Technische Universität München
Fakultät für Maschinenwesen
Lehrstuhl für Ergonomie

Automated Driving: Development of a Drowsiness Management Concept and Evaluation of Related Key Elements

Veronika Wehlack 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.-Ing. Mirko Hornung
Prüfer der Dissertation:	1. Prof. Dr. phil. Klaus Bengler
	2. Prof. Dr. Angelika Bullinger-Hoffmann

Die Dissertation wurde am 17.04.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Maschinenwesen am 06.11.2019 angenommen.

Danksagung

Diese Dissertation entstand während meiner Tätigkeit als Doktorandin der AUDI AG im Rahmen des Förderprojekts Ko-HAF (Kooperatives hochautomatisiertes Fahren).

Für die spannenden Diskussionen, die andauernde Unterstützung sowie für die lehrreiche und prägende Zeit möchte ich mich sehr herzlich bei Herrn Prof. Klaus Bengler bedanken. Für die Zweitbegutachtung danke ich Frau Prof. Bullinger-Hoffmann. Zudem möchte ich Herrn Prof. Mirko Hornung für die Übernahme des Prüfungsvorsitzes danken.

Für den fachlichen Austausch, die anregenden Diskussionen und das kritische Korrekturlesen der Arbeit, geht mein besonderer Dank an Matthias Walter, Jonas Radlmayr und Herrn Dr. Patrick Galaske.

Ganz besonders bedanken möchte ich mich bei Herrn Dr. Dirk Blaese, Herrn Tobias Muhr, Herrn Dr. Karl-Heinz Siedersberger und Herrn Dr. Martin Wimmer für die Möglichkeit, diese Arbeit verfassen zu können und für die Unterstützung während meiner Doktorandenzeit. Auch bei meiner Wegbegleiterin, Frau Dr. Katharina Reinmüller, möchte ich mich an dieser Stelle herzlich für den fachlichen Austausch bedanken. Für die Unterstützung im Rahmen der Studiendurchführung gilt mein Dank Herrn Christoph Baur, Herrn Julian Bill, Frau Christina Romanowski und Frau Dursune Gönültas.

Bedanken möchte ich mich bei meiner Familie und insbesondere bei meinem Ehemann für den kontinuierlichen Rückhalt. Durch eure Unterstützung war es mir möglich, die vorliegende Dissertation zu verfassen.

Zusammenfassung

In dieser Arbeit wird angenommen, dass der Fahrerzustand Schläfrigkeit eine andere Systemgrenze einer automatisierten Fahrfunktion (Level 3 gemäß SAE (2018)) darstellt. Um eine Beeinträchtigung der Sicherheit aufgrund des Fahrerzustands Schläfrigkeit zu vermeiden, wurde ein Schläfrigkeitsmanagementkonzept entwickelt. Dieses berücksichtigt sowohl die Nutzer- als auch die Herstellerperspektive.

Zur Beantwortung zentraler Fragestellungen des entwickelten Schläfrigkeitsmanagementkonzepts wurden drei Wizard-of-Oz-Studien mit insgesamt 126 Probanden durchgeführt. Die Ergebnisse weisen darauf hin, dass *(a)* der Einfluss von Schläfrigkeit auf die Übernahmeleistung zustandsabhängig untersucht werden sollte, *(b)* es mithilfe verschiedener Maßnahmen möglich ist, während der normalen Arbeitszeit Schläfrigkeit ohne Schlafdeprivation zu erzeugen, *(c)* Probanden in einem umgebauten Rechtslenker das Gefühl einer Level 3 Fahrt bekommen können, *(d)* Schläfrigkeit tendenziell eher zu Überreaktionen als zu verlangsamten Reaktionen führt, *(e)* bestimmte fahrfremde Tätigkeiten geeignet sind, die Schläfrigkeit zu reduzieren, *(f)* die Mehrheit der Nutzer über die Ursache einer Systemanpassung informiert werden möchte und dass *(g)* das Angebot einer Ersatzhandlung einen akzeptierten Ansatz darstellt, um mit Unsicherheiten eines Fahrerbeobachtungssystems umzugehen.

Zukünftige Forschung sollte eine Standardisierung von Wizard-of-Oz-Studien anstreben, um die Objektivität und Übertragbarkeit der Ergebnisse zu erhöhen. Weitere Forschung ist zudem erforderlich, um zu untersuchen, wie schläfrigkeitsbedingte Überreaktionen vermieden werden können.

Abstract

This thesis considers driver drowsiness as another system limit of a Level 3 (L3) Automated Driving System (ADS). To prevent potential safety impairments caused by driver drowsiness, a Drowsiness Management Concept (DMC) is developed that takes the perspectives of users and manufacturers into account.

By using a Wizard-of-Oz approach, this thesis investigates key elements of the DMC. Overall, three studies with a total of 126 participants were conducted in real traffic. The study results indicate that *(a)* a state-dependent approach is required to investigate the influence of drowsiness on take-over performance, *(b)* it is possible to induce driver drowsiness during regular working hours without sleep deprivation by taking several measures, *(c)* participants can get a feel for automated driving in the modified right-hand-drive vehicles, *(d)* drowsiness rather provokes overreactions than increases Take-over time (ToT) metrics, *(e)* specific non-driving-related tasks (NDRTs) are a useful measure to reduce drowsiness, *(f)* the majority of users wishes to be informed about the cause of a system adaption, and *(g)* offering a compensation task to bypass a system adaption represents an accepted and trusted approach to deal with uncertainty periods of a Driver Monitoring System (DMS).

Further research should focus on how Wizard-of-Oz studies can be standardized to increase the objectivity and transferability of the results and on how overreactions caused by driver drowsiness can be avoided.

Contents

List of Figures

List of Tables

Acronyms

ACC Adaptive Cruise Control

ADS Automated Driving System

BASt Bundesanstalt für Straßenwesen

DDT dynamic driving task

DL Drowsiness level

DMC Drowsiness Management Concept

DMS Driver Monitoring System

DSRS driver-state related strategy

ECG Electrocardiogram

EEG Electroencephalogram

GEQ Game Experience Questionnaire

HMI Human Machine Interface

KSS Karolinska-Sleepiness Scale

L3 Level 3

LKA Lane Keeping Assist

MRM Minimum Risk Maneuver

NDRT non-driving-related task

NHTSA National Highway Traffic Safety Administration

ODD operational design domain

OEM Original Equipment Manufacturers

SAE Society of Automotive Engineers

SBS system-based strategy

SSS Stanford-Sleepiness Scale

SuRT Surrogate Reference Task

ToT Take-over time

TTC Time To Collision

CHAPTER 1

Introduction

In 1983 Bainbridge presented *"Ironies of automation"*, emphasizing that if automation increases, the interaction between automation and the human operator can become critical to safety. This is because automated systems aim to enhance safety by replacing error-prone humans but still expect them to retake control safely in the event of system failure.

Automated Driving Systems (ADSs) aim to increase safety, user comfort, and traffic efficiency (Hoeger et al., 2011). Bengler, Winner, and Wachenfeld (2017) provided a critical view on the frequently used argument that automation will increase safety by taking error-prone humans out of the control loop. Accidents are usually rare and multi-causal events, technical systems are also not error-free, humans often act as a final safety component, and the announced automated functions rather focus on comfort in relatively safe traffic situations.

This indicates that it is challenging to enhance safety by introducing ADSs and that to achieve this goal, the interaction between such systems and users must be designed very carefully. Various researchers support this by pointing out that different human factors (e.g., trust) must be considered more seriously with an increase in automation (e.g., Parasuraman & Riley, 1997; Saffarian, Winter, & Happee, 2012). At a specific automation level, users no longer need to monitor the system. However, users must still be able to take over control safely, as they are expected to be *"fallback-ready"* (SAE, 2016). Different impact factors were already intensively investigated in this context. For example, researchers assessed the influence of traffic density (e.g., Radlmayr, Gold, Lorenz, Farid, & Bengler, 2014), take-over time (e.g., Gold, Damböck, Lorenz, & Bengler, 2013), design of the requests to intervene (e.g., Kerschbaum, Lorenz, Hergeth, & Bengler, 2015), non-driving-related tasks (NDRTs) (e.g., Radlmayr et al., 2014), age (e.g., Gold, Körber, Hohenberger, Lechner, & Bengler, 2015), and trust (e.g., Hergeth, Lorenz, Vilimek, & Krems, 2016) on take-over performance.

It is not yet known, however, whether and to what extent specific driver states could impair users' fallback readiness and thus decrease the take-over performance. Hence, this thesis first takes a closer look at requirements on a user's state concerning certain automation levels. For this, the following chapter discusses specific definitions of automation levels and legal aspects.

Three articles were published as a result of this work. These are listed below:

- *Automated driving: Subjective assessment of different strategies to manage driver drowsiness.* Human Factors and Ergonomics Society (2018) (oral presentation on September 28th, 2017 in Rome)
- *Highly automated driving: How to get the driver drowsy and how does drowsiness influence various take-over aspects?* 8.Tagung Fahrerassistenz (2017) (oral presentation on November 23th, 2017 in Munich)
- *Automated driving: The potential of non-driving-related tasks to manage driver drowsiness.* International Ergonomics Association (2019) (oral presentation on August 29th, 2018 in Florence)

CHAPTER 2

Theoretical background and state of the art

2.1 Automation levels and legal aspects

In recent years, various definitions of automation levels have been published in the context of automated driving. The most common and frequently used definitions are the definitions developed by the U.S. National Highway Traffic Safety Administration (NHTSA) (NHTSA, 2013), by the German Bundesanstalt für Straßenwesen (BASt) (Gasser et al., 2012), and by the Society of Automotive Engineers (SAE). This thesis uses the terminology provided by the SAE.

