggest that the *LED-concept* led to superior results compared to the other concepts. These included reaction times as well as parameters concerning the quality of the drivers' intervention and the overall reaction upon the RtI.

8. Summary Of The Results

So far, in the recent work the basic background and theory of the experiments have been described (see chapter 1. & 2.). The basic method of the four experiments was explained in chapter 3. In chapter 4. – 7. the four different experiments are explained in detail.

In the following section, the results of the four experiments concerning the drivers' state and the achieved results in the take-over situations will be summarized and compared.

On the one hand, in all experiments the drivers' fatigue state was influenced by different NDRTs. The resulting fatigue was measured with both objective and subjective measurements for different times over the course of the experiments. In this section of the thesis the course of emerging fatigue over all experiments will be compared and additionally an illustration will be displayed where all results will be aggregated.

On the other hand, another focus of the experiments was whether and how the resulting fatigue affects the take-over performance in *short-term* take-over situations. Therefore, in three of the four experiments, participants were confronted with at least one take-over situation. Also these results are compared and presented in summary. Both reaction times as well as driving parameters are compared.

8.1. Overall Findings Relating To the Fatigue state of the drivers

In the Experiments 1 – 3 it was examined if and how fatigue emerges in CDA in relation to different NDRTs. According to the fatigue model of May and Baldwin (2009) it is especially *passive task-related fatigue*, which is provoked through increasing monotony and increasing automation. Exactly these preconditions are given in automated driving. Thus it is expected, that *active task-related fatigue*, which is the biggest causation for fatigue in manual driving will be replaced by *passive task-related fatigue* in CDA.

In order to investigate to which extent NDRTs can be used to affect the monotony whilst driving automated, the participants in the four experiments had to engage in predefined NDRTs.

On the one hand, a monotonous monitoring task (*Pqpd-task*) was designed which is supposed to cause *passive task-related fatigue* in the experimental rides. The task was used in all four experiments to cause *passive task-related fatigue* and was either tested in contrast to other NDRTs (Experiment 1 – 3) or used specifically to fatigue all participants (Experiment 4).

On the other hand, in order to compare the resulting fatigue of the *Pqpd-task* with other activities, in three experiments NDRTs have been involved that should prevent the drivers from fatigue and keep them rather awake. In Experiment 1 (see section 4.) and Experiment 2 (see section 5.) a *Quiz-task* was used for this purpose. In the *Wizard-of-Oz* on-road experiment, Experiment 3 (see section 6.), the participants could engage in a freely selectable activity (*Free-choice-activity*).

A special case was the fourth and last experiment. In Experiment 4 (see section 7.) all participants had to deal with the fatiguing *Pqpd-task* to investigate which HMI concept best supports even a fatigued driver in a *short-term* take-over situation. Thus, in section 8.1. Experiment 4 is not discussed in detail.

In order to measure emerging fatigue due to the different NDRTs fatigue was assessed by using the subjective KSS as well as objective PERCLOS.

8.1.1. Subjective Fatigue States Over The Four Experiments - KSS

The subjective KSS was assessed for defined points of time during the experimental rides. However, in the different experiments, subjective fatigue was assessed at different points of time due to differences in the total duration of the rides.

In the following, the KSS results from the Experiments 1 - 3 are discussed. In these three experiments NDRTs were used to affect the fatigue states of the drivers.

As can be seen in Figure 44, KSS has risen especially in the experimental conditions with the monotonous monitoring task (*Pqpd-task*; see solid lines in Figure 44). In combination with shorter driving times, the *Quiz-task* (see dotted lines in Figure 44) was able to keep the drivers on a rather low level of subjective fatigue (see blue dotted line in Figure 44). This effect decreased with prolonged driving time (see green dotted line in Figure 44) as the KSS ratings revealed higher values. However, a freely selectable activity during automated driving could be used to counteract increasing fatigue in the Experiment 3 (*Wizard-of-Oz*) experiment (see black dotted line in Figure 44). The *Pqpd-task* which was used to provoke *passive task-related fatigue* has led to increased subjective fatigue assessments of the participants. All courses of the subjective fatigue states are at all times above those of the other activities.

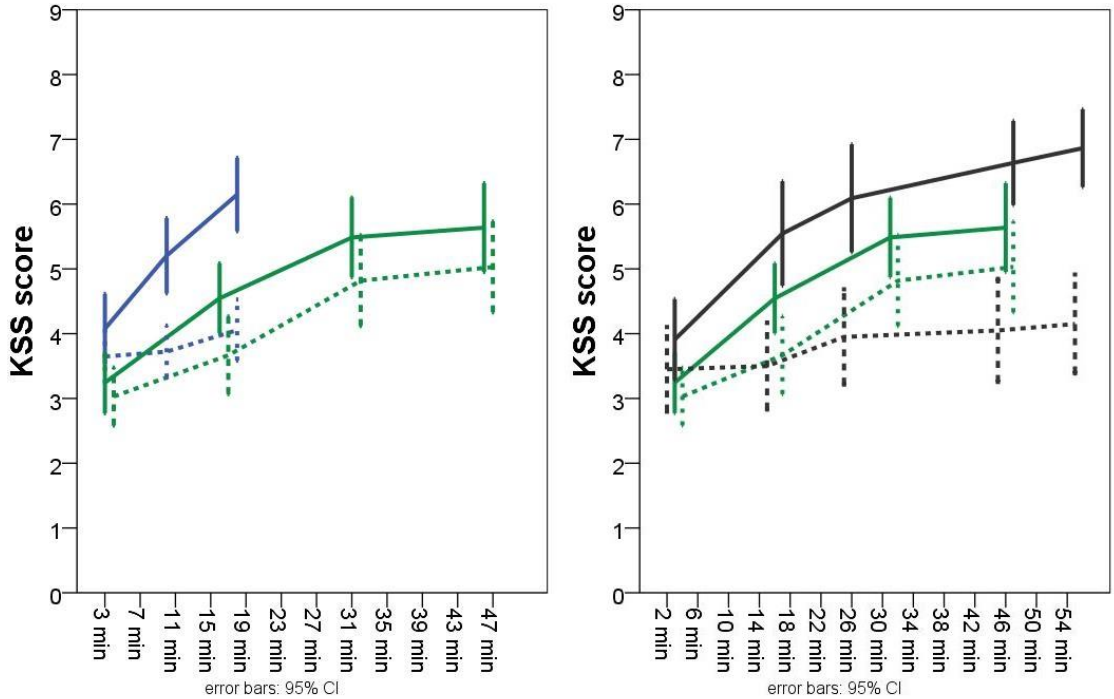


Figure 44: Course of KSS for the different experiments. **Experiment 1** (*N* = 112, blue), **Experiment 2** (*N* = 66, green), **Experiment 3** (*N* = 42, black), *Pqpd-task* (solid lines); *Quiz-task* / *Free-choice-activity* (dotted lines)

A comparison of the last measured KSS values of all rides shows that all KSS values of the participants who had to perform the monotonous monitoring task (*Pqpd-task*) were higher than those of the activating task. The highest KSS was measured in the *Wizard-of-Oz* on-road experiment after 56 min of the ride and simultaneously engaging in the monotonous monitoring task (*Pqpd-task*) with M = 6.86 (SD = 1.36). The lowest KSS value was measured in the *Quiz-task* group after 18 min of CDA (M = 4.06, SD = 1.73). An overview of each last measured KSS values and the associated statistics can be found in Table 22.

Table 22: Overview of each last assessed KSS values of the different experimental rides.

	Experiment	NDRT	N =	Environment	Duration	M	SD	Statistic	p	Effect size
**	Experiment 1	Pqpd	56	Simulator	25 min	6.15	2.08	*F* = 17.95	< .001	η_p^2 = .15
		Quiz	56			4.06	1.73			
	Experiment 2	Pqpd	33	Simulator	50 min	5.64	1.93	*F* = 2.14	.15	η_p^2 = .03
		Quiz	33			5.03	1.98			
**	Experiment 3	Pqpd	22	*Wizard-of-Oz*	60 min	6.86	1.36	*F* = 21.97	< .001	η_p^2 = .93
		Free choice activity	20			4.15	1.63			
	Experiment 4	Pqpd	64	Simulator	50 min	5.95	1.50	*No statistic calculated*		

Note. * $p < .050$, ** $p < .010$

In order to investigate which factors significantly affected the subjective KSS, a multifactor ANOVA was calculated. Each last stated value of the KSS measurement from the Experiment 1, Experiment 2 and Experiment 3 has been included in the analysis. As possible factors that may have affected subjective fatigue (KSS), the duration of the journey (short vs. long), the NDRT (active vs. passive task) and the test environment (Simulator vs. *Wizard-of-Oz*) were selected.

Results suggest that the overall model is significant, F (5, 210) = 12.14, $p < .001$, *adjusted* $R^2 = .21$, $n = 216$. It can be seen that the test environment (i.e. *Wizard-of-Oz* vs. Simulator) did not significantly affect the subjective KSS rating, F (1, 210) = .225, $p = .64$. Furthermore, also the duration of the experimental rides did not significantly affect the subjective KSS ratings, F (1, 210) = .64, $p = .43$. Depending on the NDRTs the participants had to engage in while driving automated, however, different subjective KSS ratings were reported, F (1, 210) = 51.47, $p < .001$, *partial* $\eta^2 = .2$. According to Cohen (1988) this represents a strong effect.

In addition to the main effects, also interaction effects of the individual factors were assessed. A significant interaction effect between the factors environment and task on subjective KSS could be found, F (1, 210) = 8.23, $p = .004$, *partial* $\eta 2 = .038$. Accordingly, there was a stronger effect of the NDRT in the *Wizard-of-Oz* environment. This means that subjects who could freely choose their current activity while driving automated in the *Wizard-of-Oz* environment were more awake than the subjects that had to deal with the activating *Quiz-task* in the driving simulator experiments. Furthermore it can be stated that participants who had to work on the monotonous monitoring *Pqpd-task* in the *Wizard-of-Oz* experiment were more fatigued than those in the driving simulator.

Also a significant interaction effect between the factors duration of the ride and the task could be found, F (1, 210) = 6.56, p = .011, *partial* η^2 = .03. Results suggest, that participants who had to engage in the *Pqpd-task* reported higher states of subjective fatigue as early as after 18 min. In this task-group, with increasing duration of the ride, the subjective fatigue increases only marginally. This looks different in the activating task condition. A significant increase in the subjective KSS rating can be observed according to increasing journey time.

As there was only one *Wizard-of-Oz* experiment, no interaction effect between the factors environment and duration was calculated.

According to Åkerstedt et al. (2014), KSS values > 6 go along with an impaired driving performance in manual driving. It can therefore be assumed that higher KSS values are also linked to an impaired performance in a take-over situation in CDA. Therefore, to identify how many participants reached KSS scores > 6 in the different experiments another analyses was conducted. As can be seen in Figure 45, in all experimental conditions, in which the participants had to engage in the monotonous monitoring *Pqpd-task*, about 50 % of the participants reached KSS values > 6. The experimental condition with the most participants reaching KSS > 6 was the *Wizard-of-Oz* on-road ride in combination with the *Pqpd-task* (63.6 %). The ride with the lowest rate of participants reaching KSS > 6 was the *Wizard-of-Oz* experiment in combination with the *Free-choice-activity* (5 %). Also the *Quiz-task* was able to prevent subjective fatigue relatively well in the 25 min Experiment 1 ride. Only 11.1 % of the participants indicated a KSS > 6. However, in this row of experiments the activating effect of the *Quiz-task* decreased with increasing driving time. In the Experiment 2, 27.3 % of the participants who had to engage in the *Quiz-task* for 50 min while driving automated stated a KSS value of 6 or above. See Figure 45 for the percentage of participants reaching KSS > 6 for the different experiments.

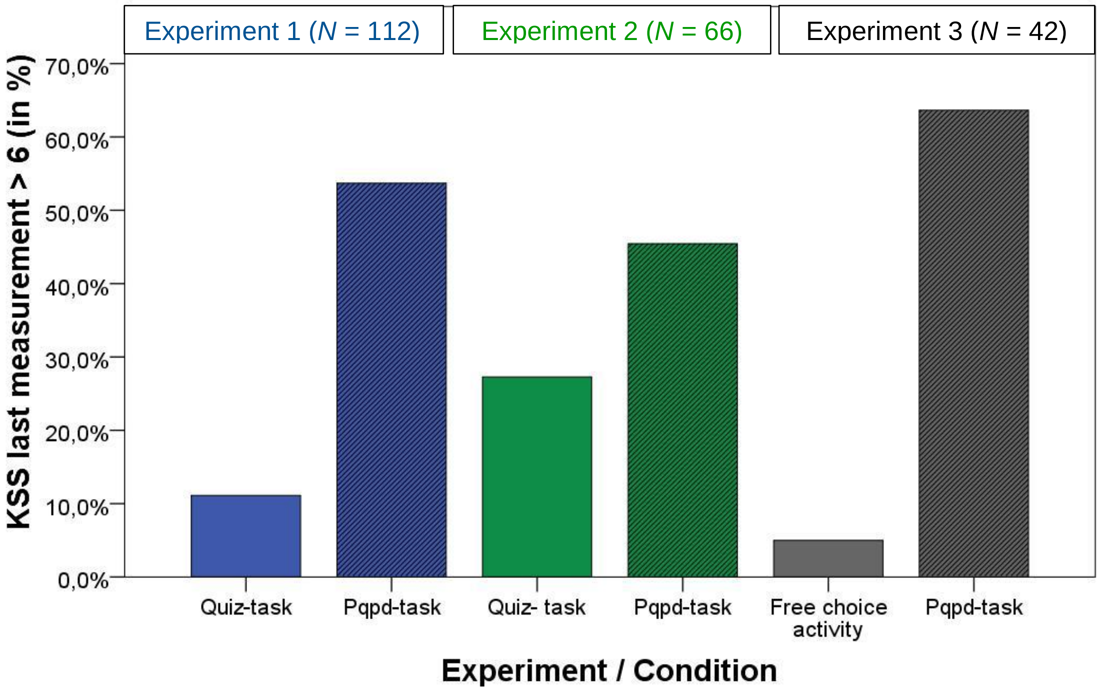


Figure 45: KSS > 6 for the different experimental conditions.

8.1.2. PERCLOS – Objective Fatigue States Over The Four Experiments

As objective parameter for measuring emerging fatigue PERCLOS was recorded over the course of the experimental rides to defined points of time in all four experiments. Similar to the KSS assessments the points of measurement differed between the four experiments. See Figure 46 for the measured PERCLOS values in Experiment 1, Experiment 2 and Experiment 3, in which the participants either had to deal with a fatiguing task (*Pqpd-task*) or an activating task (*Quiz-task* or *Free-choice-activity*). The fatiguing NDRT (*Pqpd-task*) is displayed with the solid lines and the NDRT that should prevent participants from getting fatigued is displayed with the dotted line (i.e. *Quiz-task* or *Free-choice-activity* in Experiment 3). As in Experiment 4 only the fatiguing NDRT (*Pqpd-task*) had to be processed by all participants, this experiment is not displayed for sake of clarity.