The SAE distinguishes six levels of automation (SAE, 2016). These levels range from *"No Driving Automation"* to *"Full Driving Automation"*. L3 *"Conditional Driving Automation"* is the first level in which Automated Driving Systems (ADSs) perform the entire dynamic driving task (DDT). Table 2.1 provides an overview of the definitions of the different automation levels and the DDT fallback.

> ADS: *"The hardware and software that are collectively capable of performing the entire DDT on a sustained basis, regardless of whether it is limited to a specific operational design domain (ODD); this term is used specifically to describe a level 3, 4, or 5 driving automation system."* (SAE, 2016)

L3 differs from L4 *"High Driving Automation"* in the type of DDT fallback.

In L3 users act as the DDT fallback, whereas in L4 a user may, but does not have to perform the DDT fallback and is not expected to do so. To represent a DDT fallback, users of an L3 ADS are expected to be receptive to a request to intervene and must respond appropriately to such a request.

> Request to intervene: *"Notification by an ADS to a fallback-ready user indicating that s/he should promptly perform the DDT fallback, which may entail resuming manual operation of the vehicle (i.e., becoming a driver again), or achieving a minimal risk condition if the vehicle is not drivable."* (SAE, 2018)

Further, they must be receptive in a way that they represent the DDT fallback when *"evident system failures"* occur, even without a request to intervene. Also, *"the user of an engaged level 3 ADS feature who is seated in the driver's seat of an equipped vehicle is the DDT fallback-ready user even if s/he is no longer receptive to a request to intervene because s/he has improperly fallen asleep"* (SAE, 2016). In addition, it is described that *"the driver state or condition of being receptive to alerts or other indicators of a DDT performance-relevant system failure, as assumed in level 3, is not a form of monitoring"* (SAE, 2016). In the revision of this technical report in 2018 the state of being receptive has been defined more precisely as *"an aspect of consciousness characterized by a person's ability to reliably and appropriately focus his/her attention in response to a stimulus"* (SAE, 2018).

Besides this technical report, in 2017 a traffic law (Achtes Gesetz zur Änderung des Straßenverkehrsgesetzes) was amended in Germany to provide a legal framework for the interaction between users and automated vehicles (Bundestag, 2017a). This change allows users to perform NDRTs (e.g., writing emails by using a vehicle's infotainment-system). However, also it points out that users must be receptive[1] in a way that they can understand a traffic situation and can take over control within a sufficient amount of time when requested or even without a request to intervene, for example, when the system conducts an emergency stop without any external circumstances (Bundestag, 2017b). The requirements described in this law overlap considerably with those provided by the SAE (2016, 2018).

However, this traffic law has been criticized for being imprecise, especially because of the requirement that users must still be receptive. According to Schirmer

[1] In the original law the German term *"wahrnehmungsbereit"* was used to describe the requirements which a user must fulfill when using an ADS. In this thesis, the German term *"wahrnehmungsbereit"* is considered as a synonym for the term *"receptive"*.

(2017), in a former not published draft[2], it was stated that a user who has fallen asleep could not be considered receptive. However, this specification is missing in the adopted law (Schirmer, 2017). Also, sleep was seen as a foreseeable misuse in a round-table report, which requires effective measures (Gasser et al., 2015, p. 10).

The discussion about the terms *"fallback-ready"* and *"receptive"* clearly shows that there is a challenge in defining a user state that should be ensured or avoided when an L3 ADS is used. Furthermore, it is not yet clear whether manufacturers and/or users are responsible for ensuring a driver state that allows users to safely take over control. So far it is clear that users of such an ADS must still be able to take over control in a safe and timely manner in the event of a request to intervene or when evident system failures occur, and, consequently, they are not allowed to sleep. However, it remains unclear whether and to what extent driver states before falling asleep can impair a driver's receptivity, and thus a safe transition.

[2] Entwurf eines Gesetzes zur Änderung des Straßenverkehrsgesetzes v. 27.6.2016 (Referentenentwurf; nicht veröffentlicht) as cited in Schirmer (2017).

Table 2.1: Automation levels taken from the SAE (2016)

Level	Name	Narrative definition	DDT fallback
0	No Driving Automation	The performance by the driver of the entire DDT, even when enhanced by active safety systems.	Driver
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
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 (object and event detection and response) subtask and supervises the driving automation system.	Driver
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 requests to intervene, as well as to DDT performance-relevant system failures in other vehicle systems, and will respond appropriately.	Fallback-ready user (becomes the driver during fallback)
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 request to intervene.	System
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 request to intervene.	System

2.2 Human, driver, and take-over performance

The following chapters take a closer look at factors that could reduce human, driver or take-over performance. Three main measurement techniques for the quality of human performance were identified by Wickens et al. (2013, p. 2). These are:

- *"Speed (faster is better),*
- *Accuracy (higher is better) and*
- *Attentional demand (less is generally better)."* (Wickens et al., 2013, p. 2)

A reduced attentional demand is important when other tasks need to be performed simultaneously (Wickens et al., 2013, p. 3). Besides this, Graf (1960) emphasized that human performance should always be assessed in consideration of the time needed to perform a specific task. Consequently, knowledge about the task to be investigated is required for an assessment of human performance.

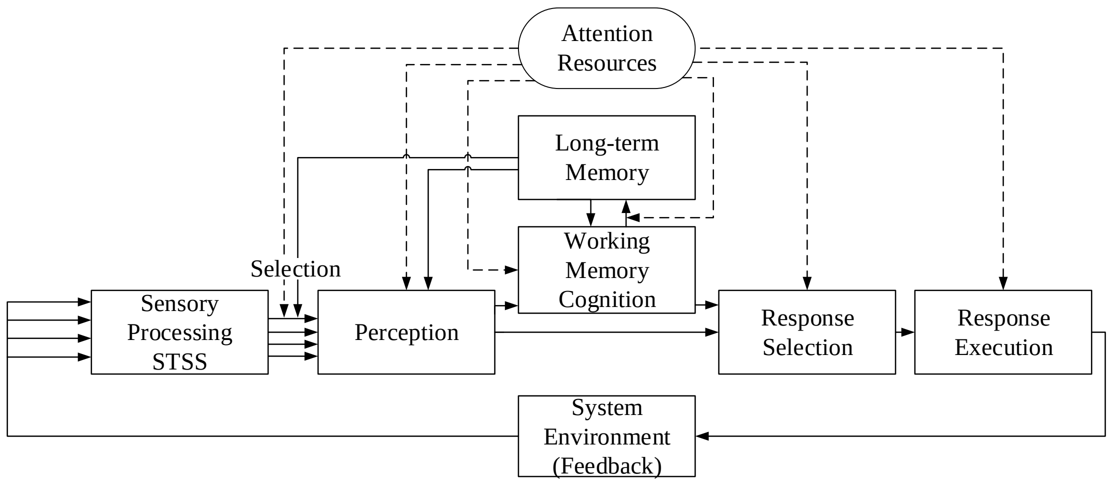


Figure 2.1: The model of human information processing taken from Wickens et al. (2013, p. 4)

The human information processing model provides a fundamental basis for a task analysis (see figure 2.1). The signal processing, the perception (determining the meaning of information), the response selection, the response execution, and feedback are relevant components of this model (Wickens et al., 2013, p. 4).

Figure 2.2 shows that also direct (e.g., lighting) and indirect (physiological arousal[3]) stressors can influence different stages of the human information processing and thus have the potential to impair human performance (Wickens et al., 2013, p. 360).

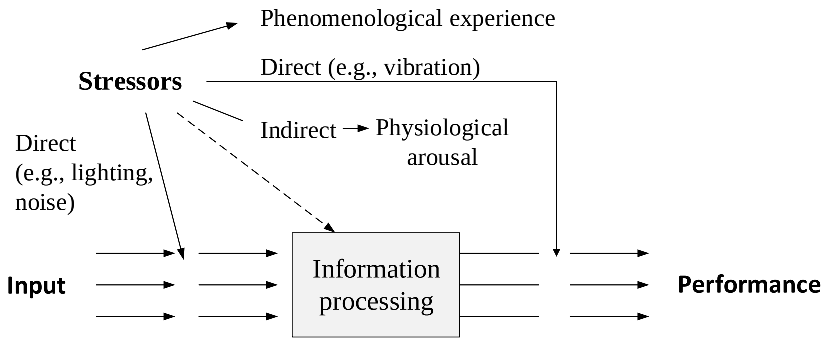


Figure 2.2: A presentation of stress effects taken from Wickens et al. (2013, p. 360)

This consideration is in line with the findings of Mullins, Cortina, Drake, and Dalal (2014), who concluded that sleepiness can reduce the processing of information. Also, a task can be completed faster and easier with an increase in experience. This relation was described by Rasmussen (1983).

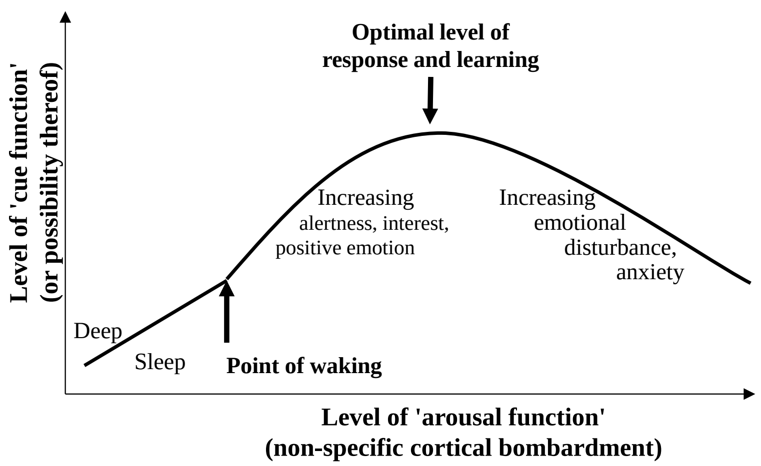


Figure 2.3: The relation between arousal and performance taken from Hebb (1955)

[3] Arousal was defined as *"an individual's level of activity, whether reflected in general behavioral states such as active wakefulness or sleep or in subjective experience, such as alertness or drowsiness"* (Wickens et al., 2013, p. 361).

Yerkes and Dodson (1908) found that there is a relation between the strength of stimuli and the speed of learning in mice experiments. The geometric orientation of the original U-curve observed by Yerkes and Dodson (1908) has been inverted by Hebb (1955) (see figure 2.3) and, according to Teigen (1994), has been applied to human performance later. According to this inverted U-curve, the arousal level influences human performance. For example, human performance is considered low when a person has fallen asleep.

In the driving context, it was emphasized that the relation between driving performance and risk is very complex and includes various impact factors, such as *"the current complexity/difficulty of the driving task"*,*"the driver's vehicle handling skills"* or *"the ability to make a correct situation assessment"* (Östlund et al., 2005, pp. 27-28).

Driver performance (in automation levels lower than L3) can be assessed in general by several metrics such as reactions times (Bengler, 2015) or other driving task-related parameters such as the minimum Time To Collision (TTC), the standard deviation of the steering wheel angle, lane exceedances, and many more (for an overview see Östlund et al. (2005, p. 32)). However, for L3 ADSs there is a paradigm shift regarding a user's task (Gold, 2016, p. 13). Consequently, also for an assessment of take-over performance, sound metrics are needed.

Gold (2016, p. 22) defined take-over performance as a *"combination of timing and quality aspects of driver's input within a take-over scenario."* As the focus of this definition is on the combination of time and quality aspects, and as both aspects can be considered extremely relevant for a safe take-over of vehicle control, this thesis follows the definition and related metrics proposed by Gold (2016, pp. 22-25). Figure 2.4 illustrates the relevant time aspects of a transition (Marberger et al., 2018). Several driving performance metrics can be used to assess take-over quality, for example, longitudinal and lateral acceleration, the time to collision, and whether or not a crash occurred (Gold, 2016, pp. 24-25). Wandtner, Schömig, and Schmidt (2018) found that users benefit from a NDRT lockout when it comes to a take-over situation. This indicates that the focus of take-over performance should be on take-over time and quality. It does not seem to make sense to focus on reduced attention during the transition phase so that users can do other tasks at the same time.

In the context of automated driving, arousal was identified as one relevant factor that could negatively influence a safe transition from automated to manual driving (Marberger et al., 2018). Notably, the fact that arousal (or drowsiness/sleepiness) can interfere with information processing (Mullins et al., 2014), which as a consequence might impair a user's receptivity, requires a closer look at this and related constructs.

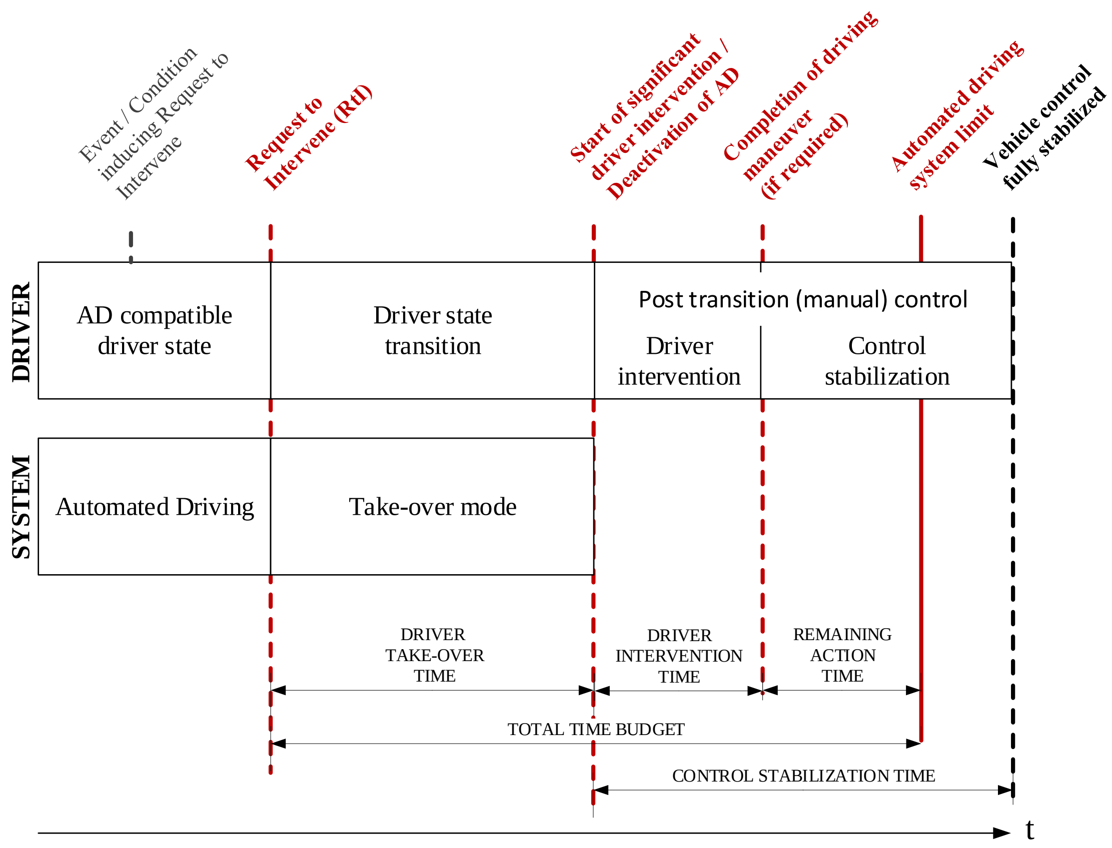


Figure 2.4: Transition model derived from Marberger et al. (2018)

2.3 Sleepiness, drowsiness, and fatigue

Researchers' opinions differ regarding the need to distinguish between constructs like sleepiness/drowsiness and fatigue. Some state that it is possible to distinguish between these constructs (e.g., Kircher, Uddman, & Sandin, 2002), whereas others use these terms interchangeably (e.g., Knipling & Wang, 1994). For an understanding of these terms, relevant definitions and models are presented and then discussed in the context of automated driving.

Sleepiness can be defined as *"a physiologic drive toward sleep"* and is often used synonymously to the term drowsiness (Ahmed & Thorpy, 2011).[4] Drowsiness or a drowsy state are defined as *"a transitional state between wakefulness and sleep"* (Johns, 1998).

Further, Knipling and Wierwille (1994) defined drowsiness in the driving context as follows:

> *"[...] drowsiness is used here to refer to the state of reduced alertness, usually accompanied by performance and psychophysiological changes, which may result in loss of alertness or being 'asleep at the wheel'."*

A fundamental model for understanding the development of sleep is the two-process model of sleep regulation provided by Borbély (1982). This model shows the relation between the time awake (sleep-dependent Process S) and the circadian rhythm (sleep-independent Process C). According to this model, the interaction between these two processes determines whether a person falls asleep. However, according to Johns (1998), the two-process model has been unable to explain why people can fall asleep within minutes just by laying down, regardless of the time of day or the time awake. Therefore, Johns (1998) evolved the model of Borbély (1982) into the *"four-process conceptual model of sleep and wakefulness"*. A wake drive and a sleep drive are the core elements of this concept. A secondary drive can influence each of these drives (see figure 2.5).