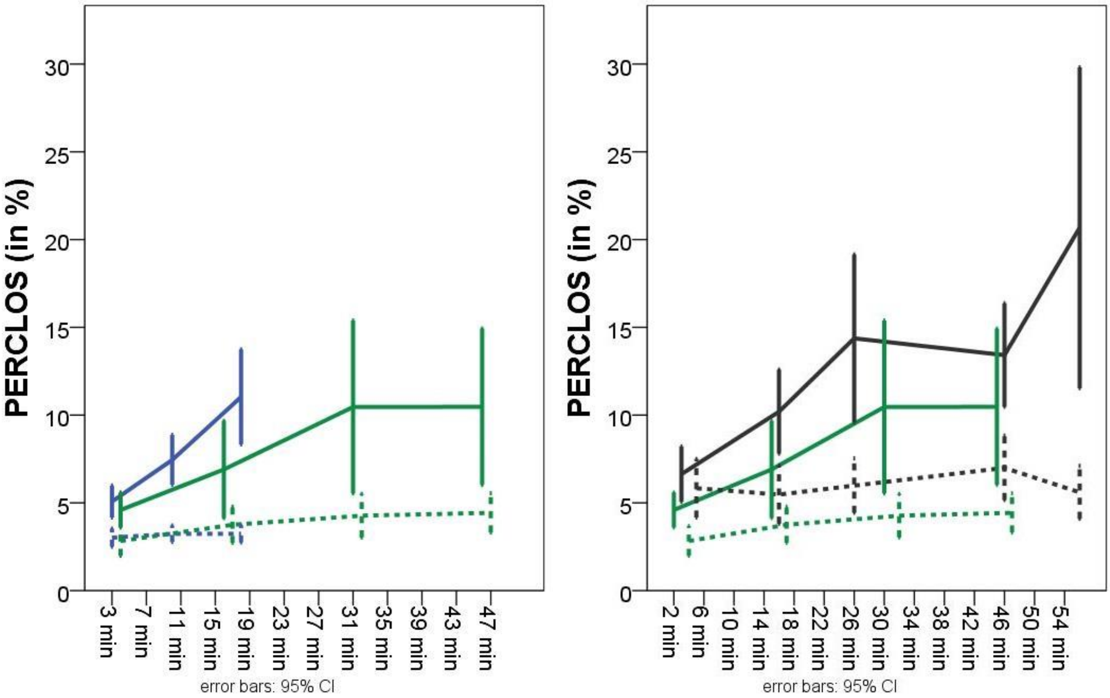


Figure 46: Course of PERCLOS for the different experiments. **Experiment 1** (*N* = 112, blue), **Experiment 2** (*N* = 66, green), **Experiment 3** (N = 42, black), *Pqpd-task* (solid lines); *Quiz-task* / *Free-choice-activity* (dotted lines)

As can be seen in Figure 46, in all conditions in which participants had to engage in the *Pqpd-task*, PERCLOS has increased over the course of the experimental ride (see solid lines in Figure 46). In all other conditions, in which participants either had to deal with the *Quiz-task* (dotted lines; Experiment 1, Experiment 2) or could freely choose their activity (Experiment 3) only a slight increase or no increase at all could be found.

The highest PERCLOS was measured in the *Wizard-of-Oz* experiment on-road in real driving environment after 60 min when participants had to monitor and process the *Pqpd-task*. However, independently from the NDRT, PERCLOS in the *Wizard-of-Oz* experiment was above the measured PERCLOS values from all experiments in the driving simulator for all points of measurement. The measured PERCLOS after 1 hr of automated driving and simultaneously engaging in the fatiguing monitoring task is at about 20 %. According to Wierwille (1994), this corresponds to the state of *fatigue* which corresponds to PERCLOS of 15 % and above. See Table 23 for further details about the different PERCLOS measurements.

Table 23: Overview of the last recorded PERCLOS values of the different experimental rides.

	Experiment	NDRT	*N* =	Environment	Duration	*M* (in %)	*SD* (in %)	Statistic	*p*	Effect size
**	Experiment 1	Pqpd	56	Simulator	25 min	11.04	9.62	*F* = 35.4	< .001	*ηp²* = .25
		Quiz	56			3.3	1.68			
*	Experiment 2	Pqpd	33	Simulator	50 min	10.47	12.57	*F* = 7.95	= .01	*ηp²* =.11
		Quiz	33			4.43	3.15			
**	Experiment 3	Pqpd	22	*Wizard-of-Oz*	60 min	20.67	20.61	*F* = 15.55	< . 001	*ηp²* = .29
		Free-choice-activity	20			5.6	3.22			
	Experiment 4	Pqpd	64	Simulator	50 min	12.62	14.83	*No statistic calculated*		

*Note. * p < .050, ** p < .010*

In Table 23, mean values of the last measured PERCLOS values are displayed. It becomes obvious that the measured PERCLOS strongly depends on the respective NDRT / activity the participant had to work on during the automated ride. In all experimental rides, in which the participants had to deal with the monotonous monitoring task (*Pqpd-task*) PERCLOS has risen above 10 %. According to Wierwille et al. (1994) this represents a fatigue level of *questionable*. With a value of more than 15 % the status can be declared as *fatigued* (Wierwille et al., 1994). Looking at the *SD* values in these groups, it becomes evident that some of the participants even reached this level of fatigue.

The situation was completely different for the participants, who worked on the activating NDRT or the *Free-choice-activity*. In these groups, PERCLOS remained below 7.5 % for all test persons, which, according to Wierwille (1994), goes hand in hand with a status of *awake*. The strongest effect of the NDRT on PERCLOS was measured in the *Wizard-of-Oz* experiment. With *ηp²* = .29 this is a large effect (Cohen, 1988).

In order to investigate which factor (environment, NDRT, and duration of the ride) significantly affected PERCLOS over all experiments, a multifactorial ANOVA was calculated. Each last recorded PERCLOS measurement from the Experiment 1, Experiment 2 and Experiment 3 has been included in this analysis.

Results suggest that the overall model is significant, F (5, 203) = 12.27, p < .001, *adjusted* R^2 = .21, n = 209. On the one hand the NDRT the participants had to engage in, significantly affected PERCLOS of the participants, F (1, 203) = 47.06, p < .001, *partial* η^2 = .19. On the other hand there was a significant main effect for the test environment (i.e. *Wizard-of-Oz* vs. Simulator). Thus, the test environment significantly affected PERCLOS, F (1, 203) = 8.73, p = .003, *partial* η^2 = .041. In contrast to these significant influencing factors, the duration of the ride did not have a significant effect on PERCLOS, F (1, 203) = .04, p < .84.

Next to the main effects, interaction effects between the factors were calculated. A significant interaction effect between the test environment and the NDRT on PERCLOS could be found, F (1, 203) = 5.51, p = .02, *partial* η^2= .026. The recorded PERCLOS values in the passive task condition were higher compared to the activating tasks. The environment *Wizard-of-Oz* further intensified this effect.

There was no significant interaction effect between the duration of the ride and the NDRT, F (1, 203) = .33, p = .57.

As there was only one *Wizard-of-Oz* experiment, no interaction effect between the factors environment and duration was calculated.

According to Wierwille et al. (1994), the threshold for PERCLOS from which one can definitely speak of a status of *fatigue* is 15 % and above. To identify, who reached such high PERCLOS values an analysis was conducted where only those participants were considered who reached such high PERCLOS values. See Figure 47 for more details. In this illustration it can be seen how many participants (in %) from each task-group reached PERCLOS > 15 %.

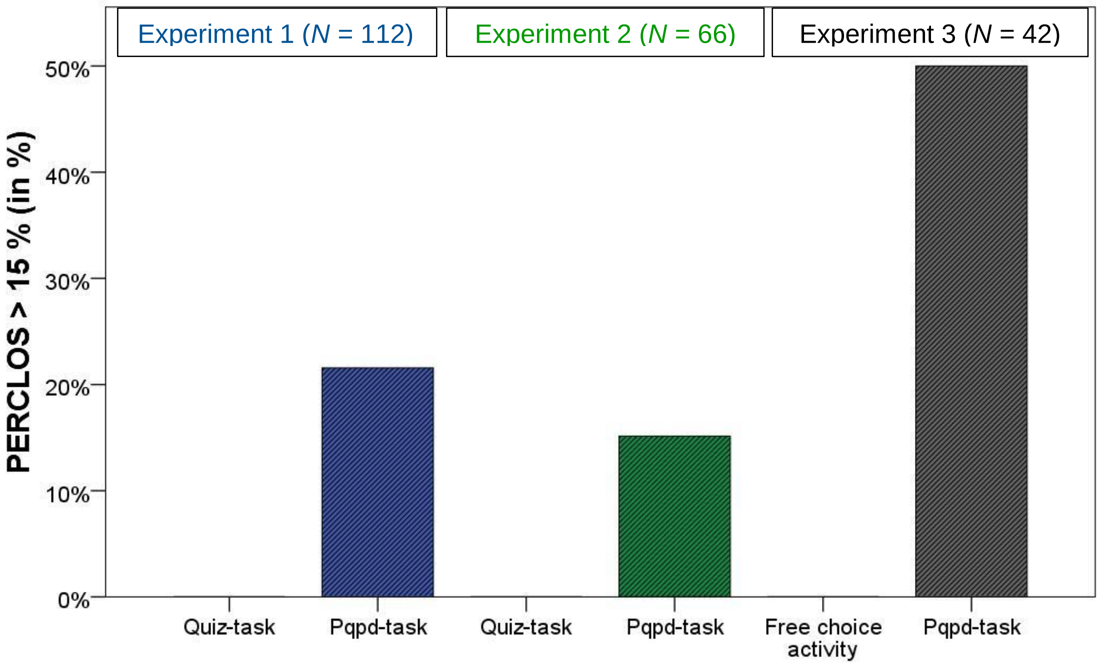


Figure 47: Participants (in %) who reached PERCLOS > 15 % above the experimental rides.

As can be seen in Figure 47, all participants who reached PERCLOS > 15 % had to engage in the *Pqpd-task* during the automated ride. Not one participant from the *Quiz-task* or from the *Free-choice-activity* reached PERCLOS > 15 %.

The highest percentage of participants reaching PERCLOS > 15 % was measured in the *Wizard-of-Oz* experiment, Experiment 3. However, the participants in this experiment already drove manually for about half an hr before the automated driving system (i.e. the *Wizard*-driver) was available and could be switched on. This can have led to an increased level of *active task-related fatigue* due to manually driving (May & Baldwin, 2009) compared to the other experimental rides. Another contributing factor for these high PERCLOS values can be the light conditions on-road. Sunlight may have affected the blinking behavior of the participants (more blinks).

Another interesting result is the following: the percentage of participants that reached PERCLOS > 15 % in Experiment 2 is lower compared to Experiment 1. This may possibly be explained by an increased self-activation of the Experiment 2 participants who had to engage in the *Pqpd-task* for about 50 min. Participants from the Experiment 1 had similar PERCLOS values after 25 min of automated driving as the participants from Experiment 2 after 50 min. Another explanation can be that PERCLOS of the Experiment 1 participants who had experienced the *Pqpd-task* in the second experimental ride after already having experienced the first ride with the *Quiz-task* for 25 min has risen faster. This group consists of 50% of the participants and therefore has a lot of influence on the total evaluation.

8.2. Overall Findings Of Take-Over Performance Measures Regarding The Reaction Times.

In all experiments except the Experiment 3 (*Wizard-of-Oz* experiment), the drivers were confronted with short term take-over situations with identical TTCs of 7 s. In this part of the thesis, a comparison will be made to determine to what extent the take-over reactions differed with regard to the reaction times (see section 8.2.) and the quality-measures of the drivers' input (see section 8.3.) due to a different driver state induced by the different NDRTs.

As in Experiment 3 no take-over situation appeared, this experiment is not taken into account in this analysis. Also Experiment 4 is disregarded in this part as the measured reaction times and quality measures are not comparable to Experiment 1 and Experiment

2. In the Experiment 4 it should be examined how to support the human driver in take-over situations, whereby clear improvements of the reaction times have been achieved. The experiment will be discussed separately later.

In a first step, the different reaction times of the drivers are considered and compared. These include t_{eyes}, t_{hands}, $t_{brake\_reaction}$, and the reaction time for the first driving maneuver which is the smaller one out of the two reaction times t_{steer} and $t_{brake\_maneuver}$.

In a further step of the evaluation the measurements that give a statement about the quality of the intervention by the drivers are compared. These include $TTC_{\_MIN}$, $Acc_{\_lat\_max}$ and $Acc_{\_long\_max}$.

8.2.1. Eyes-On-Road Reaction Time (t_{eyes})

In this analysis t_{eyes} of the Experiment 1 and Experiment 2 which included a short term take-over situation are compared. From the Experiment 1, only the results of the *accident on ego-lane* situation were included in the analysis to ensure comparability of the results as the same situation was also implemented in the Experiment 2.

A multifactorial ANOVA with the factors duration of the ride and the NDRT was calculated in order to investigate which factor significantly affected t_{eyes}.

Results suggest a significant effect of the overall model, F (3, 111) = 7.29, p < .001, *adjusted* R^2 = .14, n = 115. Furthermore, the analysis revealed that the duration of the ride, F (1, 111) = 4.16, p = .04, *partial* η^2 = .04 as well as the NDRT, F (1, 111) = 15.41, p < .001, *partial* η^2 = .12, significantly affected t_{eyes}. There was no significant interaction effect between the duration of the ride and the NDRT, F (1, 111) = .53, p = .47, *partial* η^2 = .01. Accordingly, t_{eyes} increases with increasing duration of the ride and is closely related to the NDRTs the drivers are dealing with while driving automated.

With M = 0.76 s (SD = 0.38 s) participants reacted fastest in the 25 min ride in combination with the *Quiz-task*. The slowest t_{eyes} reaction time was measured in the Experiment 2 ride when participants had to deal with the monitoring task *Pqpd-task*, M = 1.19 s (SD = 0.39 s).

See Table 24 for all measured t_{eyes} and further details for the different experimental conditions.

Table 24: Reaction times for t_{eyes} the different experimental rides.

	Experi-ment	NDRT	*N* =	Envi-ronment	Dura-tion	*M* (in s)	*SD* (in s)	Statis-tic	*p*	Effect size
	Experi-ment 1	Pqpd Quiz	27 27	Simula-tor	25 min	0.99 0.76	0.46 0.38	*F* = 3.78	= .058	*ηp²* = .07
**	Experi-ment 2	Pqpd Quiz	33 32	Simula-tor	50 min	1.19 0.85	0.39 0.31	*F* = 14.64	< .001	*ηp²* = .19

*Note. * p < .050, ** p < .010*

When considering t_{eyes} in Experiment 1 and Experiment 2, it becomes evident, that the monotonous monitoring task (*Pqpd-task*) led to impaired eyes-on-road reaction times. In both experiments, the mean reaction time was lower, when participants had to deal with the fatiguing task. However, this effect was significant only in the prolonged automated ride in the Experiment 2 in which participants had to drive for 50 min while engaging in the monotonous *Pqpd-task*. In respect to the effect size, *ηp²* = .19, a strong effect is involved here (Cohen, 1988). In Experiment 1, on the other hand, the threshold for significance is missed slightly.

To conclude it can be said, that passive tasks, like monitoring tasks can lead to impaired eyes-on-road reaction times in CDA. The measured effects are not really strong when considering the effect sizes (Experiment 1 and Experiment 2), however in such *short-term* situations a delay of 0.5 s can cause further problems, especially when considering that the vehicle travelled about 18 m during this time and the total time-budget is 7 s. See Figure 48 for an overview of the t_{eyes} times in Experiment 1 and Experiment 2.

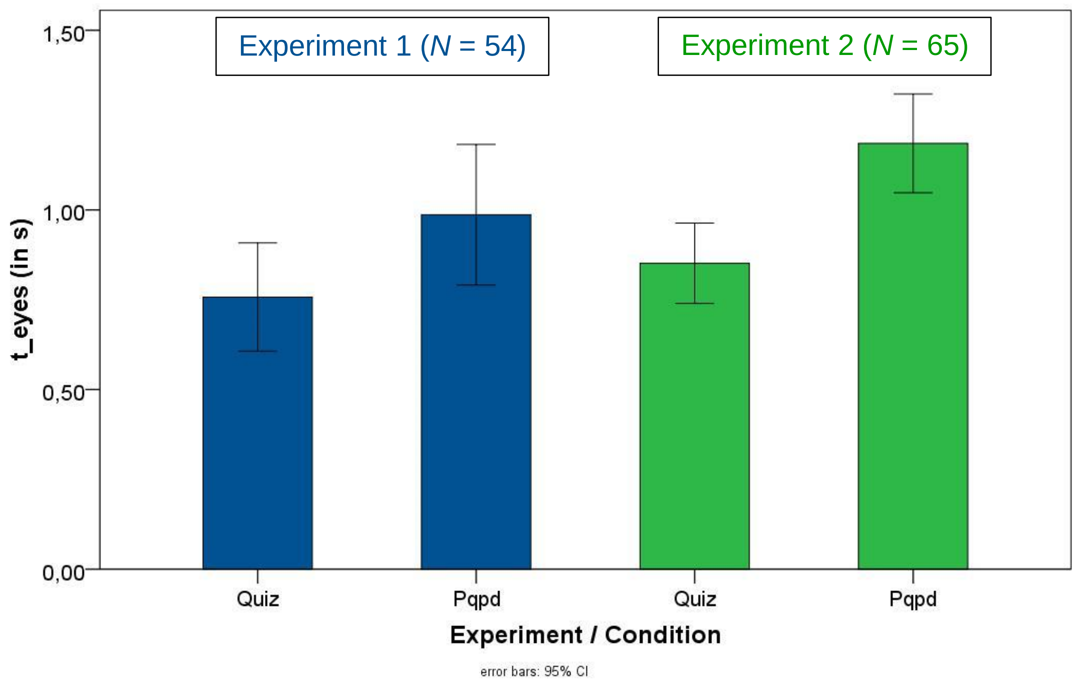


Figure 48: t_{eyes} for all experiments.

8.2.2. Hands-On Reaction Times (t_{hands}).

In this analysis t_{hands} reaction times of Experiment 1 and Experiment 2 in the take-over situation *accident on ego-lane* are compared.

In order to investigate which factor (i.e. NDRT and / or the duration of the ride) significantly affected the hands-on reaction times, a multifactorial ANOVA was calculated.

Results suggest, that the overall model cannot explain the resulting t_{hands} significantly, F (3, 108) = 1.78, p = .16, adjusted R^2 = .021, n = 112. The factor of the NDRT did not significantly affect t_{hands}, F (1, 108) = .77, p = .38. However, the factor of the duration of the ride, significantly affected t_{hands}, F (1, 108) = 4.52, p = .04, *partial* η^2 = .04. Accordingly, t_{hands} has increased with increasing driving time. However, the measured effect has to be classified as a small effect (Cohen, 1988).

There was no significant interaction effect between the factors NDRT and the duration of the ride, F (1, 108) = .17, p = .68.

The smallest t_{hands} reaction time was measured in the Experiment 1, when the participants had to deal with the fatiguing *Pqpd-task* (M = 1.00 s, SD = 0.42 s) whereas the slowest hands-on reaction time was measured in the Experiment 2 *Quiz-task* condition, (M = 1.32 s, SD = .63 s). See Table 25 and Figure 49 for further details and an overview of the achieved reaction times.

Table 25: Hands-on reaction times for all experimental rides.

Experi-ment	NDRT	*N* =	Environ-ment	Dura-tion	*M* (in s)	*SD* (in s)	*Statistic*	*p*	*Effect size*
Experi-ment 1	Pqpd Quiz	33 33	Simulator	25 min	1.00 1.14	.64 .42	*F* = .743	= .39	ηp^2 = .02
Experi-ment 2	Pqpd Quiz	28 28	Simulator	50 min	1.27 1.32	0.50 0.63	*F* = .128	= .72	$\eta p^2 < .01$

*Note. * p < .050, ** p < .010*

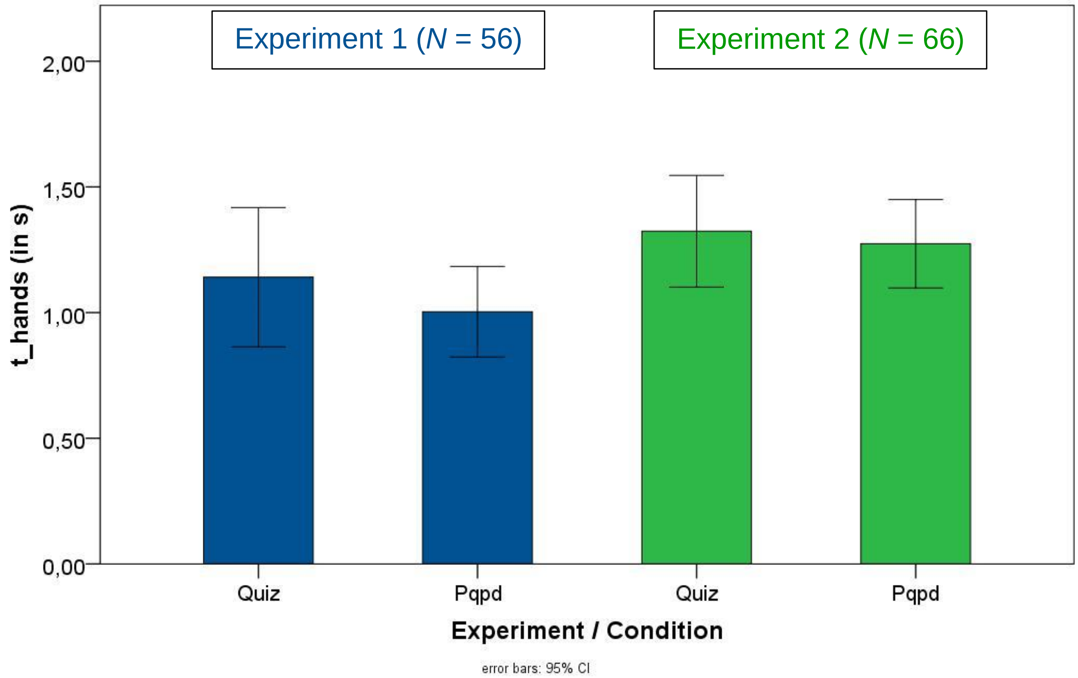


Figure 49: t_{hands} for the different NDRTs in the different experiments.

8.2.3. Braking Reaction Time (First Braking Reaction, $t_{brake\_reaction}$).

In this analysis the $t_{brake\_reaction}$ times recorded in the *accident on ego lane* situation in the experiments Experiment 1 and Experiment 2 are compared.

To investigate, which of the two factors of the NDRT and / or the duration of the automated ride significantly affected $t_{brake\_reaction}$, again a multifactor ANOVA was calculated.

Results suggest, that the overall model is significant, F (3, 118) = 3.33, p = .02, R^2 = .06, n = 122. The factor of the NDRT slightly missed a significant difference between the experimental conditions, F (1, 118) = 3.26, p = .07. However, the factor of the duration of the automated ride on the other side, revealed a significant difference, F (1, 118) = 3.99, p = .05, partial η^2 = .03. However, even if the effect size is rather small it can be stated that the braking reaction times decreased with increasing driving time.

A significant interaction effect between the NDRT and the duration of the ride could not be found, F (1, 118) = 2.26, p = .14.

The smallest and thus fastest mean braking reaction time was measured in the Experiment 2 *Quiz-task* condition (*M* = 1.71 s, *SD* = 0.94 s) whereas the slowest $t_{brake\_reaction}$ time was measured in the Experiment 1 *Pqpd-task* condition (*M* = 2.42 s, *SD* = 1.17 s). See Table 26 for the braking reaction times ($t_{brake\_reaction}$) recorded in the *accident on ego lane* take-over situations in Experiment 1 and Experiment 2. In Figure 50 further details are displayed.

Table 26: Reaction times for the first braking reaction for Experiment 1 and Experiment 2.

	Experiment	NDRT	N =	Environment	Duration	*M* (in s)	*SD* (in s)	*Statistic*	*p*	*Effect size*
	Experiment 1	Pqpd	33	Simulator	25 min	2.42	1.17	*F* = .036	= . 85	*ηp²* < .01
		Quiz	33			2.37	1.04			
*	Experiment 2	Pqpd	28	Simulator	50 min	2.33	0.97	*F* = 6.92	= . 01	*ηp²* = .10
		Quiz	28			1.71	0.94			

*Note. * p < .050, ** p < .010*

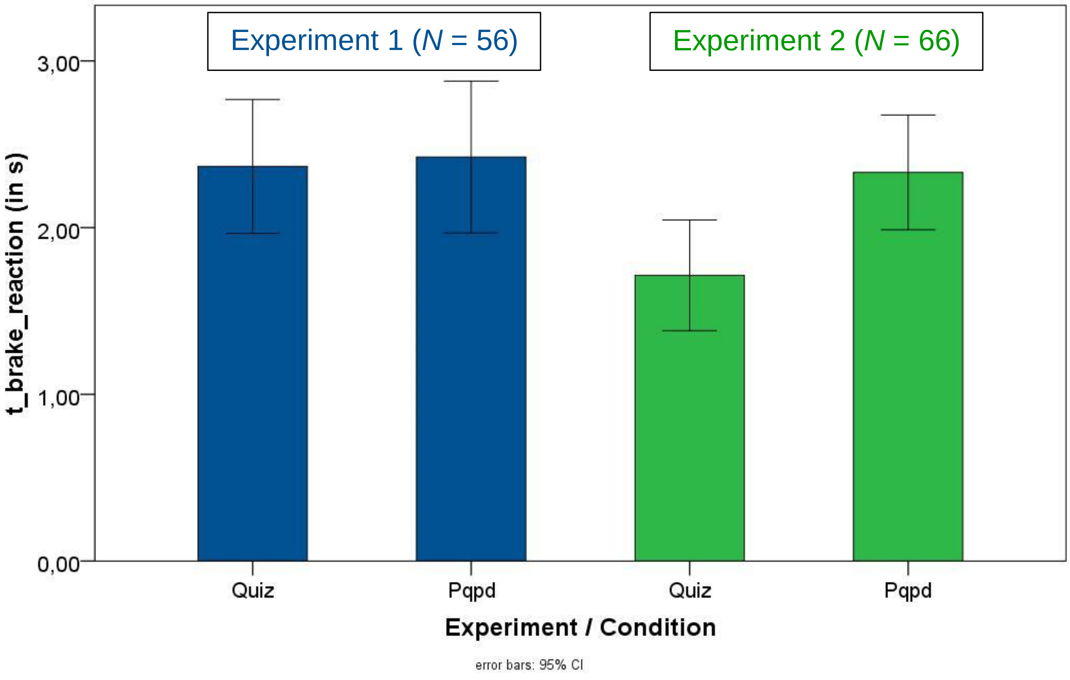


Figure 50: Overview of t_{brake} in Experiment 1 and Experiment 2.

8.2.4. First Driving Maneuver (First Reaction Out Of Steering-Wheel Input > 2° Or Braking Pedal Shift > 10%)

In this analysis reaction times for the first driving maneuver were compared. Therefore, either t_{steer} (steering-wheel input > 2°) or $t_{brake\_maneuver}$ (braking pedal shift > 10%) were considered, depending on which driving maneuver was first executed by the participant.

To figure out, if the factors duration of the ride or the NDRT affected the first driving maneuver time, a multifactor ANOVA was conducted. The reaction times from the Experiment 1 and Experiment 2 have been included in the analysis.

The results suggest, that these two factors did not significantly predict the reaction time for the first driving maneuver in the overall model, F (3, 118) = 1.08, p = .36, R^2 = .002, n = 122. Neither the factor of the NDRT, F (1, 118) = 1.25, p = .27, nor the duration of the ride, F (1, 118) = .03, p = .86, significantly affected the reaction time for the first driving maneuver.

A significant interaction effect between the NDRT and the duration of the ride could not be found, F (1, 118) = 2.22, p = .14.

The shortest reaction time for the first driving maneuver was measured in the Experiment 1 in the *Pqpd-task* condition, M = 1.97 s (SD = 0.76 s). The slowest reaction time for the first driving maneuver was in the Experiment 1 in the *Quiz-task* condition, M = 2.33 s, SD = 0.9 s. See Table 27 for further details and all driving maneuver times in the take-over situations. In Figure 51, the reaction times for the first driving maneuver are displayed.

Table 27: Reaction times for the first driving maneuver according to the different conditions.

Experiment	NDRT	*N* =	Environment	Duration	*M* (in s)	*SD* (in s)	*Statistic*	*p*	*Effect size*
Experiment 1	Pqpd Quiz	33 33	Simulator	25 min	1.97 2.33	0.76 0.90	F = 2.61	= .11	ηp^2 = .05
Experiment 2	Pqpd Quiz	28 28	Simulator	50 min	2.20 2.14	0.58 0.78	F = 0.92	= .76	ηp^2 < .01

*Note. * p < .050, ** p < .010*

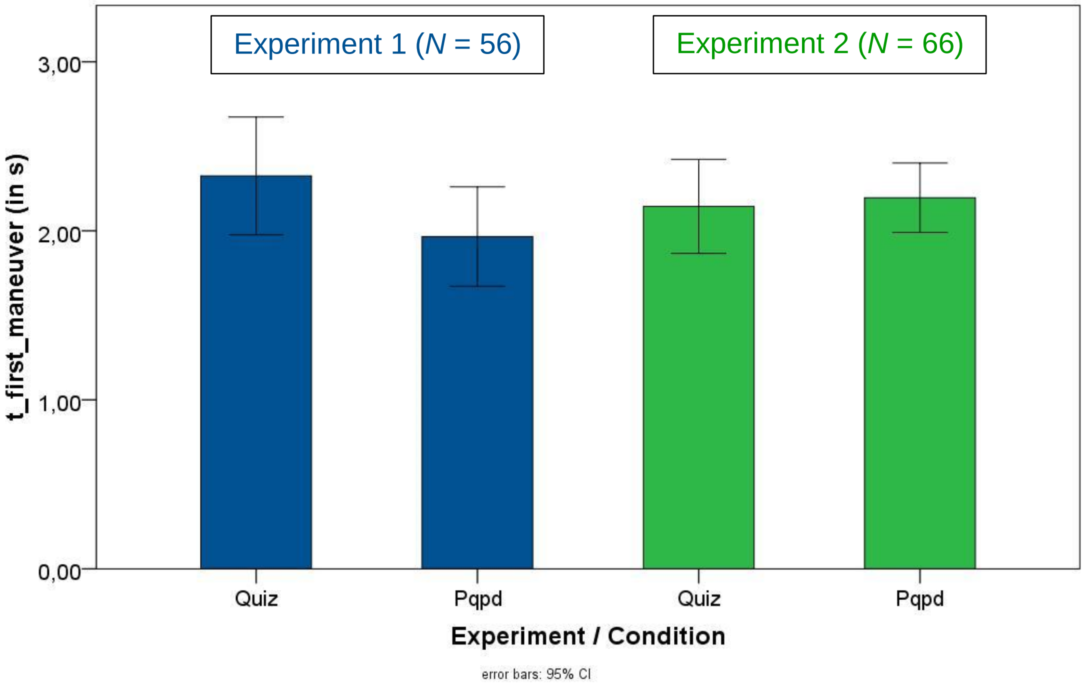


Figure 51: Reaction times for the first driving maneuvers in Experiment 1 and Experiment 2.

8.3. Overall Findings On Take-Over Performance. Differences In The Quality Of The Drivers' Intervention.

Next to the above reported reaction times, also driving performance parameters can be considered for a take-over performance assessment. Again, the two experiments in which the participants had to perform a take-over reaction after either engaging in a *Quiz-task* or a monotonous monitoring task (*Pqpd-task*) for either 25 min (Experiment 1) or 50 min (Experiment 2) have been included in the analysis. Again, the *accident on ego-lane* situation was considered.

8.3.1. $Acc_{\_long\_max}$ In The Different Experiments

One of the measurements which indicates if the human driver took over control in a qualitatively good manner is the $Acc_{\_long\_max}$. This measurement indicates how strong the human driver accelerated in the take-over situation. As the participants braked in this situation, negative accelerations were recorded. For the sake of clarity, the absolute values of the recorded accelerations were analyzed. Considered were the *accident on ego-lane* take-over situations of the Experiment 1 and Experiment 2.

To investigate, if the factor of the NDRT and / or the factor of the duration of the ride significantly affected $Acc_{\_long\_max}$ a multifactor ANOVA was calculated.

Results suggest, that these two factors significantly affected $Acc_{\_long\_max}$ when considering the overall model, F (3, 118) = 7.73, p < .001, R^2 = .14, n = 122. The analysis further revealed, that the factor NDRT did not significantly affect $Acc_{\_long\_max}$, F (1, 118) = .96, p = .32 but that the factor duration of the ride had significant effects on $Acc_{\_long\_max}$, F (1, 118) = 22.2, p < .001, *partial* η^2 = .16. According to these results, participants from the prolonged automated ride experiment, Experiment 2, reacted by stronger braking reactions compared to the participants from the shorter automated ride, Experiment 1.

There was no significant interaction effect between the NDRT and the duration of the ride on $Acc_{\_long\_max}$, F (1, 118) = .03, p = .86.