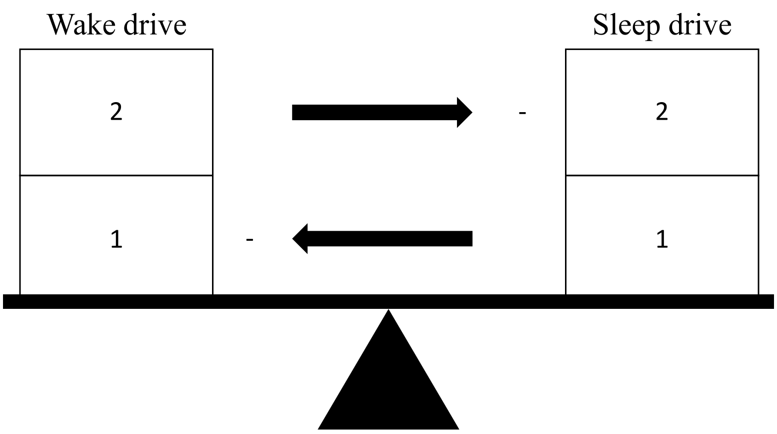


Figure 2.5: Flip-Flop: Sleep and wake drives with the primary and secondary components, according to Johns (1998)

The primary and secondary wake drives result in the total wake drive. The

[4] in the following the term drowsiness is used synonymously to sleepiness

primary wake drive represents the Process C in conformity with Borbély (1982). Several factors modulate the secondary wake drive (e.g., posture or visual stimuli (Johns, 2007)). The term *"somnificity"* joins the factors that can influence the secondary wake drive (Johns, 2010). Further, the degree of somnificity differs among different activities (e.g., lying down to rest in the afternoon reaches higher somnificity scores than sitting and talking to someone) (Johns, 2002).

Thus, the somatosensory input can influence the secondary wake drive as it affects the afferent nervous system. Based on this connection, Johns (2010) suggested naming this process *Process A*. As Saper, Cano, and Scammell (2005) found that also emotional and cognitive input have a major impact on this self-influenceable process, George (2018) added these impact factors to the *Process A* and therefore proposed to name this process the *"psycho-sensory wake drive"*.

Johns (1998) divided the total sleep drive into the primary sleep drive, which depends on the activation of the central nervous system (CNS), and into the secondary sleep drive, which represents the Process S in conformity with Borbély (1982). Whenever the total sleep drive exceeds the total wake drive a person would fall asleep and vice versa (see figure 2.5) (Johns, 1998).

In addition, Johns (2007) stated that distinguishing between drowsiness/ sleepiness and fatigue is of great importance and that this distinction is needed to decide how driver drowsiness should be managed. Johns (2007) characterizes drowsiness by short-term changes and by periods without awareness. This state is entered whenever a person falls asleep and might last longer if this state should be avoided, for example, while driving. In contrast to drowsiness, fatigue does not change rapidly and gets worse as the duration of a demanding task increases. A further difference between drowsiness and fatigue is that drowsiness can only be relieved by sleep, whereas fatigue can already be reduced by rest (Johns, 2007). A driving simulator study examining the influence of breaks during a prolonged manual drive on sleepiness (assessed by the Karolinska-Sleepiness Scale (KSS) and Electroencephalogram (EEG)) and fatigue (assessed by the POMS (Profile of Mood States)) supports this. Breaks helped minimize fatigue. However, such a positive effect was not observed for sleepiness (Phipps-Nelson, Redman, & Rajaratnam, 2011).

In the medical context, sleepiness can be distinguished from severe fatigue by looking at how patients react to a nap. Patients experiencing fatigue will, for example, still feel weak after a nap, whereas patients experiencing sleepiness will feel better (Rosenthal, Majeroni, Pretorius, & Malik, 2008). Additionally, Shen, Barbera, and Shapiro (2006) pointed out that sleepiness and fatigue are complex and interrelated but distinct constructs. In their article, they define fatigue as follows: *"is an overwhelming sense of tiredness, lack of energy and a feeling of exhaustion, associated with impaired physical and/or cognitive functioning."*

Also, other researchers mainly focused on task-related aspects. For example:

1. Fatigue was defined *"as a reduced inclination for activity, due to excessive extension in time or intensity of that activity"* by (Åkerstedt, 2011).

2. Physiological fatigue was defined *"as a subjectively experienced disinclination to continue performing the task at hand"* by (Brown, 1994).

These definitions can be considered in line with the fatigue definition provided by Johns (2007).

So far, it has not been difficult to distinguish between sleepiness/drowsiness on the one hand and fatigue on the other. However, following the fatigue definition, Brown (1994) assumed that *"the most frequent cause of general attentional impairment is the eye closure that accompanies sleepiness"*. Johns (2000) questioned Brown's assumption as it mixes the states fatigue and sleepiness. Another definition of fatigue was proposed by Grandjean (1979), who categorized fatigue into physical and mental fatigue. As physical fatigue is related to muscular fatigue, it is not considered relevant here. According to Grandjean (1979), *"mental fatigue [...] is a diffuse sensation which is accompanied by feelings of indolence and disinclination for any kind of activity"*. This definition also focuses on an activity that one does not like to continue due to suffering from fatigue. However, later in this article Grandjean (1979) described that there are different functional states (*"deep sleep; light sleep, drowsy; weary, hardly awake; relaxed, resting; fresh, alert; very alert, stimulated; in a state of alarm."*) and considers that *"mental fatigue is a functional state which grades in one direction into sleep, and in the opposite direction into a relaxed, restful condition"*. When taking the distinctions between drowsiness and fatigue into account, it becomes clear that the terms drowsiness/sleepiness and fatigue are mixed here. However, the presentation of the different functional states supports the idea that drowsiness is a state that is passed shortly before falling asleep. Further, fatigue is considered to be *"distressing"* when one is not allowed to relax (Grandjean, 1979). This is supported by the affect grid, which shows that fatigue and drowsiness are states of less activation (see figure 2.6). However, a feeling of being drowsy is in the quadrant of *"low-activation positive affect"*, whereas a feeling of fatigue is in the quadrant of *"low-activation negative affect"* (Warr, 2013).

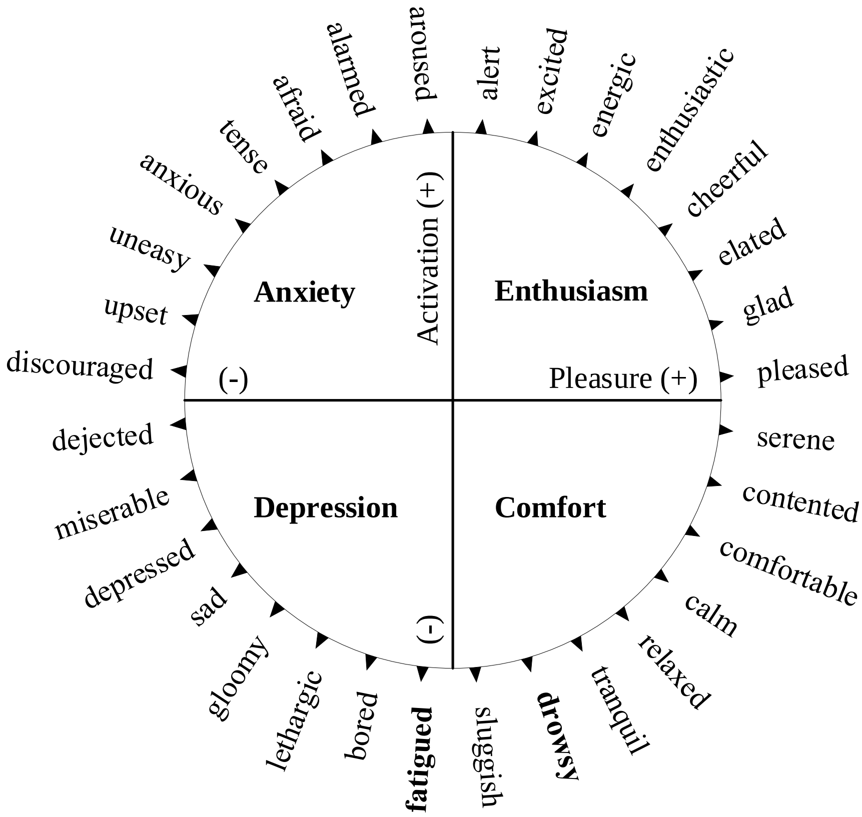


Figure 2.6: Assignment of drowsiness and fatigue in the affect circumplex taken from Warr (2013)

Taking all of this into account, this thesis concludes that drowsiness/sleepiness can be distinguished from fatigue. Table 2.2 summarizes the derived differences between these constructs. However, the question remains why these terms are frequently used synonymously. A possible explanation for this might be that these constructs are often experienced at the same time (Johns, 2007). Another explanation might lie in the fatigue model provided by May and Baldwin (2009). This model (see figure 2.7) may have contributed to the general confusion between the terms fatigue and sleepiness/drowsiness, especially in the context of automated

Table 2.2: The distinction between fatigue and sleepiness/drowsiness

Fatigue	**Sleepiness/drowsiness**
• increases with the duration of a demanding task (Johns, 2007; Åkerstedt, 2011) • does not change rapidly (Johns, 2007) • humans suffering fatigue experience a disinclination to perform the task at hand (Brown, 1994; Åkerstedt, 2011) • is a low-activation negative affect (Warr, 2013) • attention and vigilance problems are likely to occur (Brown, 1994) • can be relieved by rest (not necessarily by sleep) (Johns, 2007; Rosenthal et al., 2008)	• is a state of reduced alertness usually accompanied by performance and psychophysiology changes (Knipling & Wierwille, 1994) • is a transitional state between wakefulness and sleep (Johns, 1998; Grandjean, 1979); is a state shortly before sleep (Grandjean, 1979) • is a low-activation positive affect (Warr, 2013) • can result in a loss of alertness or being "asleep at the wheel" (Knipling & Wierwille, 1994), periods without awareness occur (Johns, 2007), • is characterized by short-term changes (Johns, 1998) • can be influenced by somatosensory (Johns, 1998) and by emotional and cognitive input (Saper et al., 2005) • can be relieved by sleep (Johns, 2007; Rosenthal et al., 2008)

driving. Many researchers are referring to this fatigue model, regardless of the automation level they are investigating.