The smallest longitudinal acceleration was measured in the Experiment 1 *Quiz-task* condition, M = 5.58 m / s², SD = 3.59 m / s². The largest acceleration was measured in the Experiment 2 *Pqpd-task* condition, M = 8.69 m / s², SD = 2.10 m / s². According to Gail, Lorig, Gelau, Heuzeroth, and Sievert (2001), an emergency braking maneuver with a high deceleration rate corresponds to values of about 7 m / s². See Table 28 as well as Figure 52 for further details according to the $Acc_{\_long\_max}$ in the two experiments.

Table 28: Longitudinal accelerations in the take-over situations for the different conditions.

Experiment	NDRT	*N* =	Environment	Duration	*M* (in m/s²)	*SD* (in m/s²)	Statistic	*p*	Effect size
Experiment 1	Pqpd Quiz	28 28	Simulator	25 min	6.22 5.58	3.68 3.59	*F* = .434	= .51	*ηp²* < .01
Experiment 2	Pqpd Quiz	33 33	Simulator	50 min	8.69 8.25	2.10 2.55	*F* = .593	= .44	*ηp²* < .01

*Note. * p < .050, ** p < .010*

In Experiment 1 and Experiment 2, where the NDRTs were used to affect the fatigue state of the driver, no significant differences between the task-groups for $Acc_{\_long\_max}$ could be found. However, it seems like the participants from the fatiguing *Pqpd-task* braked a bit harder compared to the participants from the activating *Quiz-task* condition. However, when considering the effect sizes, only a small effect was recorded.

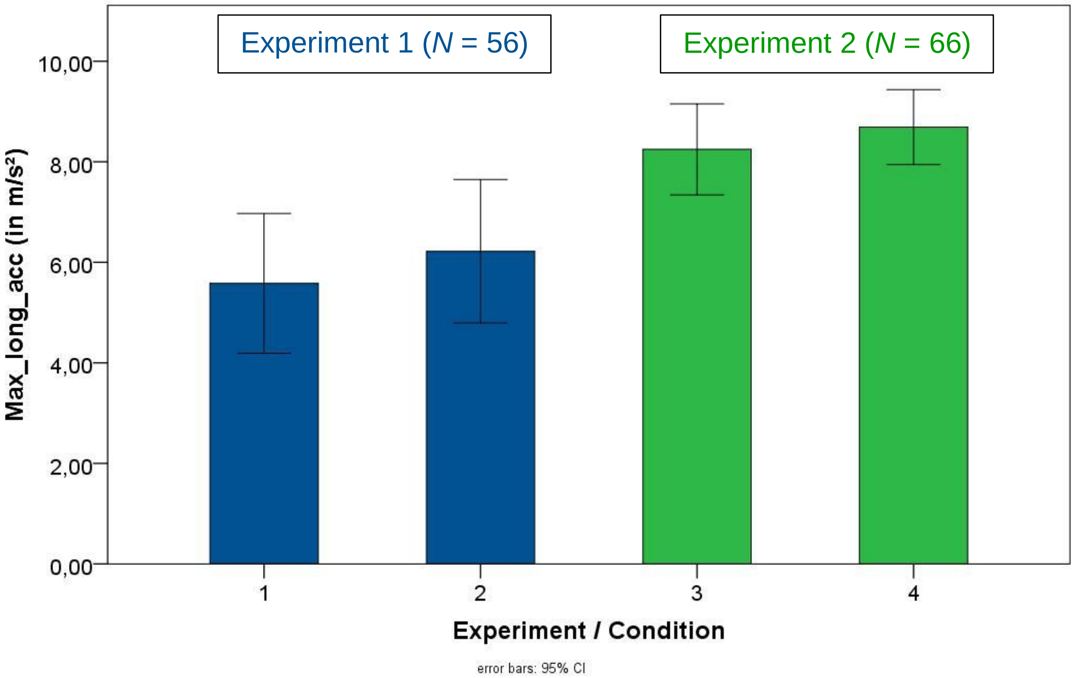


Figure 52: Overview of $Acc_{\_long\_max}$ values of Experiment 1 and 2 (absolute values).

8.3.2. Mean Maximum Lateral Accelerations The Different Experiments

Next to the acceleration in the longitudinal direction, the $Acc_{\_lat\_max}$, was recorded in the take-over situations. $Acc_{\_lat\_max}$ indicates, with which intensity participants steered after the RtI. Higher values indicate a more dynamic maneuver. In this analysis, the *accident on ego-lane* situations from Experiment 1 and Experiment 2 were considered.

To investigate whether the NDRT and / or the duration of the ride significantly affected $Acc_{\_lat\_max}$, as in the analysis before a multifactor ANOVA was used.

The results of the ANOVA suggest, that $Acc_{\_lat\_max}$ values differed significantly between the different experimental conditions and according to the NDRT and the duration of the ride, F (3, 118) = 2.29, $p < .001$, $R^2 = .17$, $n = 122$. The effect of the factor NDRT was not significant, F (1, 118) = 2.45, $p = .12$. However, the effect of the duration of the ride revealed a significant difference between the experimental conditions, F (1, 118) = 25.43, $p < .001$, *partial* $\eta^2 = .18$.

There was no significant interaction effect between the factors NDRT and duration of the ride, F (1, 118) = .01, $p = .91$.

As can be seen in Table 29, the lowest $Acc_{\_lat\_max}$ value was measured in Experiment 1 in the *Pqpd-task* condition, M = 1.3 m / s², SD = 0.84 m / s², whereas the highest $Acc_{\_lat\_max}$ values were recorded in the Experiment 2 in the *Quiz-task* condition, M = 3.62

m / s², *SD* = 2.71 m / s². Only in the Experiment 1, significant differences in the $Acc_{\_lat\_max}$ were found according to the NDRTs: the participants from the *Pqpd-task* steered less pronounced compared to the *Quiz-task* participants. However, the effect is rather small (Cohen, 1988). In Table 29 and Figure 53 the results are displayed in more detail.

Table 29: $Acc_{\_lat\_max}$ values for the different experimental conditions.

	Experiment	NDRT	*N* =	Environment	Duration	*M* (in m/s²)	*SD* (in m/s²)	Statistic	*p*	Effect size
*	Experiment 1	Pqpd Quiz	28 28	Simulator	25 min	1.30 1.89	.84 .85	*F* = 6.85	= .01	ηp^2 = .11
	Experiment 2	Pqpd Quiz	33 33	Simulator	50 min	3.11 3.62	2.28 2.71	*F* = .68	= .41	ηp^2 < .01

Note. * *p* < .050, ** *p* < .010

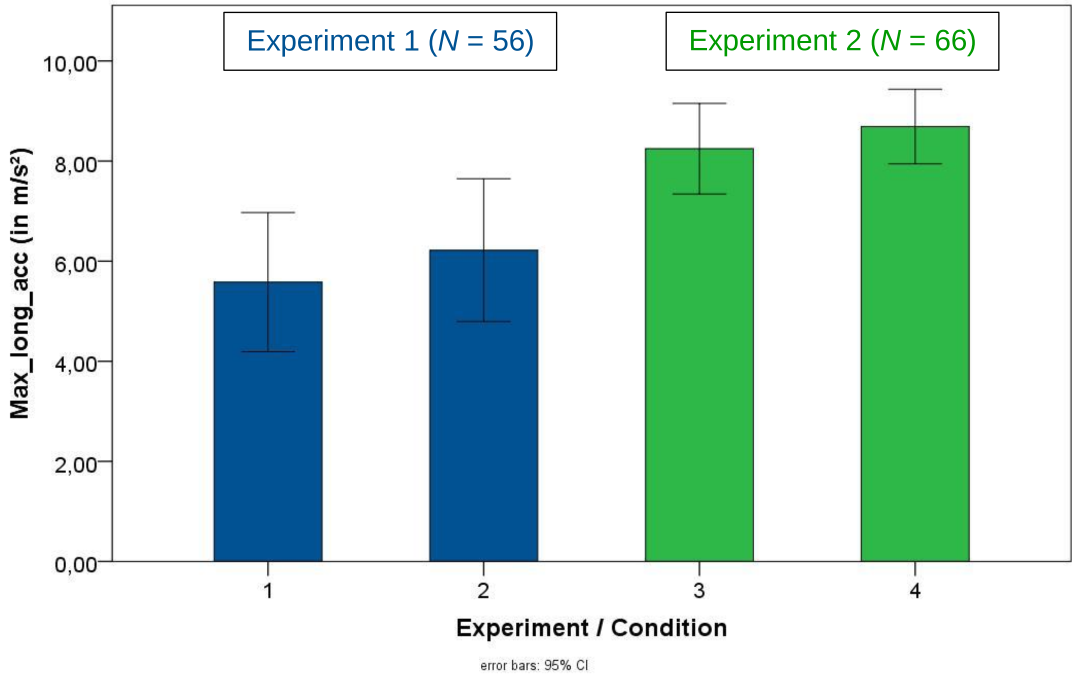


Figure 53: Overview of $Acc_{\_lat\_max}$ values in Experiment 1 and 2.

8.3.3. Comparison Of The $TTC_{\_MIN}$ For The Different Experimental Rides

Next to the reaction times and accelerations in longitudinal and lateral direction, $TTC_{\_MIN}$ were recorded in the take-over situations. $TTC_{\_MIN}$ indicates, how much time remains until the ego vehicle collides with the obstacle ahead, or in the case of the experiments discussed here, the *accident on the ego lane* in front. Thus, a high $TTC_{\_MIN}$ indicates a less dangerous situation for the driver. The other way around, a low $TTC_{\_MIN}$ indicates a dangerous situation for the driver.

For this analysis the $TTC_{\_MIN}$ values of the Experiment 1 and the Experiment 2 *accident on ego-lane* take-over situation were compared. Cases involving an accident (i.e. $TTC_{\_MIN}$ = 0 s) were not included in the analysis.

To identify, if the factor of the NDRT and the factor of the duration of the ride significantly affected $TTC_{\_MIN}$, a multifactor ANOVA was conducted.

The results suggest, that these two factors significantly affect $TTC_{\_MIN}$ in the overall model, F (3, 109) = 3.61, p = .02, R^2 = .07, n = 113. When considering the main effect of the factor NDRT, no significances could be found, F (1, 109) = 2.31, p = .13. However, the factor of the duration of the ride significantly affected $TTC_{\_MIN}$, F (1, 109) = 7.4, p = .008, *partial* η^2 = .064. Interestingly, the participants that had to drive automated for a longer period of time (Experiment 2) reacted in a way that the $TTC_{\_MIN}$ was higher compared to the Experiment 1 participants. However, participants from the Experiment 2 braked more strongly compared to the Experiment 1 participants (see section 8.3.2.).

A significant interaction effect between the duration of the ride and the NDRT could not be found, F (1, 109) = .93, p = .34.

The highest $TTC_{\_MIN}$ was recorded in the Experiment 2, when the participants had to deal with the *Pqpd-task*, M = 2.98 s, SD = 1.19 s. The lowest $TTC_{\_MIN}$ was measured in Experiment 1, when participants had to deal with the *Quiz-task*, M = 2.14 s, SD = 0.7 s. This parameter indicates that participants from the activating task group in the experiment with the shorter automated driving times had shorter $TTC_{\_MIN}$ and thus experienced potentially more hazardous situations. However, thus these participants also were those with the most time spending for their execution of the driving maneuver. With regard to the model of Marberger et al. (2017) these participants maybe used the additional time for preparing and executing the take-over maneuver. See Table 30 for the different $TTC_{\_MIN}$ achieved in the experimental rides.

Table 30: $TTC_{\_MIN}$ values for the take-over situations in the experiments.

	Experiment	NDRT	*N* =	Environment	Duration	*M* (in s)	*SD* (in s)	Statistic	*p*	Effect size
*	Experiment 1	Pqpd Quiz	28 27	Simulator	25 min	2.63 2.14	1.00 0.70	F = 4.37	= .04	*ηp²* = .08
	Experiment 2	Pqpd Quiz	29 31	Simulator	50 min	2.98 2.87	1.19 1.19	F = .126	= .72	*ηp²* < .01

*Note. * p < .050, ** p < .010*

As you can see in the table above, the $TTC_{\_MIN}$ values significantly differed in the Experiment 1 between the different NDRTs. In this experiment, the *Pqpd-task* participants had significant higher $TTC_{\_MIN}$ values compared to the *Quiz-task* participants. The effect size is medium. This effect could not be found in the Experiment 2.

For an overview of the recorded $TTC_{\_MIN}$ values in the *accident on ego lane* situation in Experiment 1 and Experiment 2, see Figure 54.

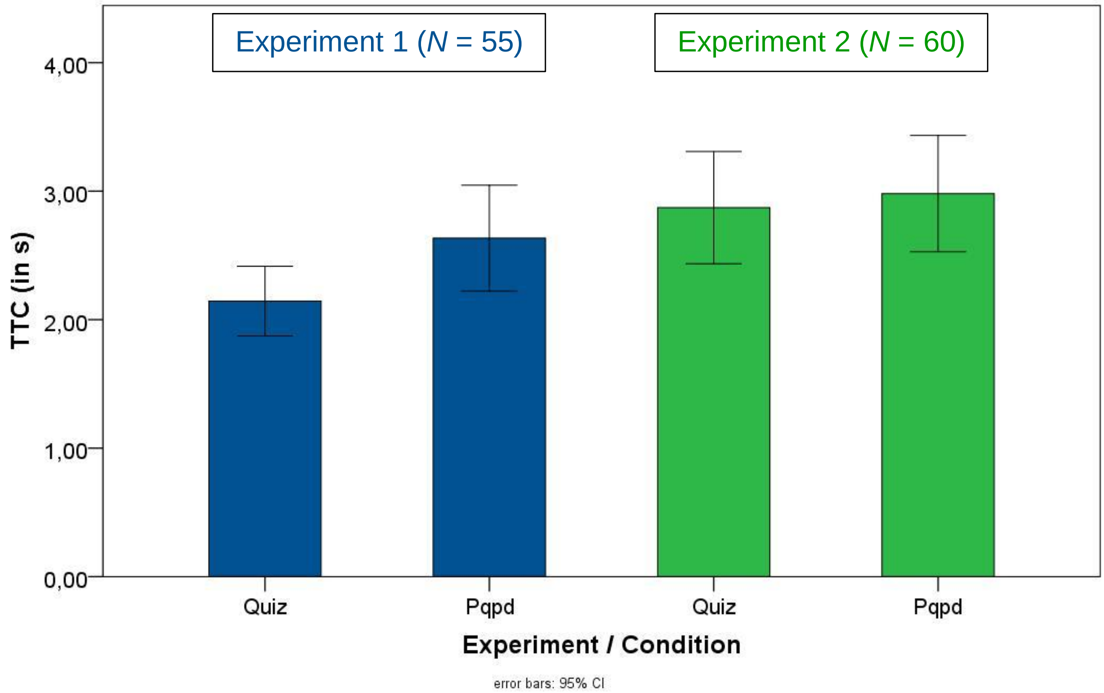


Figure 54: $TTC_{\_MIN}$ values for the different experimental rides in Experiment 1 and Experiment 2.

8.4. Overview Of The Findings

In this part of the thesis, an overview of the factors that significantly affected either the driver state or take-over performance is displayed.

For the overview of the drivers' state, Experiment 1, Experiment 2 (both were conducted in the motion base driving simulator) and the Experiment 3 (*Wizard-of-Oz* experiment) were considered. Thus, three different factors that could have affected the drivers' state have been included in the analyses: *duration of the ride* (*short* in Experiment 1 vs. > *long* in Experiment 2 and Experiment 3), *NDRT* (*activating* vs. *passive*) and the *environment* (*driving simulator* vs. *Wizard-of-Oz*). Subjective KSS and objective PERCLOS were analyzed.

For the analysis of the take-over performance measurements, the take-over situations in Experiment 1 and Experiment 2 were taken into account. In both experiments, the participants were confronted with the same take-over situation, whereby it can be analyzed if either the *duration of the ride* or the *NDRT* the participants had to engage in, affected take-over performance. As both experiments were conducted in the motion base driving simulator, the factor of the environment could not be analyzed regarding significant differences. Reaction times (t_{hands}, t_{eyes}, $t_{brake\_reaction}$ and *time for first driving maneuver*) as well as measurements indicating the quality of the drivers' input ($Acc_{\_long\_max}$, $Acc_{\_lat\_max}$ and $TTC_{\_MIN}$) were considered. Experiment 3 did not contain any take-over situation and therefore was excluded from the analysis of take-over performance measures.

It has to be stated, that the results are not based on causal relationships, but instead should be seen as descriptive as only the factors that have been manipulated were considered in the analysis. Other factors, like for example the season / the time of the year, which may also have affected the measures, have not been included in the analysis.

Table 31: Overview of the findings.

		Measurement	Duration of the ride (25 vs. 50 min)	NDRT (activating / fatiguing)	environment
Driver state measures		KSS	n.s.; $p = .43$	** $p < .001$, *partial* $\eta^2 = .2$	n.s.; $p = .64$
		PERCLOS	n.s.; $p < .84$.	** $p < .001$, *partial* $\eta^2 = .19$	** $p = .003$, *partial* $\eta^2 = .041$
Take-over performance measures	reaction times	t_{eyes}	* $p = .04$, *partial* $\eta^2 = .04$	** $p < .001$, *partial* $\eta^2 = .12$	
		t_{hands}	* $p = .04$, *partial* $\eta^2 = .04$	n.s.; $p = .38$	
		$t_{brake\_reaction}$	* $p = .05$, *partial* $\eta^2 = .03$	n.s.; $p = .07$	
		First driving maneuver	n.s.; $p = .86$	n.s.; $p = .27$	
	quality measures	$Acc_{\_long\_max}$	** $p < .001$, *partial* $\eta^2 = .16$	n.s.; $p = .32$	
		$Acc_{\_lat\_max}$	** $p < .001$ *partial* $\eta^2 = .18$	n.s.; $p = .12$	
		$TTC_{\_MIN}$	** $p = .008$, *partial* $\eta^2 = .064$	n.s.; $p = .13$	

Note. * $p < .050$, ** $p < .010$

In the table above, all findings of the Experiment 1, Experiment 2 and Experiment 3 are summarized. As can be seen in this table, especially the duration of the ride significantly affected the take-over performance of the participants. Both categories, the measurements indicating the quality of the drivers input, as well as the reaction times revealed significant differences for the duration of the ride but no significances for the factor of the NDRT the participants had to engage in. Only in one reaction time, t_{eyes}, a significant difference contributed to the NDRT could be found. When considering the effect sizes, results suggest, that the reaction times were less affected than the measurements indicating the quality of the drivers' intervention. The effect sizes for the reaction times are medium compared to big effect sizes for the quality measurements.

The factor of the NDRT also resulted in significant differences. However, these differences could not be found in the driving performance measurements but rather in the driver state measurements that are an indicator for the fatigue state of the driver.

The development of fatigue while driving CDA, which was another focus of the experiments, on the other hand, was less affected through the factor of the duration of the ride but rather from the engagement in the respective NDRT or activity.

Concluded it can be said, that an engagement in a passive task quickly can lead to a state of fatigue in CDA. However, this state of fatigue increasingly becomes problematic with increasing driving time. In the experiments reported in this thesis, a take-over situation after an automated driving time of 25 min seemed to be feasible for the drivers. Things changed, when the automated ride took about 50 min when the participants were confronted with a take-over situation and previously had to engage in the predefined activities or tasks for the entire ride. More than 10 % of the participants were not able to adequately take over control of the vehicle and crashed or lost control of the vehicle.

Results suggest, that prolonged automated driving can go along with hazardous situations for the drivers. Especially after prolonged CDA rides, the human drivers need to be supported by ADAS when it comes to a take-over situation.

8.5. TOC-Ratings For The Take-Over Situations In Experiment 1 And Experiment 2

Next to the reaction times and the measures for the quality of the drivers' intervention, all take-over situations (in the *accident on ego lane situation*) were rated by three trained raters by using the TOC-rating tool. This tool was developed in order to be able to evaluate a take-over performance in one integrated rating, as it has been shown in the past that an isolated analysis of measured values such as reaction times or driving parameters are not very indicative. Different dimensions such as longitudinal guidance, lateral guidance, securing behavior, expression of the driver etc. are included in this overall assessment. See section 3.4.3. for a detailed explanation of the tool.

In Table 32 the different TOC-ratings of Experiment 1 and Experiment 2 (in the *accident on ego lane* take-over situation) are displayed. Additionally, different statistical evaluations are given. The TOC-rating covers a range of numbers from 1 - 10 whereby a *1* indicates a very good take-over performance (i.e. *perfect*) and a *10* indicates a very poor take-over performance (i.e. *loss of control / crash*).

To identify, whether the NDRT or the duration of the ride significantly affected the TOC-rating, four conditions (duration of the ride: short / long; NDRT: Quiz / Pqpd) were compared using the non-parametric Kruskal-Walis-test.

The results revealed no significant differences between the different experimental conditions, $Chi^2 = .97$, $p = .8$.

See Table 32 for the TOC-ratings in the different experimental conditions.

Table 32: TOC-ratings of the take-over situation in Experiment 1 and Experiment 2.

Experiment	NDRT	*N* =	Environment	Duration	*Md*	*SD*	Statistic	*p*	*Effect size*
Experiment 1	Pqpd	28	Simulator	25 min	5.00	1.71	*U* = 332.5	= .93	*r* = .01
	Quiz	28			5.00	1.58			
Experiment 2	Pqpd	33	Simulator	50 min	4.67	2.21	*U* = 460.0	=.38	*r* = .11
	Quiz	33			5.17	1.92			

Note. * $p < .050$, ** $p < .010$

As can be seen from the results in the table above, the subjective rater evaluations within one experiment did not differ significantly between the two task-groups. Take-over performances were best rated in the Experiment 2 in the *Pqpd-task* condition, *Md* = 4.67, whereas the worst ratings were assigned in the same experiment, in the *Quiz-task* condition, *Md* = 5.17.

These findings would indicate that the participants who were in the *Quiz-task* condition in Experiment 2 reacted worst, although the participants in the *Pqpd-task* condition caused more accidents and uncontrollable events and basically reacted worse (when for example considering t_{eyes} or $t_{brake\_reaction}$). This clearly reveals the problem of the TOC-rating: As soon as a participant reacts by a braking maneuver (as it was the case in the *Pqpd-task* condition in Experiment 2), many evaluation categories of the TOC-rating are no longer taken into account for the assessment (e.g. categories like *Lateral vehicle control* or *Lane change / Lane choice*). For participants who react with a steering maneuver, more categories can be included in the assessment, which makes a good result in the TOC-rating less likely.

In Figure 55, the TOC-ratings of the take-over situations involving an *accident on ego-lane* situation in Experiment 1 and Experiment 2 are displayed in a boxplot diagram for further information. No big difference between the two experiments can be seen. However, it looks like especially the higher TOC-values could be more often observed in the Experiment 2 that goes along with the prolonged driving time.

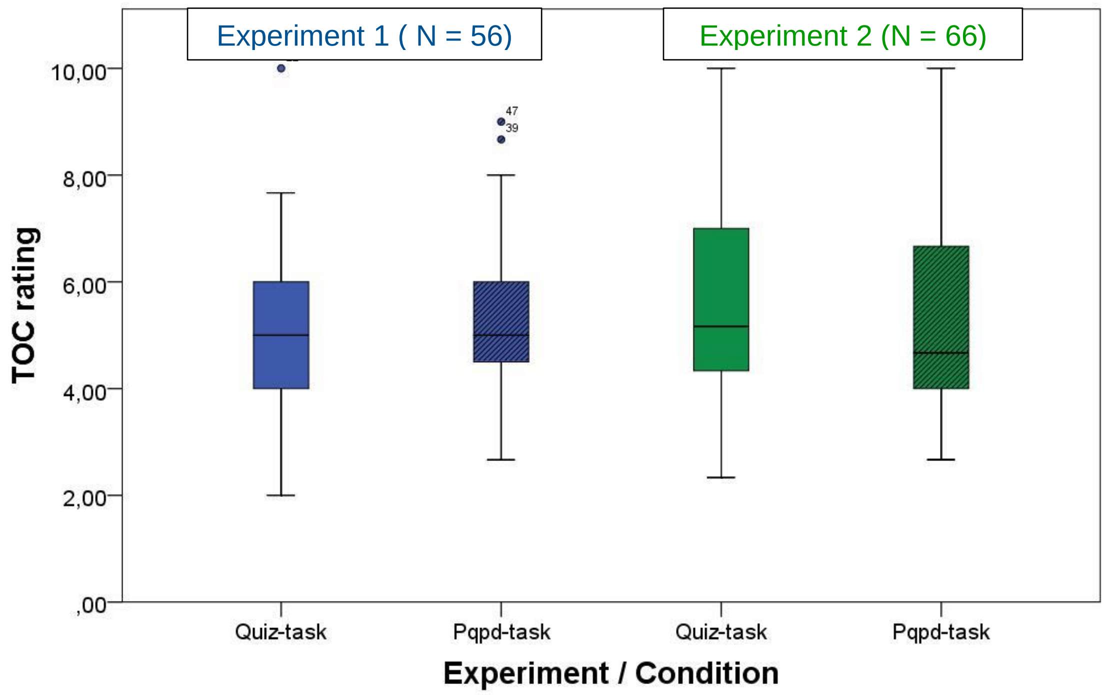


Figure 55: TOC-ratings for the take-over situations in the different experiments.

8.6. Occurrence Of Dangerous Situations And Accidents

In all experiments which contained take-over situations, some of the participants were not able to take over and control the vehicle in an adequate manner. Some lost control of the vehicle and skidded, others collided with the accident situation or other road users and some others were only able to solve the situation by putting themselves or others at serious danger.

The participants who were involved in an accident or were only able to solve the situation with an endangerment are easy to identify using the TOC-rating. Test persons who were involved in an accident were rated with a *10* and the participants that could only solve the situation by an endangerment have TOC-rating scores between *7* and *9*. These participants were taken into account in the following graphic, Figure 56. In this figure, all participants that were rated with a TOC-rating score between *7 – 9* (*endangerment*; orange) or a *10* (*accident / loss of control*; red) in Experiment 1 and Experiment 2 are illustrated. The Experiment 4 is discussed separately as in this experiment it should be investigated how the human driver can best be supported in a take-over situation. Therefore the results are less comparable with those from the Experiment 1 and the Experiment 2.

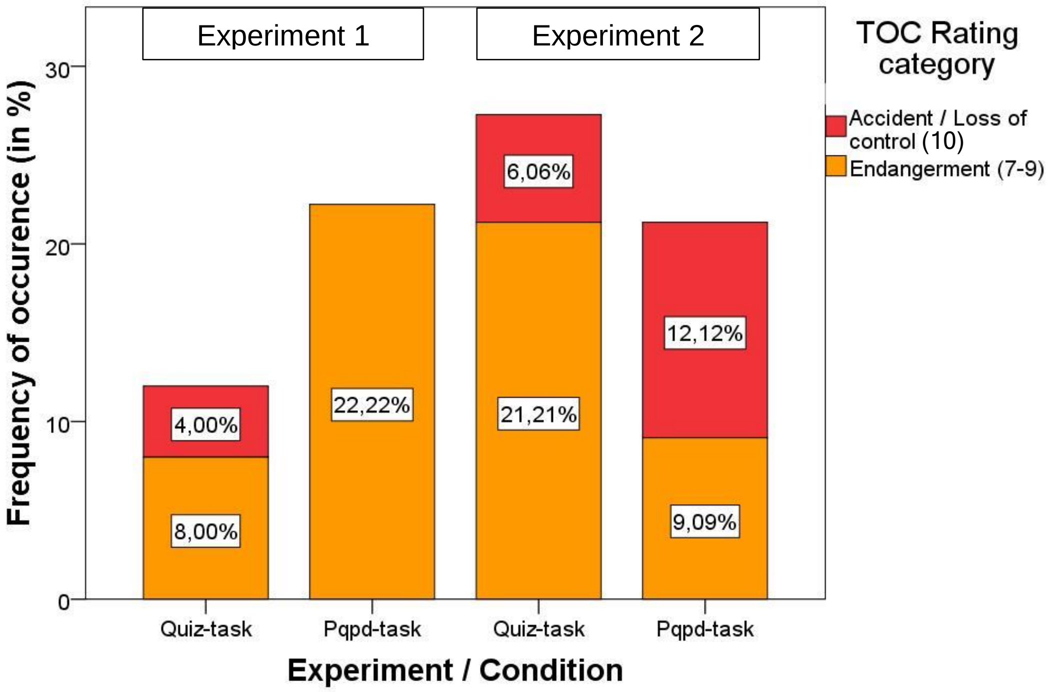


Figure 56: Overview of endangerments and not controllable events in the different experiments.

As can be seen in the figure above, in all experiments and all conditions endangerments in the take-over performance have occurred. However, accidents and losses of control (i.e. TOC-rating *10*) have not occurred in all experimental conditions. The highest number of accidents occurred in the Experiment 2. Especially in the *Pqpd-task* condition, a high number of participants (*n* = 4; 12 %) was not able to regain control of the vehicle after 50 min of the automated drive. In the *Quiz-task* condition, *n* = 2 (6 %) participants were not able to regain control after the RtI. Thus, a total of $n = 6$ participants (9.1 %) in the Experiment 2 were not able to control the situation compared to $n = 1$ (2 %) in the Experiment 1. This indicates that a prolonged automated driving time with a subsequent take-over situation can lead to serious problems and endanger the safety in CDA.

However, the system for CDA that was simulated in the experiments was designed to measure how human drivers react in take-over situations and if they are feasible of taking over control even after prolonged driving with the system. Therefore, to be able to measure human performance, when the RtI was issued no driver assistance supported the driver in the take-over situation. Additionally, the tasks that had to be processed by the participants do not correspond to natural tasks people would like to do when driving automated. However, a pure monitoring of automated driving systems is also likely to induce states of passive fatigue to the drivers.

In order to investigate how to best support the human driver in *short-term* take-over situations, another experiment was conducted that focused on exactly this topic. See section 8.7. for an overview of the results from this experiment.

8.7. Potential Of Different HMI Concepts – A Possibility To Support The Driver In Take-Over Situations

In contrast to the other experiments, in Experiment 4, the independent variable was represented through the HMI concept which warns the driver in case of a take-over situation. Therefore three different concepts were tested against each other. A *LED-concept*, a *SPEECH -concept* and a *BASELINE-concept* were used in a between-subject design to warn the driver in the respective take-over situations.

Each participant had to solve two different take-over situations. In the first situation, the ego-lane was blocked by an accident. However, a lane-change maneuver was possible, as the lanes left of the ego-vehicle were free of other road users. Thus, a lane-change maneuver was the appropriate reaction in this situation. Also in the second situation the ego-lane was blocked by an accident. However, in this situation, the lane left was blocked trough a number of other road users. In this situation, a braking reaction was the appropriate reaction of the driver to solve the situation.

In all recorded reaction times upon the RtI, the *LED-concept* could help to improve the reactions of the drivers compared to the other two concepts. The *SPEECH-concept* on the other hand side, could not improve the reactions of the drivers. Rather a negative effect must be assumed when considering the *SPEECH-concept*. Compared to the *BASELINE-concept*, all measured reaction times were slower in the *SPEECH-concept* condition.

Also in the measurements of the quality of the drivers' input, the *LED-concept* had benefits compared to the other two concepts. The $TTC_{\_MIN}$ was highest in the *LED-concept* group in both situations. In the situation, in which a braking maneuver was required to solve the situation, participants of the *LED-concept* group showed the highest mean $Acc_{\_long\_max}$. In the other situation, in which a lane-change maneuver was required, $Acc_{\_long\_max}$ did not differ between the concepts significantly. The measurement $Acc_{\_lat\_max}$ did not differ significantly according to the different HMI concepts. However, the lowest steering-wheel inputs were measured in the *LED-concept* group in both situations.

Concluded it can be said, that the *LED-concept* helped to improve the reaction times of the drivers. Especially in the situation, in which a braking input was required, this concept had benefits compared to the other concepts.

9. Conclusion

In the following last section a final summarization of the results of the experiments that were conducted as part of the thesis is given and the research questions will be answered in detail.