In general, this model provides a good overview of various factors that can influence task-related fatigue, and thus can be helpful to identify countermeasures to avoid a decrease in driving performance due to attention or vigilance[5] problems. According to this model (see figure 2.7), active task-related-fatigue can occur due to a demanding task such as driving in adverse weather conditions, whereas passive

[5] Vigilance can be assessed by specific tasks *"in which attention is directed to one or more sources of information over long, unbroken, periods of time, for the purpose of detecting small changes in the information being presented. Such tasks are also known as 'monitoring' or 'watch-keeping' tasks"* (Davies & Parasuraman, 1982, p. 3).

task-related fatigue can occur, for example, in monotonous driving situations or when automated systems are used. The description of active and passive task-related fatigue can be considered in line with the definitions of fatigue mentioned above (see table 2.2).

It is, however, questionable whether this model remains valid also for L3 ADSs. May and Baldwin (2009) pointed out that

> *"Passive fatigue is produced when a driver is mainly monitoring the driving environment over an extended period of time when most or the entire actual driving task is automated."* (May & Baldwin, 2009)

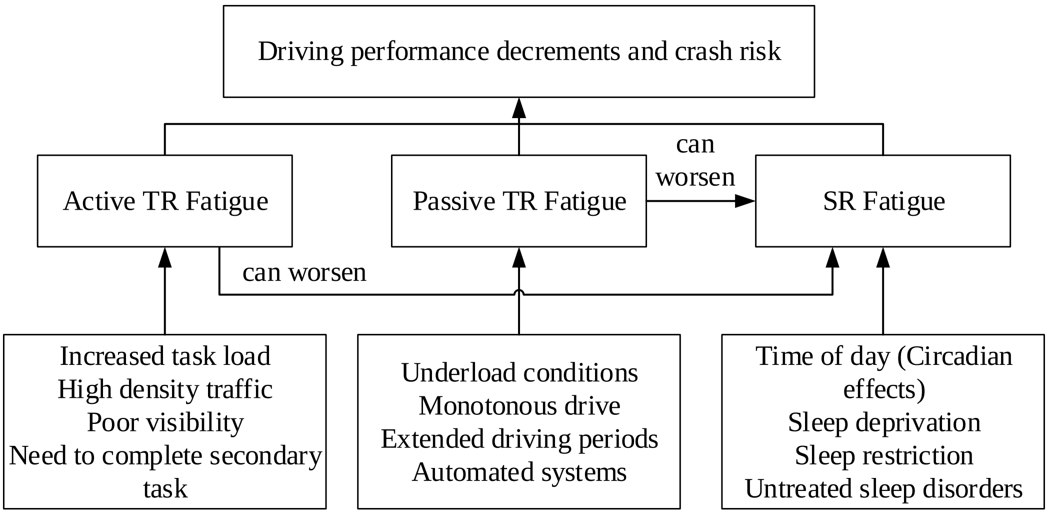


Figure 2.7: Model of fatigue taken from May and Baldwin (2009)

However, an L3 ADS does not require monitoring (SAE, 2016, 2018) (see section 2.1). Also, it is unlikely that the model provided by May and Baldwin (2009) has already considered the paradigm shift of a user's task for L3 ADS. In 2009, research focused mainly on partially automated systems and mostly started to focus on L3 ADSs in about 2013 (e.g., Gold et al., 2013). Therefore, automated systems in this model are likely to relate to L2 rather than to L3 systems.

Taking all of this together, this thesis assumes that passive task and active task-related fatigue lose relevance as long as an L3 ADS is used (unless it is of interest to decrease the probability of error in an NDRT).

Further, the fatigue model provided by May and Baldwin (2009) appears to present a model for automation levels below L3, where a user is responsible for the entire DDT and must constantly monitor a system and the driving environment. In contrast, it appears that driver sleepiness/drowsiness might even increase in relevance when L3 ADSs are used. Various studies have shown that users want to look out of the window, relax, and even sleep while using an ADS (e.g., Pfleging, Rang, & Broy, 2016; Yang, Klinker, & Bengler, 2019). However, relaxing has high somnificity scores and is therefore likely to increase drowsiness (Johns, 2002). The wish to sleep is not addressed further in this work, as this driver state has already been identified as a state that is not allowed when using an L3 ADS (see section 2.1). The evaluation of the influence of driver drowsiness on take-over performance requires appropriate methods to assess this driver state.

2.4 Drowsiness assessment

There are several different procedures for the evaluation of drowsiness, which can be sub-categorized into subjective, behavior-based, and physiological methods (Sahayadhas et al., 2012; Platho, Pietrek, & Kolrep, 2013). Different methods and relevant assessment tools are presented in the following. Table 2.3 provides an overview of the different methods.

Table 2.3: Overview of drowsiness assessment methods

Subjective assessments (e.g., Stanford-Sleepiness Scale (SSS), Karolinska-Sleepiness Scale (KSS))
Behavior-based assessments (e.g., steering wheel behavior, expert ratings of driver drowsiness)
Physiological assessments (e.g., eye-tracking, Electrocardiogram (ECG), Electroencephalogram (EEG))

2.4.1 Subjective assessments

Useful overviews about different tools for the subjective assessment of drowsiness were presented by various researchers (for an overview see Shen et al. (2006), Drake (2011) or Platho et al. (2013)). The SSS and the KSS have predominantly been used for the subjective assessment of driver drowsiness. The SSS is a seven-point scale, ranging from 1 (feeling active and vital, alert; wide awake) to 7 (almost in reverie; sleep onset soon; lost struggle to remain awake) (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973). A factor analysis revealed that, among other components, activation and sleepiness are assessed by the SSS (MacLean, Fekken, Saskin, & Knowles, 1992). Based on this finding, Drake (2011) estimated that this scale is more likely to assess a global state, including sleepiness and fatigue, and thus is not appropriate for a specific assessment of sleepiness. Nevertheless, a high correlation was found between SSS ratings and the performance of a mental task (Hoddes et al., 1973). The KSS (Åkerstedt & Gillberg, 1990) ranges from 1 (extremely alert) to 9 (extremely sleepy - fighting sleep) (see table 2.4).

Table 2.4: KSS according to Åkerstedt and Gillberg (1990)

1	extremely alert
2	
3	alert
4	
5	neither alert nor sleepy
6	
7	sleepy - but no difficulty remaining awake
8	
9	extremely sleepy - fighting sleep

For the KSS high correlations were found between subjective and physiological measures (Åkerstedt & Gillberg, 1990; Kaida et al., 2006) and between subjective and performance measures (Gillberg, Kecklund, & Åkerstedt, 1994; Kaida et al., 2006). However, physiological indicators of sleepiness could only be reliably detected when participants reached a KSS score between 7 and 9 (Åkerstedt & Gillberg, 1990). Further, the probability of incidents and accidents increased strongly when KSS ratings were greater than 7 (Ingre et al., 2006). Besides this, for KSS ratings greater than 7 the driver was considered to be in a state of drowsiness (Johns, 2009). Positive aspects of such subjective sleepiness scales are that their use is cost and time effective (Drake, 2011) and as described above, there is a correlation between subjective ratings, other sleepiness measures, and performance metrics. However, some criticisms need to be taken into account when using such scales. One is that self-assessment could become inaccurate with increasing driver drowsiness. For example, E. A. Schmidt et al. (2009) observed a mismatch between subjective rating (after about 3h) and objective measures of sleepiness at the end of a prolonged monotonous drive. In addition, completing of such questionnaires might be stimulating and could thus influence the rating (Brown, 1994). One study addressed these issues and found that the stimulating effect triggered by the verbal assessment disappeared within two minutes (E. A. Schmidt, Schrauf, Simon, Buchner, & Kincses, 2011).

2.4.2 Behavior-based assessments

Changes in different driving performance metrics (e.g., steering angles) are measures often used to the detect driver drowsiness (for an overview see Sahayadhas et al., 2012; Platho et al., 2013). Various studies found high recognition rates for driver drowsiness detection through steering wheel behavior (e.g., Krajewski, Sommer, Trutschel, & Edwards, D., Golz, M., 2009; McDonald, Schwarz, Lee, & Brown, 2012). However, due to the automation of the entire DDT, such metrics to detect driver drowsiness are not available when an ADS is used (Hoeger et al., 2011; Gonçalves & Bengler, 2015). Hence, the importance of reliable detection based on other non-intrusive metrics increases. In the experimental field, expert ratings of driver drowsiness[6] were found to be a reliable, consistent and non-intrusive approach (Wierwille & Ellsworth, 1994). This approach makes it possible to assess different levels of drowsiness based on various indicators such as body movements (e.g., stretching, eye rubbing) or eye-related characteristics (e.g., eye closure time, blinking rate). On the basis of these indicators, Wierwille and Ellsworth (1994) distinguished five levels of drowsiness (not drowsy, slightly drowsy, moderately drowsy, very drowsy, and extremely drowsy). This procedure graded the assessed drowsiness levels in one-minute video intervals. Furthermore, high inter-rater reliability ($r = .81$) was found. In another experiment, a strong correlation was found between participants' KSS ratings and expert ratings of drowsiness ($r = .437$) as well as a correlation between expert ratings and blink duration ($r = .360$) (Anund et al., 2013). However, this study observed that sudden changes in driver drowsiness were not reflected in experts' drowsiness ratings. According to Anund et al. (2013), this might be due to the 5-minute interval at which drowsiness was assessed. Maybe observers cannot remember all the relevant actions for this rather long period (Anund et al., 2013). The original scale for the assessment of drowsiness provided by Wierwille and Ellsworth (1994) was slightly

[6] some call this method observer-rated sleepiness (Anund, Fors, Hallvig, Åkerstedt, & Kecklund, 2013)

extended (Wiegand, McClafferty, McDonald, & Hanowski, 2009) or adapted for the assessment of driver drowsiness. For example, in the above study, three levels of driver drowsiness were assessed and four levels were assessed by Karrer-Gauß (2012, p. 66). A limitation of this approach is that it strongly depends on raters' expertise (Platho et al., 2013). Further, Platho et al. (2013) recommended that ratings should be performed at least by two independent raters. Besides, such ratings can only be used under experimental conditions to investigate participants' drowsiness level and for the development of DMSs.