The current thesis addressed how NDRTs affect the drivers' state and the drivers' performance in *short-term* take-over situations in CDA. In particular, the following research questions should be answered (see also section 2.9.):

1. How do NDRTs affect the drivers' state during CDA?

2. How can emerging fatigue be detected in CDA?

3. How does fatigue affect take-over performance of the driver, when a short term take-over situation occurs?

During the research-process and based on the results of the first experiments, further research-questions have arisen:

4. Does fatigue emerge in real traffic environment on-road as quickly as in the driving simulator when driving CDA?

5. How can the human driver be best supported in *short-term* take-over situations?

These have been evaluated during the course of the present thesis as well.

To investigate the topic of the current thesis four experiments were conducted. The overall sample consisted of N = 235 participants. In total, the participants experienced 16790 min of CDA and thus also a similar time of engaging in predefined NDRTs or activities. In the context of the thesis, take-over behavior of the participants in 313 situations was investigated. The experimental CDA rides each took between 25 min – 60 min. Three of the experiments were conducted in a motion base driving simulator and one in a *Wizard-of-Oz* vehicle.

The first question that should be addressed in the context of this thesis was whether NDRTs in CDA affect the state of the driver. To answer this question and to investigate effects of different NDRTs, the drivers' fatigue state was assessed at different points of time during all experimental rides using subjective (KSS) as well as objective methods

(PERCLOS). In the experimental rides, the participants were assigned to different NDRTs, which should have affected the state of the drivers.

On the one hand, a monotonous monitoring task was used to induce *passive task-related fatigue* (*Pqpd-task*) amongst the participants. Due to increasing monotony and the new role for the human driver in CDA, especially this form of fatigue is expected to occur in automated driving (May & Baldwin, 2009; Neubauer, Matthews, Langheim et al., 2012).

On the other hand, NDRTs or activities that should prevent the human drivers from emerging fatigue were used. Therefore either a classic *Quiz-task* or a *Free-choice-activity* were selected.

Thus, with the help of these different NDRTs and activities, the fatigue state of the drivers should be directly affected. This was also necessary to answer the other research questions: on the one hand to be able to investigate how emerging fatigue in CDA can be measured and detected, and on the other hand to be able to investigate to what extent the drivers' state can affect the performance of the human driver in a take-over situation.

The results of Experiment 1 have shown that fatigue in CDA can occur quickly and that emerging fatigue is closely linked to the respective NDRT the human driver works on while driving automated. Both, subjective as well as objective fatigue measurements have increased during the rides in which the participants had to deal with the monotonous monitoring *Pqpd-task*, that was used to induce task-related fatigue. In the other rides, in which the participants had to deal with the activating tasks, fatigue did not emerge. Neither subjective nor objective fatigue has significantly increased in these experimental condition. In this experiment, the participants each experienced two rides, each lasting 25 min. In each of the two rides, they worked either on the *Pqpd-* or the *Quiz-task*.

The results of Experiment 2 confirmed the effects of the NDRTs on the drivers' state that were found in Experiment 1. The *Pqpd-task*, that was used to provoke *passive task-related fatigue* amongst the participants led to an increase in the objective PERCLOS as well as to an increase in the subjective KSS. On the other hand, participants that had to deal with the *Quiz-task*, which was designed to prevent participants from higher states of fatigue, PERCLOS did not increase significantly and rather stayed on a low level in the *Quiz-task* condition. However, the subjective fatigue measurement, KSS, could not confirm these findings. Participants that had to deal with the *Quiz-task* stated higher states of subjective fatigue, too. The difference in the KSS between the two NDRTs was not significant in this experiment.

A possible explanation for this result may be that participants have difficulties estimating their own subjective fatigue. This can be especially the case if you have to work on a predefined task, which can lead to boredom after a certain time. Next to that, it is also uncertain whether participants can distinguish between perceived fatigue and boredom.

Therefore, in Experiment 3 that was conducted in the Wizard-of-Oz vehicle the task that was used to prevent participants from emerging fatigue was amended. Instead of the predefined *Quiz-task*, participants had the possibility to work on freely chosen activities (*Free-choice activity*) during the automated ride. The NDRT that should lead to fatigue on the other hand remained the same, the *Pqpd-task*. As result it can be concluded, that the *Free-choice activity* could prevent participants from emerging fatigue. PERCLOS as well as subjective KSS did not increase in this task-condition at all and remained at one level during the ride. Participants that had to deal with the Pqpd-task on the other hand clearly showed signs of emerging fatigue. Compared to the results from Experiment 1 and Experiment 2, these participants even were the ones with the highest subjective as well as objective fatigue states.

Concluded it can be stated that the drivers' fatigue state in CDA indeed highly depends on the respective NDRT or the activity the driver is working on while driving automated. Especially preset NDRTs like permanent system-monitoring lead to fatigue quickly. On the other hand, it could be shown that the possibility of a Free-choice activity could almost completely prevent drivers' fatigue in CDA. In contrast to the experiments conducted by Omae et al. (2006) or Feldhütter et al. (2018) not one participant fell asleep while driving automated in all experiments presented in this thesis. In the experiments conducted by Omae et al. (2006) or Feldhütter et al. (2018), several drivers fell asleep, when they could not engage in any NDRT. Thus it can be said, that NDRT-engagement is preferable to doing nothing.

The second aim of this thesis was to investigate, how emerging fatigue in CDA can be detected and measured. Therefore, two methods were used to evaluate the fatigue state of the drivers. On the one hand, subjective KSS was questioned for different points of time during the experimental rides and on the other hand, objective PERCLOS was recorded over the course of the experiment.

In the first experiment, significant differences in the drivers' state could be found for both fatigue measurements. The measured KSS values as well as PERCLOS indicated higher fatigue states for the participants that had to deal with the *Pqpd-task* which should lead to *passive task-related fatigue* while driving automated. In Experiment 2, in which participants experienced prolonged automated rides of 50 min, however, the findings on emerging fatigue were less conclusive. Whilst the objective PERCLOS measurement again was significantly affected by the factor of the NDRT, a significant difference in the KSS between the two NDRT groups could not be found. This can again be attributed to the fact that an assigned activity can lead to boredom with increasing time-on-task. Thus it can be conceivable that participants rather rated their perceived boredom instead of their experienced fatigue. PERCLOS, on the other hand, seems to be supportive to determine the actual level of fatigue as it cannot be directly affected by the participants.

This leads to the question, which fatigue measurement really measures emerging fatigue: While the PERCLOS cannot be affected by the participants, the subjective KSS rating rather corresponds to a subjectively experienced state of the drivers and therefore seems to be inaccurate and erroneous. Thus, when an activity leads to boredom, this may also lead to higher KSS ratings as the participants tend to rate their subjective boredom and not their fatigue state. The objective PERCLOS on the other hand, cannot be affected by the participants and should therefore rather reflect the actual state of fatigue of the drivers.

Similar findings were also reported by Schmidt (2010). In his experiments, subjective fatigue assessments were compared to objective measurements. He claimed, that In contrast to the objective assessment, the verbal fatigue assessment of the drivers' state has to be interpreted with caution due to issues of validity and intrusion (Schmidt, 2010).

Therefore, subjective fatigue assessments in CDA do not seem to be appropriate for a usage in series production cars. On the one hand it is uncertain if the participants can estimate their actual fatigue level and on the other hand it is uncertain if fatigued drivers' would indicate their fatigue level to the system.

Things are different with PERCLOS. This measurement seems to be an appropriate method to detect and measure the fatigue state of the drivers in CDA. In all experiments significant differences in the PERCLOS measurement could be detected relating to the respective NDRTs.

The eye-tracking system that was used in the experiments was a professional labor-setting eye-tracking method. Additionally, an extensive preparation and post-processing of the recorded eye-tracking data was necessary in order to be able to make reliable statements about PERCLOS. Of course this is not conceivable for a series production car. Thus, for future usage in automated driving vehicles, a camera-based system integrated in the interior of the vehicle that can detect emerging fatigue based on PERCLOS or other eye-lid based methods gets strongly recommended.

How such a system can be integrated in series production cars, and how such a system can work with the required robustness has to be investigated in future work.

The third research question that should be answered within the framework of this thesis is whether fatigue in CDA affects the take-over behavior of the human drivers when being confronted with a *short-term* take-over situation. Therefore, in all experiments except Experiment 3, participants were confronted with take-over situations with TTC = 7 sec. These take-over situations occurred in the end of the rides, whereby higher fatigue states of the *Pqpd-task* participants should be achieved.

In Experiment 1, the take-over situations occurred after 25 min of the start of the rides. It can be stated that after this time-period the fatigue induced by the NDRTs had almost no effect on the take-over performance of the participants. In this experiment almost all participants could solve the take-over situation in a good manner, independently from their fatigue level. Relevant significant differences in take-over performance measurements including reaction times and measurements that are related to the quality of the drivers' intervention could not be found in this experiment. Only one participant from the *Quiz-task* group collided with another road user when executing a lane change maneuver.

In the future, however, longer periods, in which a CDA system takes over control of the vehicle, are conceivable. Therefore, in the second experiment, the time of the automated ride was extended to 50 min. The remaining part of the experimental setup was similar to Experiment 1 (concerning the NDRTs, the mock-up and the take-over situation).

Results of this second experiment suggest, that although the fatigue level of the participants did not further increase compared to the 25 min rides from Experiment 1, take-over performances were undoubtedly affected after the prolonged automated ride. Independently from the NDRT, the recorded reaction times upon the RtI were slower, when participants had to regain vehicle control after 50 min compared to the 25 min ride. Next to the differences in the reaction times, similar findings could be shown for the metrics that are related to the quality of the drivers input. Significant differences could be found for longitudinal and lateral acceleration measures as well as for the measured TTCs. Interestingly, the participants from Experiment 2 reached higher $TTC_{\_min}$ compared to Experiment 1.

In a first sight, higher TTCs go along with a lower risk for the human drivers as the measured time-period until they would collide with the obstacle is higher. However, when comparing the previous reactions of the drivers in the prolonged automated ride, it gets obvious, that the drivers took longer for their first gaze-reaction and all following steps until regaining vehicle control. From these observations it can be concluded, that the participants from the prolonged automated ride took more time for perceiving the take-over situation but in the following steps reacted faster and more extremely compared to the participants that experienced the shorter automated ride. Of course, faster reaction times also seem to be preferable compared to slower reaction times. However, a reaction, even if it is fast, will only be good when the reaction itself is appropriate in the current situation and the execution of the reaction is of good quality.

Another possibility to evaluate the take-over behavior is to look at the specific reactions the drivers fulfilled. In this second experiment, six participants could not control the vehicle within the available time and collided. Four of them previously had to engage in the monotonous Pqpd-task compared to two from the Quiz-task. This is in obvious contrast to only one crashed driver from the first experiment with a driving time of 25 min.

Additionally, video data was analyzed: When looking at the reactions of the drivers in the take-over situation in the prolonged experimental rides two things have attracted attention: on the one hand, participants that had problems in the take-over situations often intervened hastily and on the other hand, some others have fallen into a kind of shock-state, whereby they showed a fast first gaze-reaction but no further ongoing reactions. Both are known phenomenon that are related to panic or stress reactions. In such a take-over situation, when the participant gets suddenly torn from his current NDRT due to an acoustically and visual warning signal, this can lead to a rush of adrenaline. Well known consequences of such a sudden increase of adrenaline are impaired ability to perceive as well as to control the environment (Jamson & Smith, 2003). In other domains, panic reactions are also connected to either an complete absence of any reaction at all (Muir et al., 1996) or to an overcompensating reaction (Dingus et al., 1998).

Exactly such behavior patterns of the participants could be observed in the prolonged automated rides. Participants that reacted in an overcompensating manner when the RtI was triggered often lost control over the vehicle or collided with other road users.

First approaches on how to better support a human driver in such a situation were examined in Experiment 4. In such *short-term* take-over situations, especially a concept that supported the driver in his gaze reaction seem to be advantageous. A red flashing LED-light in the direction of the accident which caused the RtI led to improved reaction times and an improved take-over quality. On the other hand, a speech-output which warned the driver could not result in any improvements compared to a baseline concept.

Concluded it can be said that fatigue in CDA which occurred due to NDRT-engagement indeed can affect take-over performance of the human drivers. However, it seems like the factor of the task is less crucial than the factor of the duration of the ride in CDA. Especially after prolonged automated driving, the drivers had problems with regaining vehicle control.

So far, the main research questions of the thesis have been answered. However, in the course of the work further questions arose, which will be answered in the following.

One of these additional research questions was, if the measured fatigue level of the participants can rather be attributed to the environment of the driving simulator than on the automated driving and simultaneously engaging in NDRTs.

As Experiment 1 and Experiment 2 were designed to lead to fatigue (especially in the *Pqpd-task* conditions) as good as possible, the simulated driving scenario was monotonous. During the rides, participants experienced barely any take-over situations, the traffic density was low and also the weather conditions should rather cause than prevent fatigue. By using this setting, it could be shown that fatigue can occur quickly in CDA. However, things can look different in reality. The question that arose was: Does fatigue emerge in CDA in real driving environment as it was the case in the driving simulator experiments. To answer this question, a simulated CDA ride in real traffic environment was used with the help of a *Wizard-of-Oz* approach (Experiment 3).

For this purpose, participants were again subdivided into two NDRT groups. One group had to deal with the *Pqpd-task*, which had been used in Experiment 1 and Experiment 2 before, to induce *passive task-related fatigue*. The other group could freely choose their activity (*Free-choice-activity*) while driving automated. The experimental rides took about 60 min and during the entire ride, participants had to engage in the predefined NDRT.

For different times during the ride, fatigue was assessed by PERCLOS as well as KSS measurements.

Results of this experiment confirm the results of the first two experiments conducted in the driving simulator. Participants that had to engage in the monotonous monitoring *Pqpd-task*, were more fatigued compared to the participants that could freely choose their activity. This was obvious in PERCLOS as well as in the subjective KSS. The mean PERCLOS of the *Pqpd-task* group, which was over 20 % after 60 min of the ride, can be classified as a state of *fatigue* according to Wierwille et al. (1994). Also the subjective KSS after 60 min of the ride ranged in an area in which the ability to drive is clearly restricted (Åkerstedt, Anund, Axelsson, & Kecklund, 2014).

Contrary to what was expected, the measured fatigue states of the *Pqpd-task* participants were higher in this on-road experiment than in the driving simulator experiments before. A possible explanation for this observation can be that the participants were already fatigued due to manual driving to the starting point of the experimental session from which CDA / the wizard driver was available. The fatigue model of May and Baldwin (2009) states that *active task-related fatigue* (caused through manual driving) can increase *passive task-related fatigue*.

On the other hand, when participants could engage in the *Free-choice-activity*, the measured fatigue state was the lowest compare to all other experimental rides. These results clearly show the positive effects of a freely selectable activity in CDA on the fatigue state of the driver.

Concluded, it must be said, that fatigue in CDA may also quickly emerge in real driving environment on-road. The measured effects of a monotonous monitoring task (*Pqpd-task*) on the drivers' state in Experiment 1 and Experiment 2 are therefore not sole to be attributed to the driving simulator, but can also occur on-road in real driving environment. It must therefore be assumed that the take-over performance in real traffic environment can also be impaired in a similar way as it occurred in the driving simulator experiments.

On the other hand, the potential of freely choosable activities or NDRTs in CDA could be demonstrated. Participants from this task-group did not get fatigued at all.

Based on these results, a predefined activity, such as monitoring the automated driving system, cannot be recommended. This could be conceivable in earlier stages of driving automation such as *Partial Driving Automation* (SAE, 2018) or also in CDA when required by law.

This leads to the final research question, which should be answered in this thesis. In the experiments that were conducted as part of this thesis, especially after prolonged CDA rides, the take-over performance of the participants was impaired. As CDA itself can contribute to drivers' fatigue due to increasing monotony, HMI concepts that can support even fatigued drivers in take-over situations have to be designed. Therefore, in Experiment 4, three different HMI concepts were tested regarding their ability to warn and request the driver to intervene.