2.4.3 Physiological assessments

Besides experts observing eye-related characteristics to determine drowsiness, eye-tracking systems can also objectively measure various indicators associated with an increase in drowsiness. However, Ebrahim (2016, pp. 25-26) pointed out that such systems face several difficulties. These are the time-consuming calibration before a measurement can be performed, problems in distinguishing between off-road gaze shifts and eye closures due to drowsiness, and light/reflection due to adverse light conditions or glasses (Friedrichs & Yang, 2010), which can degrade the detection quality (Ebrahim, 2016). Further, it appears that higher drowsiness levels can be more reliably detected than lower drowsiness levels (Hu & Zheng, 2009). Besides an analysis of eyelid-related parameters, an EEG can be recorded. The development of drowsiness can be observed based on the output of the EEG. However, this approach is very time consuming and intrusive (for an overview of these methods see Platho et al. (2013) or Ebrahim (2016, pp. 16-20)). Also, based on heart-rate/heart rate variability metrics an ECG can provide information about a driver's drowsiness level (Jung, Shin, & Chung, 2014). Though an ECG is also intrusive, it is considered less intrusive than an EEG (Sahayadhas et al., 2012).

2.5 Driver Monitoring Systems

The *"Euro NCAP 2025 Roadmap"* highlights the relevance and safety potential of drowsiness and attention monitoring systems. According to this roadmap, an incentive for DMSs is planned. The assessment of such systems will include the reliability and accuracy of the state detection and the related actions that a vehicle would perform depending on the detection function (Euro NCAP, 2017).

Further, it emphasizes that:

> *"Effective driver monitoring will also be a prerequisite for automated driving, to make sure that, where needed, control can be handed back to a driver who is fit and able to drive the vehicle. This item will be taken on board under the HMI requirements for Automated Driving."* (Euro NCAP, 2017)

Besides, Yoshida (2016) pointed out that:

> *"Driver monitoring is no longer a nice-to-have feature. In the era of self-driving cars, it becomes a must-have technology, because drivers, some of the time, are still going to be required behind the steering wheel."*

According to Hörwick (2011, p. 37), two different approaches exist for driver monitoring. One approach is based on observing a driver's behavior while the other focuses on specific forced driver inputs. These approaches are presented below.

For assessing driver behavior, various directly observable metrics (e.g., head position, eyelid closure, vital parameters) and several indirect metrics (e.g., detection of steering errors, hands-on detection, pedal actuation) can be assessed. Hence, behavior-based and physiological metrics are considered for this approach (Hörwick, 2011, p. 37).

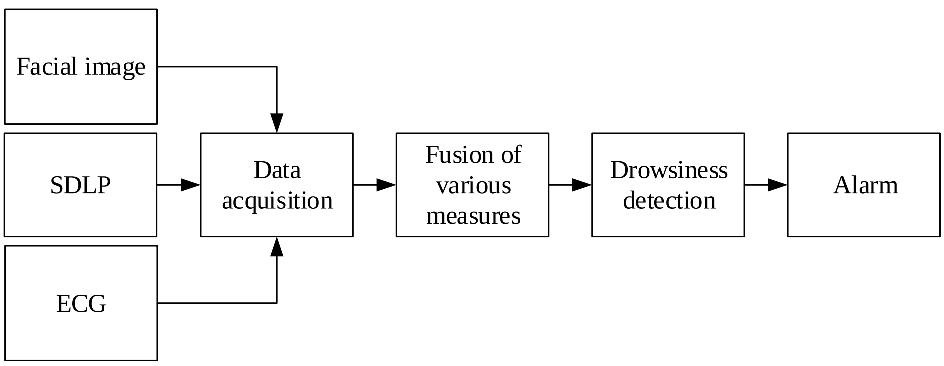


Figure 2.8: A hybrid drowsiness detection system (figure based on Sahayadhas et al. (2012))

However, as the introduced methods have different advantages and disadvantages (see section 2.4), several researchers have proposed the use of hybrid drowsiness detection systems (or DMSs) (figure 2.8 provides an example of such an approach) (Rauch, Kaussner, Krüger H.-P., Boverie, & Flemisch, 2009; Sahayadhas et al., 2012; Čolić, Marques, & Furht, 2014).

Various solutions of DMSs have been developed and implemented in series-vehicles for years. For example, among other metrics, changes in the steering wheel behavior are used to predict driver drowsiness (e.g., Mercedes Benz (n.d.) and Volkswagen (n.d.)) (for an overview see Čolić et al. (2014)).

Especially with increased automation, it becomes evident that only a few metrics remain for an estimation of the driver state (Hoeger et al., 2011; Gonçalves & Bengler, 2015). Besides this, it was found that informative and warning DMSs without additional system interventions have the potential to change users behavior in the context of manual driving (Eichinger, 2011, p. 23). Users of a L3 ADS will not be allowed to sleep (see section 2.1). Thus, it can be deduced that the recommended nature of DMSs is likely to become more restrictive when an L3 ADS is used.

Further, Bubb, Bengler, Grünen, and Vollrath (2015, p. 565) pointed out that of the direct driver monitoring methods, only camera-based driver monitoring remains an option for series vehicles. So far, driver monitoring cameras were integrated into some production vehicles that offer L2 or lower systems. In 2006, the Lexus GS 450h was the first vehicle that was equipped with an additional driver monitoring camera (Lexus Europe Newsroom, 2016). This camera was located on the steering column (Toyota Global Newsroom, 2008). Since then, other manufacturers have equipped their vehicles with driver monitoring cameras (DS7 Crossback (DS, 2017) and the Cadillac CT6 (Cadillac, 2017)).

The second approach forces users to perform specific actions. Based on whether a user responds to a reaction request, it is determined whether a user's state is still sufficient (Hörwick, 2011, p. 36). This concept is known as the "dead man's switch" in rail transport and requires a train driver's input to ensure that an emergency stop does not need to be performed (Patentschrift, 1922).

Consequently, this approach would be a monitoring system without a need for camera-based state detection. However, researchers proposed using hybrid driver state assessment approaches even for lower automation levels (Hoeger et al., 2011; Wimmer, 2014). Hence, besides responses to specific triggers, camera-based and environmental metrics are also considered (Hoeger et al., 2011; Wimmer, 2014; J. Schmidt, Braunagel, et al., 2016). The ”potential-trigger” concept of Wimmer (2014, pp. 74-79) aims to ensure a driver’s motor responsiveness and an appropriate level of situation awareness while using an L2 system.[7] For this, a driver must press a specific button when a potential level is fallen short of. A user’s potential level is estimated based on the TTC, on a driver’s HMI interaction, and the driver’s gaze behavior. When the driver presses the correct button, the potential level will be refilled. Operating errors and ToT metrics did not differ between the frequent (31.6s) and the medium (44.3s) trigger intervals (Wimmer, 2014, p. 164). Therefore, Wimmer (2014, p. 165) suggested using the medium trigger interval of 44.3s for further functional development. The most prolonged, and thus least frequent interval of 81.0s led to significant longer take-over times compared to the frequent condition in one of four test situations.

Figure 2.9 shows the algorithm used for an alert request of an L3 ADS. In a driving simulator study, the influence of different time intervals (30s compared to 180s between two requests) on driver drowsiness and ToT were investigated. The development of driver drowsiness and ToT did not differ significantly between these two intervals (J. Schmidt, Stolzmann, & Karrer-Gauß, 2016).

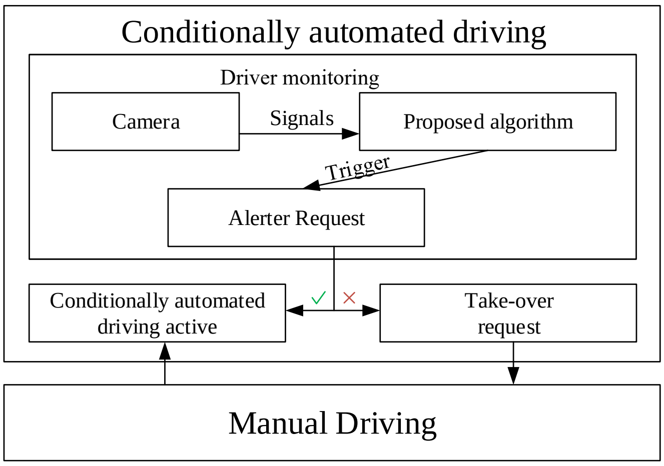


Figure 2.9: Alertness request (figure taken from J. Schmidt, Braunagel, et al. (2016))

Overall, it can be concluded that DMSs are gaining in relevance, especially as automation increases. However, several limitations of DMSs are already known

[7] original named VA (Vollautomation)

for lower automation levels and must be taken into account.