Therefore, a *LED-concept* which should direct the gaze of the driver as quickly as possible towards the obstacle, a *SPEECH-concept* which should warn the driver by a speech-output and a *BASELINE-concept* which warned the driver via a text in the instrument cluster were used. Each participant experienced only one of these concepts in two different *short-term* take-over situations. Again, take-over performance was measured in these situations in order to be able to make statements with regard to the benefits of these concepts.

Results suggest, that especially in *short-term* take-over situations a concept which directs the gaze of the driver is beneficial. With this concept, almost all reaction times of the drivers could be accelerated compared to the other two concepts. The SPEECH-concept on the other hand lead to slower reactions compared to the other concepts. Furthermore, it can be stated that participants from the *LED-concept* group reacted the most adapted to the situations. This was especially the case, when a braking maneuver was required to solve the situation.

However, further research is needed to gain more meaningfulness within the framework of HMI concepts, which are intended to support the driver in such short-term take-over situations. In addition to adaptations in the HMI concept, driver assistance systems are conceivable that support the human driver in such take-over situations.

In order to come to a final conclusion, the main findings are briefly summarized.

In CDA, NDRT-engagement can affect the fatigue state of the drivers. In this context, it can be stated that monotonous monitoring tasks can lead to fatigue within only 20 minutes. On the other hand, activities or tasks the human driver can freely choose have the potential, to keep the fatigue level of the human driver on a rather low level.

For the measurement or the detection of fatigue in CDA, the measurement of eye-lid closure over time (PERCLOS) seemed to be superior compared to subjective fatigue assessment. However, there is still research to be done on how to measure PERCLOS in a series production vehicle. The method that was used in the experiments presented in this thesis was too complex and not adapted to a series production vehicle.

Results also have shown, that fatigue has to be detected in CDA. Especially after prolonged automated driving, the participants had problems to take over vehicle control. Six participants were unable to control the vehicle within the available time budget and crashed with other road users when they were requested to intervene after 50 min of automated driving. This was especially the case, when the participants had to engage in the monotonous *Pqpd-task* that was used to induce *passive task-related fatigue*.

With the help of a *Wizard-of-Oz* approach, the transferability of the results regarding the drivers' state found in the simulator could be confirmed for on-road real driving environment. The course of drivers' fatigue was similar in both environments. Thus, it has to be expected, that also in on-road environment, the take-over performance of the drivers can be impaired due to emerging fatigue.

Furthermore it could be shown that different HMI concepts, which should warn the driver, have a lot of potential to support the driver. With the help of a led light signal, for example, a faster reaction of the drivers could be caused.

All in all, it has to be stated, that although the first CDA vehicles are supposed to come to the public market within the next few years, still a lot of research hast to be conducted.

Until now, experiments on the take-over behavior in short-term take-over situation have almost without exception only be conducted in driving simulators. Therefore, the Wizard-of-Oz approach that was used in Experiment 3 can be recommended for future research on this topic. Of course, testing of such *short-term* take-over situations in on-road environment still seems to be hazardous. A combination of the Wizard-of-Oz approach with test areas and other simulated road users can be beneficial here.

In any case, methods must be developed to be able to investigate the behavior of human drivers in *short-term* takeover situations in real traffic environments. The transferability of the results of the tests from the driving simulator seems to be too uncertain, especially regarding the reactions of the drivers.

References

Åkerstedt, T., Anund, A., Axelsson, J., & Kecklund, G. (2014). Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *Journal of sleep research*, *23*(3), 242–254.

Åkerstedt, T., & Gillberg, M. (1990). Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*, *52*(1-2), 29–37.

Åkerstedt, T., Peters, B., Anund, A., & Kecklund, G. (2005). Impaired alertness and performance driving home from the night shift: A driving simulator study. *Journal of Sleep Research*, *14*(1), 17–20.

Akin, M., Kurt, M. B., Sezgin, N., & Bayram, M. (2008). Estimating vigilance level by using EEG and EMG signals. *Neural Computing and Applications*, *17*(3), 227–236.

Alpern, M., & Minardo, K. (2003). Developing a car gesture interface for use as a secondary task. In G. Cockton & P. Korhonen (Eds.), *Chi 2003* (p. 932). New York: ACM Press.

Anund, A., Kecklund, G., Peters, B., Forsman, Å., Lowden, A., & Åkerstedt, T. (2008). Driver impairment at night and its relation to physiological sleepiness. *Scandinavian journal of work, environment & health*, 142–150.

Automotive News Europe. (2018). Mercedes plans advanced self-driving tech for S class. Retrieved from https://europe.autonews.com/article/20181011/ANE/181009659/mercedes-plans-advanced-self-driving-tech-for-s-class

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779.

Barón, A., & Green, P. (2006). *Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review.* (No. UMTRI 2006-5) University of Michigan.

Baum, L. F. (1900). The Wonderful Wizard of Oz, George M. *Hill Company, Chicago*.

Bengler, K., Winner, H., & Wachenfeld, W. (2017). No Human–No Cry?. *at-Automatisierungstechnik*, *65*(7), 471-476.

Bibby, K. S., Margulies, F., Rijnsdorp, J. E., Withers, R. M.J., & Makarov I. M. (1975). Man's role in control systems. *IFAC Proceedings Volumes*, *8*(1), 664–683.

Bishop, H., Madnick, B., Walter, R., & Sussman, E. D. (1985). *Potential driver attention monitoring system development.* (No. DOT-HS-806-744). United States. National Highway Traffic Safety Administration.

Bishop, R. (2005). Intelligent vehicle R&D: A review and contrast of programs world-wide and emerging trends. *Annales Des Télécommunications*, *60*(3), 228–263.

BMW AG. (2016). BMW Group, Intel and Mobileye Team Up to Bring Fully Autonomous Driving to Streets by 2021. Retrieved from https://www.press.bmw-group.com/global/article/detail/T0261586EN/bmw-group-intel-and-mobileye-team-up-to-bring-fully-autonomous-driving-to-streets-by-2021?language=en

Brickenkamp, R. (1962). Test d2: Aufmerksamkeits-Belastungs-Test.

Buld, S., Tietze, H., & Krüger, H. P. (2005). Auswirkungen von Teilautomation auf das Fahren. *Fahrerassistenzsysteme mit maschineller Wahrnehmung* (pp. 161-187). Springer, Berlin, Heidelberg.

Cohen, J. (1988). Statistical power analysis for the social sciences.

Čolić, A., Marques, O., & Furht, B. (2014). *Driver drowsiness detection: Systems and solutions*. *Springer briefs in computer science*. Cham: Springer.

Damböck, D. (2013). *Automationseffekte im Fahrzeug – von der Reaktion zur Übernahme* (Doctoral dissertation, Technische Universität München).

Damböck, D., Kienle, Martin, Bengler, Klaus, & Bubb, H. (2011). The H-Metaphor as an Example for Cooperative Vehicle Driving. International Conference on *Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments* (p. 376–385). Springer, Berlin, Heidelberg.

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

Deutscher Bundestag. (2017). Beschlussempfehlung und Bericht des Ausschusses für Verkehr und digitale Infrastruktur (15.Ausschuss). *Drucksache 18/11776*. Retrieved from http://dip21.bundestag.de/dip21/btd/18/117/1811776.pdf

DIN Deutsche Institut für Normung e.V. (2015). DIN EN ISO 9000:2015 *Qualitätsmanagementsysteme - Grundlagen und Begriffe* (ISO 9000:2015); Deutsche und Englische Fassung (EN ISO 9000:2015)

DIN Deutsches Institut für Normung e.V. (1998). DIN V 19233:1998-07 (Vornorm) – *Leittechnik - Automatisierung mit Prozessrechensystemen - Begriffe* (DIN V 19233:1998-07)

Dingus, T. A., Hardee, H., & Wierwille, W. (1987). Development of models for on-board detection of driver impairment. *Accident Analysis & Prevention*, *19*(4), 271–283.

Dingus, T. A., Jahns, S. K., Horowitz, A. D., & Knipling, R. (1998). Human factors design issues for crash avoidance systems. *Human factors in intelligent transportation systems*, *3*.

Elisabeth Schmidt. (2019). Effects of Thermal Stimulation During Passive Driver Fatigue (Doctoral Dissertation, Technische Universität Chemnitz).

Ergoneers GmbH. (2014). D-Lab 3 Manual EN. Retrieved from http://www.ergoneers.com/faq/index.php?action=artikel&cat=16&id=189&artlang=en

Ergoneers GmbH. (2015). Dikablis Professional: Eye Tracking Glasses. Retrieved from https://www.ergoneers.com/wp-content/uploads/2015/12/Dikablis-Professional-Flyer.pdf

Eriksson, A., Petermeijer, S. M., Zimmermann, M., Winter, J. C. F. de, Bengler, K., & Stanton, N. A. (2019). Rolling Out the Red (and Green) Carpet: Supporting Driver Decision Making in Automation-to-Manual Transitions. *IEEE Transactions on Human-Machine Systems*, *49*(1), 20–31.

Erwin, C. W. (1976). Studies of drowsiness: final report. *Durham, North Carolina: The National Driving Center.*

Feldhütter, A., Gold, C., Schneider, S., & Bengler, K. (2017). How the Duration of Automated Driving Influences Take-Over Performance and Gaze Behavior. *Advances in ergonomic design of systems, products and processes: Proceedings of the annual meeting of GfA 2016* (Vol. 57, pp. 309–318). Berlin, Germany: Springer.

Feldhütter, A., Hecht, T., & Bengler, K. (2018). Fahrerspezifische Aspekte beim hochautomatisierten Fahren: Schlussbericht zum Forschungsprojekt FE 82.0628/2015.

Feng, R., Zhang, G., & Cheng, B. (2009). An on-board system for detecting driver drowsiness based on multi-sensor data fusion using Dempster-Shafer theory. In *International Conference on Networking, Sensing and Control 2009* (pp. 897–902). Piscataway, NJ: IEEE.

Flemisch, F., Nashashibi, F., Rauch, N., Schieben, A., Glaser, S., & et al. (2010). Towards Highly Automated Driving: Intermediate report on the HAVEit-Joint System. *3rd European Road Transport Research Arena, TRA 2010.*, Brussels, Belgium.

Forster, Y., Naujoks, F., Neukum, A., & Huestegge, L. (2017). Driver compliance to take-over requests with different auditory outputs in conditional automation. *Accident; Analysis and Prevention*, *109*, 18–28.

G.tec Medical Engineering GmbH. (2018). g.Sensors and Utilities – Instructions for Use. Retrieved from http://www.gtec.at/Products/Electrodes-and-Sensors/g.Sensors-Specs-Features

Gail, J., Lorig, M., Gelau, C., Heuzeroth, D., & Sievert, W. (2001). *Optimization of rear signal pattern for reduction of rear-end accidents during emergency braking maneuvers.* Tech. Rep., Federal Highway Research Institute, Bergisch-Gladbach, Germany.

Gasser, T., Arzt, C., Ayoubi, M., Bartels, A., Bürkle, L., Eier, J., Flemisch, F., Häcker, D., Hesse, T., Huber, W., Lotz, C., Maurer, M., Ruth-Schumacher, S., Schwarz, J., Vogt, W. (2012). Rechtsfolgen zunehmender Fahrzeugautomatisierung - Gemeinsamer Schlussbericht der Projektgruppe. *Berichte der Bundesanstalt für Straßenwesen. Unterreihe Fahrzeugtechnik*. (83). Bergisch-Gladbach.

Gimeno, P. T., Cerezuela, G. P., & Montanes, M. C. (2006). On the concept and measurement of driver drowsiness, fatigue and inattention: implications for countermeasures. *International journal of vehicle design*, *42*(1-2), 67–86.

Gold, C. (2016). *Modeling of Take-Over Performance in Highly Automated Vehicle Guidance.* (Doctoral dissertation, Technische Universität München).

Gold, C., & Bengler, Klaus. (2014). Taking over control from highly automated vehicles. Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics, AHFE 2014, AHFE Conference Krakow, Poland, 2014.

Gold, C., Berisha, I., & Bengler, Klaus. (2015). Utilization of Drivetime – Performing Non-Driving Related Tasks While Driving Highly Automated. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1666–1670.

Gold, C., Körber, M., Lechner, D., & Bengler, Klaus. (2016). Taking Over Control From Highly Automated Vehicles in Complex Traffic Situations: The Role of Traffic Density. *Human Factors*, *58*(4), 642–652.

Gold, C., Naujoks, F., Radlmayr, J., Bellem, H., & Jarosch, O. (2017). Testing Scenarios for Human Factors Research in Level 3 Automated Vehicles. *Advances in Human Aspects of Transportation: Proceedings of the AHFE 2017 International Conference on Human Factors in Transportation*, Springer, p. 551 - 559.

Graw, P., Kräuchi, K., Knoblauch, V., Wirz-Justice, A., & Cajochen, C. (2004). Circadian and wake-dependent modulation of fastest and slowest reaction times during the psychomotor vigilance task. *Physiology & Behavior*, *80*(5), 695–701.

Hahn, S. (1993). Switching Between Autonomous And Conventional Car Driving A Simulator Study. In *Proceedings of the Intelligent Vehicles '93 Symposium* (pp. 25–30). IEEE.

Hardee, H., Dingus, T. A., & Wierwille, W. (1986). Driver Drowsiness Detection Using Subsidiary Task and Driving Performance Measures. *Proceedings of the Human Factors Society Annual Meeting*, *30*(4), 398–402.

Helton, W. S., & Warm, J. S. (2008). Signal salience and the mindlessness theory of vigilance. *Acta psychologica*, *129*(1), 18–25.

Hergeth, S., Lorenz, L., & Krems, J. F. (2017). Prior familiarization with takeover requests affects drivers’ takeover performance and automation trust. *Human factors*, *59*(3), 457–470.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of Sleepiness: A New Approach. *Psychophysiology*, *10*(4), 431–436.

Holloway, E. R. (1966). An automatic speed controller for automobiles. (Doctoral dissertation, Georgia Institute of Technology).

Hummel, M., Kühn, M., Bende, J., & Lang, A. (2011). Fahrerassistenzsysteme. Ermittlung des Sicherheitspotenzials auf Basis des Schadengeschehens der Deutschen Versicherer. *Forschungsbericht/Unfallforschung der Versicherer (UDV), Reihe Fahrzeugsicherheit*. (03).

Ingre, M., Åkerstedt, T., Peters, B., Anund, A., & Kecklund, G. (2006). Subjective sleepiness, simulated driving performance and blink duration: examining individual differences. *Journal of sleep research*, *15*(1), 47–53.

Ingre, M., Åkerstedt, T., Peters, B., Anund, A., Kecklund, G., & Pickles, A. (2006). Subjective sleepiness and accident risk avoiding the ecological fallacy. *Journal of sleep research*, *15*(2), 142–148.

ISO International Organization for Standardization (2013). ISO 15623:2013 *Intelligent transport systems – forward vehicle collision warning systems – performance requirements and test procedures* (Vol. 03.220.01; 35.240.60; 43.040.99) (No. 15623:2013). ISO International Organization for Standardization.

ISO International Organization for Standardization (2018). ISO/TR 21959-1:2018 *Road vehicles — Human performance and state in the context of automated driving — Part 1: Common underlying concepts* (Vol. 43.020) (No. ISO/TR 21959-1:2018). ISO International Organization for Standardization.

Jamson, A., Merat, N., Carsten, O. M. J., & Lai, F. C. H. (2013). Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation research part C: emerging technologies*, *30*, 116–125.

Jamson, H., & Smith, P. (2003). Drivers' perception of and response to brake failure. *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Park City, UT.*

Jarosch, O., Bellem, H., & Bengler, Klaus. (2019). Effects of Task-Induced Fatigue in Prolonged Conditional Automated Driving. *Human factors*, *61*(7), 1186-1199.

Jarosch, O., Kuhnt, M., Paradies, S., & Bengler, Klaus. (2017). It's Out of Our Hands Now! Effects of Non-Driving Related Tasks During Highly Automated Driving on Drivers' Fatigue. *Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design,* 319–325.

Jarosch, O., Paradies, S., Feiner, D., & Bengler, K. (2019). Effects of non-driving related tasks in prolonged conditional automated driving–A Wizard of Oz on-road approach in real traffic environment. *Transportation research part F: traffic psychology and behaviour*, *65*, 292-305.