In 1998 Dinges and Mallis (1998) warned that *"there is a risk of rush toward widespread use of technologies that do not reliably detect fatigue"* and that these systems need to be investigated by engineering criteria like validity, reliability, generalization, sensitivity, and specificity. Also, Anund and Kecklund (2011) pointed out that the effectiveness of implemented DMSs has not yet been demonstrated in independent studies. Besides that, a high rate of *"false-alarms"* (alarm, though the driver is not drowsy) and *"misses"* (no alarm, though the driver is drowsy) is still a major problem of all drowsiness detection systems - regardless of the method used (Bubb et al., 2015, p. 565).

Bengler et al. (2014) pointed out that although such systems have already been launched, the *"effects and consequences of drowsiness detection, especially in terms of acceptance by the driver, are, however, not yet profoundly researched."*

Also, misuse of such drowsiness detection systems must be considered, for example, it was found that some professional drivers consider the reactivation nature of the drowsiness feedback useful, as they can continue to drive even though they are experiencing drowsiness (Karrer-Gauß, 2012, p. 161). Therefore, Karrer-Gauß (2012, p. 161) emphasized that it is important to reactivate a driver temporarily by feedback, but that also solutions are needed to ensure that such detection systems are not misused as a kind of alarm clock. In addition, the feedback must be designed in a way that it alerts drivers without causing startled responses (Knipling & Wierwille, 1994, p. 8).

2.6 Drowsiness and take-over performance

Various accident reports and simulator studies show that there is a link between an increase in drowsiness and a decrement in driving performance. A decrease in driving performance, for example, can be observed by delayed or missing driver responses (e.g., Sagberg, 1999; Filtness, Armstrong, Watson, & Smith, 2017). As this relation was extensively investigated for years in the context of manual driving, this section focuses on the interaction between drowsiness, L3 ADSs, and take-over performance.

Several studies focused on the development and the influence of driver drowsiness or fatigue on take-over performance when using an L3 ADS. Hence, the similarities and differences between these studies are analyzed. Results of this review are presented in the following.

It is apparent that the different studies often do not point out whether they aim to investigate drowsiness or fatigue and to what definition they refer. Also, terms are used interchangeably with one another without further explanation. For example, fatigue is assessed by using a sleepiness scale (in these cases by the KSS) (Jarosch, Kuhnt, Paradies, & Bengler, 2017; Kreuzmair, Gold, & Meyer, 2017). Therefore, in the following, the terms originally used by the various researchers are also used to represent the similarities and dissimilarities of the studies and the corresponding results. Also, researchers (e.g., Jarosch et al., 2017) refer to the model provided by May and Baldwin (2009) and build up their studies based on it. This might not be an optimal approach (see section 2.3), as it makes a clear distinction difficult or sometimes even impossible between the investigated factors and their impact on take-over performance (drowsiness, fatigue or type of NDRT). The methods for assessing the influence of drowsiness or fatigue on take-over performance can be categorized into two methodical approaches.

One is the manipulation of the automation duration (referred to as fixed-time approach in the following). In this approach, requests to intervene are triggered after a particular predetermined time. Feldhütter, Gold, Schneider, and Bengler (2017) found that after 18 minutes of automation, participants needed more time to look away from the NDRT to the driving scene than after three minutes. In this context, the authors pointed out that this is either due to the automation duration or the task. Other take-over performance metrics were not significantly influenced by the automation duration (Feldhütter et al., 2017). The fixed-time approach was also used by Jarosch et al. (2017). In this study, a system limit was reached after 24 minutes. Different fatigue states were induced by two types of NDRTs. One was a monotone monitoring task in which participants were asked to tap a tablet whenever the lowercase p was displayed. The other NDRT was a quiz that aimed to be motivating. Fatigue was assessed by the KSS and several eye-tracking parameters (PERCLOS, blink rate, blink duration). The KSS ratings increased significantly over time and fatigue was significantly greater considering the behavioral assessment in the monitoring compared to the quiz condition. Also, no significant influence of the different fatigue states (induced by task and automation duration) on take-over performance metrics was found in this study (Jarosch et al., 2017).

The second approach (hereafter referred to as a state-dependent approach) triggers take-over situations depending on a user's drowsiness (or fatigue) level. Hence, due to the state-dependence, the automation duration differs among participants. A driving simulator study was aimed to get users tired or sleepy by using an L3 ADS (Kreuzmair et al., 2017). Also, to induce higher levels of fatigue, this experiment started at 6 a.m. and participants were asked not to go to bed the previous night. Participants subjectively assessed their fatigue level on the KSS. Besides this, investigators rated participants' fatigue level also on the KSS. Take-over situations were triggered depending on these expert ratings (baseline: KSS $\leq$ 5, slight fatigue: KSS = 6 or 7, fatigue: KSS = 8 or 9 and KSS = 10). Experts assessed the beginning of sleep. Thus, the KSS was extended for this study by adding a level 10. No statistically significant influence of the different fatigue levels on take-over time or quality was found in this study. Also, J. Schmidt, Dreißig, Stolzmann, and Rötting (2017) did not observe an influence of fatigue on ToT and quality. Their study lasted for about 2h 51m. In total, five take-over situations were implemented in this study. Situation 5 was triggered between 108 and 130 minutes depending on whether the participants missed two alertness requests displayed for 5 seconds. Between 130 and 180 minutes after the beginning ot the experiment, the take-over situation was triggered if the participants missed responding to only one request. After that time span, the take-over situation was triggered automatically. During this experiment, mean KSS ratings increased from 4.29 to 7.88 in the last situation (J. Schmidt et al., 2017). Take-over requests were also triggered state-dependent in another driving simulator study (Gonçalves, Happee, & Bengler, 2016). In this study a take-over situation was triggered after participants assessed their drowsiness level as 5 on the SSS (*"foggy; losing interest in remaining awake; slowed down")* in the drowsy condition and after 3 minutes in the reference conditions. No significant influence of drowsiness was found compared to the reference condition regarding take-over time metrics. However, drowsiness significantly increased the lateral acceleration in a take-over situation compared to the reference condition (Gonçalves et al., 2016).

To sum up, this review shows that the constructs drowsiness and fatigue are often mixed in related literature. Besides the different approaches to assessing driver state changes, the measures to induce driver drowsiness or fatigue and the take-over scenarios used in these studies differed. Table 2.5 provides an overview of the main characteristics of state induction and the investigated take-over situations. Further, table 2.5 shows that the development and the influence of driver drowsiness (or fatigue) on take-over performance have been limited to driving simulator studies. However, according to Körber and Bengler (2014) there still is a great need for real-traffic studies, especially to study automation effects.

Further, when looking at the ToTs of the different studies, no adverse influence of drowsiness or fatigue on take-over time metrics was observed. Figure 2.10 presents the mean values of ToT or hands-on time metrics and their link to the automation duration (also including task effects) and to the different drowsiness or fatigue levels found in the various studies (if these were explicitly reported). Also, the determined take-over time metrics were similar to those of other studies that did not aim to induce drowsiness or fatigue. For example, a mean hands-on time of 1.45s was found for a total time budget of 5 seconds for non-drowsy drivers. For a total time budget of 7 seconds, the mean hands-on time slightly increased to 1.70s (Gold et al., 2013).

Table 2.5: Take-over situations and approaches for drowsiness and/or fatigue generation

Study	**Take-over situation and state induction**
(Feldhütter et al., 2017) N = 30	• Driving simulator study • Total time budget: 6s • 120km/h, left lane, stationary obstacle • fatigue induction by manipulation of automation duration (5 minutes compared to 20 minutes.) • NDRTs: no task and Surrogate Reference Task (SuRT)
(Jarosch et al., 2017) N = 56	• Driving simulator study • Total time budget: n.a. • accident on the ego lane and a sensor failure in a bend • fatigue induction by automation duration (24 minutes) and by the type of NDRTs (monitoring task compared to a quiz task)
(Kreuzmair et al., 2017) N = 22	• Driving simulator study • Total time budget: 5.28s • 120km/h, simple situation: drive through a roadwork, complex situation: roadwork requires the performance of a lane change • fatigue induction by several measures: experiment started at 6 a.m. (participants were also asked not to go to bed before midnight), no coffee consume before the experiment and long monotone highway drive with an L3 ADS
(Gonçalves et al., 2016) N = 31	• Driving simulator study • Total time budget: 5s • a stationary obstacle in the ego lane • fatigue induction by the time of day (between 2 p.m. and 4 p.m.) and eating before the start of the experiment (no further NDRT - devices were removed before the start of the experiment)