Jewett, M. E., Dijk, D.-J., Kronauer, R. E., & Dinges, D. F. (1999). Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. *Sleep*, *22*(2), 171–179.

Johns, M. (2007). Drowsy Driving and the Law: A Submission to the Tasmania Law Reform Institute, vol. *Issues Paper*. (12).

Johns, M. (1998). Rethinking the assessment of sleepiness. *Sleep Medicine Reviews*, *2*(1), 3–15.

Johns, M. (1991). A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*, *14*(6), 540–545.

Johns, M., Chapman, R., Crowley, K., & Tucker, A. (2008). A new method for assessing the risks of drowsiness while driving. *Somnologie-Schlafforschung und Schlafmedizin*, *12*(1), 66–74.

Käppler, W.D. (1993). Views on the role of simulation in driver training. *Proceedings of the 12th European Annual Conference on Human Decision Making and Manual Control*, Kassel, Germany, 5.12-5.17.

Kerschbaum, P., Lorenz, L., & Bengler, K. (2014). Highly automated driving with a decoupled steering-wheel. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1686–1690.

Kerschbaum, P., Lorenz, L., & Bengler, K. (2015). A transforming steering-wheel for highly automated cars. In *2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, Korea* (pp. 1287–1292). Piscataway, NJ: IEEE.

Kienle, M. (2015). Kooperative Interaktionskonzepte zum Fahren eines teilautomatisierten Fahrzeugs. (Doctoral dissertation, Technische Universität München).

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching - A review. *Psychological bulletin*, *136*(5), 849.

Klemmer, S. R., Sinha, A. K., Chen, J., Landay, J. A., Aboobaker, N., & Wang, A. (2000). Suede. *Proceedings of the 13th annual ACM Symposium on User Interface Software and Technology, San Diego, California* (pp. 1–10). New York: ACM Press.

Körber, M., Cingel, A., Zimmermann, M., & Bengler, K. (2015). Vigilance decrement and passive fatigue caused by monotony in automated driving. *Procedia Manufacturing*, *3*, 2403–2409.

Lee, E. (1745). Self-regulating Wind Machine. *British Patent (old series)*, *615*, 1745.

Lenné, M. G., Triggs, T. J., & Redman, J. R. (1997). Time of day variations in driving performance. *Accident Analysis & Prevention*, *29*(4), 431–437.

Lin, C.-T., Chang, C.-J., Lin, B.-S., Hung, S.-H., Chao, C.-F., & Wang, I.-J. (2010). A real-time wireless brain–computer interface system for drowsiness detection. *IEEE transactions on biomedical circuits and systems*, *4*(4), 214–222.

Lorenz, L., Hergeth, S., Kerschbaum, P., Gold, C., & Radlmayr, J. (2015). Der Fahrer im hochautomatisierten fahrzeug. vom dual-task zum sequential-task paradigma: Ein Rückblick über Fahrsimulatorstudien. *7.Tagung Fahrerassistent.* München.

Mackworth, N. H. (1950). Researches on the measurement of human performance. (Med. Res. Council, Special Rep. Ser. No. 268.).

Manstetten, D., Krautter, W., Grothkopp, B., Steffens, F., & Geutner, P. (2001). Using a Driving Simulator to Perform a Wizard-of-Oz Experiment on Speech-Controlled Driver Information Systems. *Proceedings of the 1st Human-Centered Transportation*.

Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung. *Ergebnisse und Anwendungen der Psychophysiologie. Enzyklopädie der Psychologie Bd.* (7th ed., pp. 799–864). Göttingen: Hogrefe.

Marberger, C., Mielenz, H., Naujoks, F., Radlmayr, J., Bengler, K., & Wandtner, B. (2018). Understanding and Applying the Concept of “Driver Availability” in Automated Driving. *International Conference on Human Factors in Transportation, Los Angeles, California, USA* (Vol. 597, pp. 595–605). Cham: Springer.

Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology Section A*, *55*(2), 659–686.

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.

Merat, N., & Jamson, A. (2009). How Do Drivers Behave in a Highly Automated Car? *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design : Driving Assessment 2009* (pp. 514–521). Iowa City, Iowa: University of Iowa.

Merat, N., Jamson, A., Lai, F. C. H., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human factors*, *54*(5), 762–771.

Meyer, G., & Deix, S. (2014). Research and innovation for automated driving in Germany and Europe. In *Road vehicle automation* (pp. 71–81). Springer.

Miller, D., Sun, A., Johns, Mishel, Ive, H., Sirkin, D., Aich, S., & Ju, W. (2015). Distraction Becomes Engagement in Automated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1676–1680.

Mitler, M. M., Gujavarty, K. S., & Browman, C. P. (1982). Maintenance of wakefulness test: a polysomnographic technique for evaluating treatment efficacy in patients with excessive somnolence. *Electroencephalography and clinical neurophysiology*, *53*(6), 658–661.

Muir, H. C., Bottomley, D. M., & Marrison, C. (1996). Effects of motivation and cabin configuration on emergency aircraft evacuation behavior and rates of egress. *The International Journal of Aviation Psychology*, *6*(1), 57–77.

Murphy-Chutorian, E., & Trivedi, M. M. (2010). Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness. *IEEE Transactions on intelligent transportation systems*, *11*(2), 300–311.

National Center for Statistics and Analysis. (2017). Drowsy Driving 2015 (Crash Stats: Drowsy Driving 2015 (Crash Stats Brief Statistical Summary. Report No. DOT HS 812 446). Washington, DC: National Highway Traffic Safety Administration.

National Highway Safety Administration. (2017). ODI Resume: PE 16-007. Retrieved from https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.PDF

National Transportation Safety Board. (2018a). Preliminary Report: HWY18FH011. Retrieved from https://ntsb.gov/investigations/AccidentReports/Reports/HWY18FH011-preliminary.pdf

National Transportation Safety Board. (2018b). Preliminary Report: HWY18MH010. Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf

Naujoks, F., Befelein, D., Wiedemann, K., & Neukum, A. (2018). A Review of Non-driving-related Tasks Used in Studies on Automated Driving. *Proceedings of the AHFE 2017 International Conference on Human Factors in Transportation, Los Angeles, California, USA* (Vol. 597, pp. 525–537). Cham: Springer.

Naujoks, F., Forster, Y., Wiedemann, K., & Neukum, A. (2016). Speech improves human-automation cooperation in automated driving. *Mensch und Computer 2016 – Workshopband*.

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

Naujoks, F., Wiedemann, K., Schömig, N., Jarosch, O., & Gold, C. (2018). Expert-based controllability assessment of control transitions from automated to manual driving. *MethodsX*, *5*, 579–592.

Netter, F. (2017). Künstliche Intelligenz im Auto — Applikationen, Technologien und Herausforderungen. *ATZelektronik*, *12*(S1), 20–25.

Neubauer, C., Matthews, G., Langheim, L., & Saxby, D. (2012). Fatigue and voluntary utilization of automation in simulated driving. *Human Factors*, *54*(5), 734–746.

Neubauer, C., Matthews, G., & Saxby, D. (2012). The Effects of Cell Phone Use and Automation on Driver Performance and Subjective State in Simulated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1987–1991.

Neubauer, C., Matthews, G., & Saxby, D. (2014). Fatigue in the Automated Vehicle: Do Games and Conversation Distract or Energize the Driver? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 2053–2057.

Omae, M., Fujioka, T., Hashimoto, N., & Shimizu, H. (2006). The application of rtk-gps and steer-by-wire technology to the automatic driving of vehicles and an evaluation of driver behavior. *IATSS research*, *30*(2), 29–38.

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

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.

Petermann, I., & Schlag, B. (2010). Auswirkungen der Synthese von Assistenz und Automation auf das Fahrer-Fahrzeug System. *Proceedings of the AAET*, 257–266.

Petermann, I., & Kiss, M. (2010). Die Rolle des Fahrers im Spektrum von Automation und Transition: Nutzerzentrierte Gestaltung von Übergabe- und Übernahmeprozessen zwischen Fahrer und drei aufeinander aufbauenden Automationsstufen. *ATZ extra*, *15*(2), 90–95.

Petermann-Stock, I., Hackenberg, L., Muhr, T., & Mergl, C. (2013). Wie lange braucht der Fahrer - Eine Analyse zu Übernahmezeiten aus verschiedenen Nebentätigkeiten während einer hochautomatisierten Staufahrt. 6. *Tagung Fahrerassistenzsysteme. Der Weg zum automatischen Fahren*.

Philip, P., Sagaspe, P., Moore, N., Taillard, J., Charles, A., Guilleminault, C., & Bioulac, B. (2005). Fatigue, sleep restriction and driving performance. *Accident Analysis & Prevention*, *37*(3), 473–478.

Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Åkerstedt, T., Charles, A., Bioulac, B. (2005). Fatigue, sleepiness, and performance in simulated versus real driving conditions. *Sleep*, *28*(12), 1511–1516.

Radcliffe, W. (1828). *Origin of the new system of manufacture: commonly called" power-loom weaving," and the purposes for which this system was invented and brought into use, fully explained in a narrative, containing William Radcliffe's struggles through life to remove the cause which has brought this country to its present crisis*: J. Lomax.

Riley, V. (1989). A General Model of Mixed-Initiative Human-Machine Systems. *Proceedings of the Human Factors Society Annual Meeting*, *33*(2), 124–128.

Rossi, R., Gastaldi, M., & Gecchele, G. (2011). Analysis of driver task-related fatigue using driving simulator experiments. *Procedia - Social and Behavioral Sciences*, *20*, 666–675.

SAE, J. (2018). 3016: Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. *Society of Automotive Engineers*.

Salber, D., & Coutaz, J. (1993). Applying the Wizard of Oz technique to the study of multimodal systems. *Lecture Notes in Computer Science: Vol. 753. Human-computer interaction: Third international conference,* (Vol. 753, pp. 219–230). Berlin, New York: Springer-Verlag.

Saradadevi, M., & Bajaj, P. (2008). Driver fatigue detection using mouth and yawning analysis. *International journal of Computer science and network security*, *8*(6), 183–188.

Saxby, D. J., Matthews, G., Hitchcock, E. M., & Warm, J. S. (2007). Development of Active and Passive Fatigue Manipulations Using a Driving Simulator. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *51*(18), 1237–1241.

Saxby, D. J., Matthews, G., Hitchcock, E. M., Warm, J. S., Funke, G. J., & Gantzer, T. (2008). Effect of Active and Passive Fatigue on Performance Using a Driving Simulator. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(21), 1751–1755.

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.

Schieben, A., Heesen, M., Schindler, J., Kelsch, J., & Flemisch, F. (2009). The theater-system technique. In A. Schmidt (Ed.), *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (p. 43). New York, NY: ACM.

Schmidt, E. A. (2010). Die objektive Erfassung von Müdigkeit während monotoner Tagfahrten und deren verbale Selbsteinschätzung durch den Fahrer (Doctoral dissertation, Universität Düsseldorf).

Schmidt, E. A., Schrauf, M., Simon, M., Buchner, A., & Kincses, W. E. (2011). The short-term effect of verbally assessing drivers' state on vigilance indices during monotonous daytime driving. Transportation research part F: traffic psychology and behaviour, 14(3), 251-260.

Schmidt, J. (2018). Detektion der Reaktionsbereitschaft beim hochautomatisierten Fahren. (Doctoral dissertation, Technische Universität Berlin).

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*, 6652–6659.

Skipper, J. H., Wierwille, W., & Hardee, L. (1984). *An investigation of low level stimulus-induced measures of driver drowsiness.: IEOR Department Report 8402.*: Vehicle Simulation Laboratory, Human Factors Group. Virginia Polytechnic Institute and State University. Blacksburg. VA.

Stanton, N. A., Dunoyer, A., & Leatherland, A. (2011). Detection of new in-path targets by drivers using Stop & Go Adaptive Cruise Control. *Applied Ergonomics*, *42*(4), 592–601. https://doi.org/10.1016/j.apergo.2010.08.016

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

Statistisches Bundesamt. (2018). Verkehrsunfälle - *Zeitreihen 2017*.

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

Tesla Motors. (2016). All Tesla Cars Being Produced Now Have Full Self-Driving Hardware. Retrieved from https://www.tesla.com/blog/all-tesla-cars-being-produced-now-have-full-self-driving-hardware

Thoma, S. (2010). Mensch-Maschine-Interaktionskonzepte für Fahrerassistenzsysteme im Kreuzungsbereich. (Doctoral dissertation, Technische Universität München).

Trimble, T. E., Bishop, R., Morgan, J. F., & Blanco, M. (2014). *Human factors evaluation of level 2 and level 3 automated driving concepts: Past research, state of automation technology, and emerging system concepts.* (Report No. DOT HS 812 043). Washington, DC: National Highway Traffic Safety Administration.

Van den Beukel, A. P., & van der Voort, M. C. (2013). The influence of time-criticality on Situation Awareness when retrieving human control after automated driving. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC) (*pp. 2000–2005). Piscataway, NJ: IEEE.

Van den Beukel, A. P., & van der Voort, M. C. (2011). Human-centered challenges and contribution for the implementation of automated driving. In *Advanced Microsystems for Automotive Applications 2011* (pp. 225–235). Springer.

Vogelpohl, T., Vollrath, M., Kühn, M., Hummel, T., & Gehlert, T. (2016). *Übergabe von hochautomatisiertem Fahren zu manueller Steuerung: Teil 1. Review der Literatur und Studie zu Übernahmezeiten*. *Gesamtverband der Deutschen Versicherungswirtschaft [Tb1]: Forschungsbericht: Nr. 39*. Berlin: Unfallforschung der Versicherer, GDV.

Walch, M., Lange, K., Baumann, M., & Weber, M. (2015). Autonomous driving: Investigating the Feasibility of Car-Driver Handover Assistance. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 11–18). New York, New York: Association for Computing Machinery.

Wallace, R., & Silberg, G. (2012). Self-driving cars: the next revolution. *KPMG and CAR (Center for Automotive research), source: http://www. kpmg. com/Ca/en/IssuesAndInsights/ArticlesPublications/Documents/selfdrivingcars-next-revolution. pdf.*

Wandtner, B., Schömig, N., & Schmidt, G. (2018). Secondary task engagement and disengagement in the context of highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 253–263.

Warm, J. S., Parasuraman, Raja, & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human factors*, *50*(3), 433–441.

Weinbeer, V., Bill, J.-S., Baur, C., & Bengler, Klaus. (2018). Automated driving: subjective assessment of different strategies to manage drowsiness. *Proceedings of the human factors and ergonomics society europe*, 188.

Weinbeer, V., Muhr, T., & Bengler, Klaus. (2019). Automated Driving: The Potential of Non-driving-Related Tasks to Manage Driver Drowsiness. *Advances in Intelligent Systems and Computing: volume 823. Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)* (Vol. 823, pp. 179–188). Cham: Springer.

Weller, W. (2008). *Automatisierungstechnik im Überblick: [was ist, was kann Automatisierungstechnik?]* (1. Aufl.). *Maschinenbau : Wissen*. Berlin, Wien, Zürich: Beuth.

Wierwille, W., Wreggit, S. S., Kirn, C. L., Ellsworth, L. A., & Fairbanks, R. J. (1994). *Research on vehicle-based driver status/performance monitoring; development, validation, and refinement of algorithms for detection of driver drowsiness. Final report*. (Report Nr. DOT HS 808 247). Washington, DC: National Highway Traffic.

Zeeb, K., Härtel, M., Buchner, A., & Schrauf, M. (2017). Why is steering not the same as braking? The impact of non-driving related tasks on lateral and longitudinal driver interventions during conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *50*, 65–79.

Zhang, X., Zheng, N., Mu, F., & He, Y. (2009). Head pose estimation using isophote features for driver assistance systems. In *Ieee Intelligent Vehicles Symposium, 2009: Iv ; 3 - 5 June 2009, Xi'an, China* (pp. 568–572). Piscataway, NJ: IEEE.

Zimmermann, M., Bauer, S., Lutteken, N., Rothkirch, I. M., & Bengler, K. (2014). Acting together by mutual control: Evaluation of a multimodal interaction concept for cooperative driving. *Proceedings of the 2014 International Conference on Collaboration Technologies and Systems, Minneapolis, Minnesota* (pp. 227–235). Piscataway, NJ: IEEE.