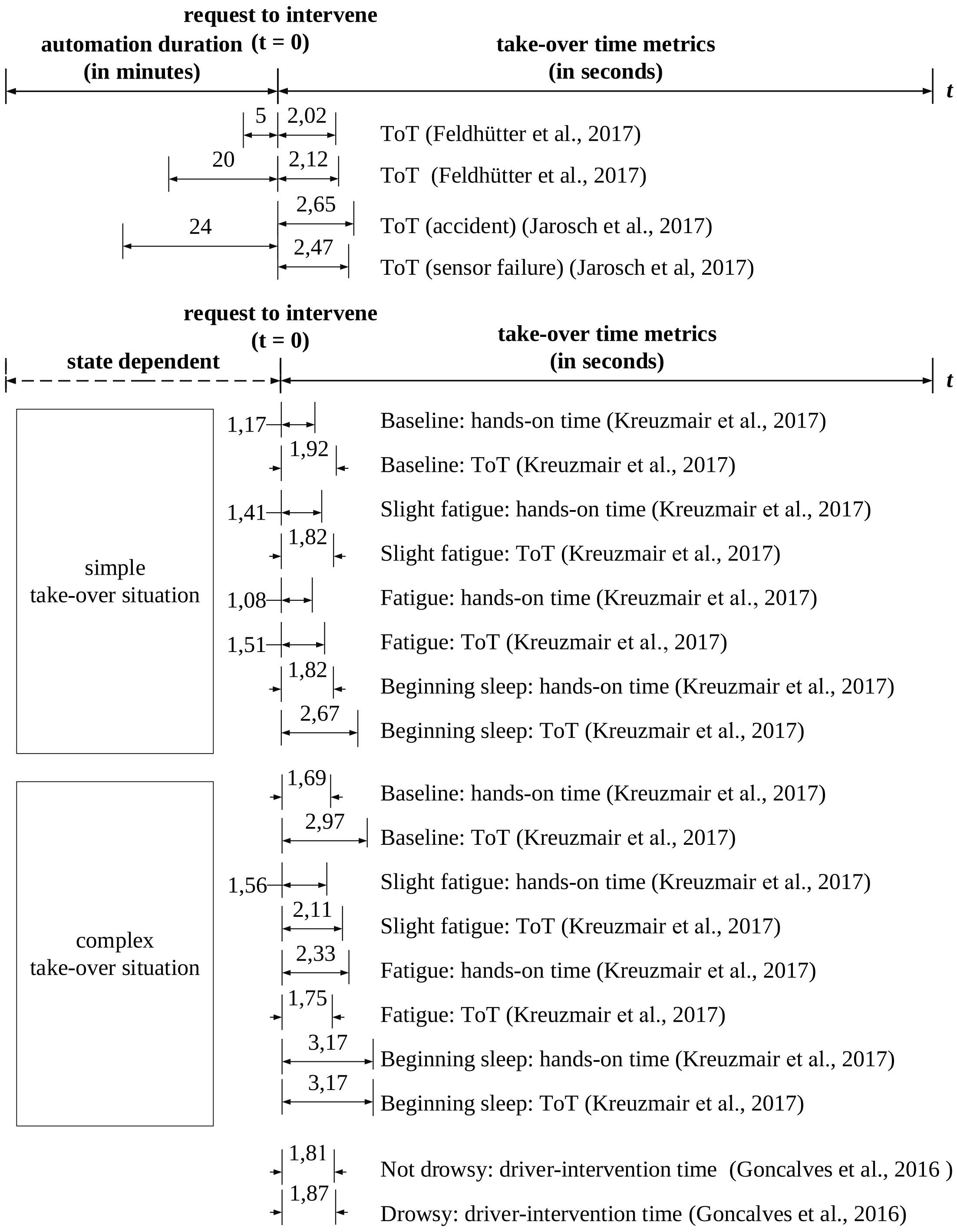


Figure 2.10: An overview of the different approaches and their influence on ToT metrics

2.7 Problem statement

In summary, drowsiness was identified in the previous chapters as a factor that has the potential to impair users' receptivity and thus a safe transition from L3 to lower automation levels. As users of an L3 ADS are still not allowed to sleep (see section 2.1) and so far only an intrusive EEG – which does not come into question for series vehicles (Bubb et al., 2015, p. 565) – can assess sleep stages, as well as based on the findings of Gonçalves et al. (2016), this thesis assumes that driver drowsiness as a *"transitional state between wakefulness and sleep"* (Johns, 1998) represents another system limit. It also is clear that as users of an L3 ADS are still required to represent a DDT fallback, DMSs will gain in relevance to ensure that the users' driver state is appropriate. However, such systems face several limitations (see section 2.5).

Consequently, the interaction between an L3 ADS and a DMS must be carefully designed. It is also crucial to identify effective strategies for managing driver drowsiness when using an L3 ADS. For this, appropriate test methods to induce and assess this specific driver state are essential. Besides these methodological challenges, it should be borne in mind that the development and the influence of driver drowsiness on take-over performance have been limited to driving simulator studies.

CHAPTER 3

Development of a Drowsiness Management Concept (DMC)

3.1 Need and challenges of a DMC

A user of an L3 ADS who has passed beyond the driver state drowsiness and has fallen asleep is considered *"unresponsive"* and likely cannot take over control safely, especially in a complex situation. This poses a potential security risk. Thus, the driver state drowsiness, which is passed through shortly before one falls asleep, is considered another system limit of an L3 ADS and must be managed. Further, users wish to use ADSs, in particular when they are tired (Payre, Cestac, & Delhomme, 2014). Hence, there might be a conflict between user expectations and L3 ADS requirements on the users' driver state. In summary, a system that hands back the driving task to a user whose driver state was classified as *"extremely drowsy"*, appears unacceptable from a safety, the user's, other road user's and Original Equipment Manufacturers (OEM) perspective.

Thus, a DMS should be integrated into a vehicle with L3 ADSs functionality to determine whether the system limit driver drowsiness is reached and to be able to take effective measures. In this context, manufacturers must decide if and what kind of DMS they will integrate.

Further, they may choose between a conservative functional design that would already consider a lower level of driver drowsiness as a system limit or a less conservative design that would consider only extreme levels of driver drowsiness as a system limit. Consequently, the latter approach would provide an L3 ADS over a longer period, but might impair a safe transition or the subsequent drive when automation levels lower than L3 are used.

To sum up, strategies are needed to provide solutions to manage the driver state drowsiness from a user's and manufacturer's perspective. The equally important perspective of other road users is not addressed further in this thesis. An investigation of the perspective of other road users will be meaningful only at the point when knowledge about safe, likely accepted (from the user perspective) and effective strategies exist.

3.2 Elements of a strategy[1]

According to Drucker (2009, p. 352) strategic decisions *"involve either finding out what the situation is, or changing it, either finding out what the resources are or what they should be"*. Rumelt (2013, p. 2) pointed out that *"discovering the critical factors in a situation and designing a way of coordinating and focusing actions to deal with those factors"* are core elements of strategic work. Concluding from these definitions, the identification of strategies for managing driver drowsiness in the context of automated driving will require *(a)* an analysis of the present situation and *(b)* an identification of critical factors.

Analysis of the present situation

An assessment of the current situation requires knowledge about the ADS and a user's drowsiness state. The usefulness for a DMS to be integrated into an L3 ADS vehicle has already been derived in the previous section. Thus, the ADS (according to the SAE (2018)) is supplemented by a DMS. The L3 ADS must also be able to detect system limits and trigger requests to intervene. Besides the system limit driver drowsiness, other system limits, such as reaching a motorway exit or stationary objects in front of the vehicle can occur (Bahram, Aeberhard, & Wollherr, 2015).

[1] the following sections are mainly based on the following publication: Weinbeer et al. (2018)

Identification of critical factors

In chapter 2, two types of critical reactions of drowsy users of an L3 ADS were identified when it comes to a take-over situation accompanied by a request to intervene. First, drowsy drivers might need more time for a sufficient understanding of the current situation. Second, drowsy drivers might react startled in case of an unexpected take-over situation.

3.3 Strategies to manage driver drowsiness

To identify suitable strategies for managing driver drowsiness when an L3 ADS is used, the basic mechanisms of the sleep and wake drive must be considered. For this, the four-process model provided by Johns (1998) is taken into account (see section 2.3). In the context of automated driving, the secondary wake drive appears to be extremely relevant, as this is the drive which can be influenced by somatosensory input (Johns, 1998) and by emotional and cognitive inputs (Saper et al., 2005). Further, Johns pointed out that in most cases the secondary wake drive determines whether a person falls asleep and that this drive can change within seconds (Johns, 2000).

3.3.1 Driver-state related strategy

A driver-state related strategy (DSRS) aims to minimize a driver's drowsiness level to avoid the system limit driver drowsiness. Anund and Kecklund (2011) derived sleepiness countermeasures for lower automation levels following the taxonomy of Michon (2985), who categorized a road user's task into a strategic, tactical, and operative level (see table 3.1).

For example, in a driving simulator study, it was observed that hitting milled rumble strips had a short alerting effect. However, indicators of sleepiness occurred again after 5 minutes (Anund, Kecklund, Vadeby, Hjälmdahl, & Åkerstedt, 2008). Further, different studies investigated the influence of, for example, caffeine (J. A. Horne & Reyner, 1996; De Valck & Cluydts, 2001), naps (J. A. Horne & Reyner, 1996), alertness-maintaining tasks (Oron-Gilad, Ronen, & Shinar, 2008), cooling (E. Schmidt, Decke, Rasshofer, & Bullinger, 2017), and warning modalities (Gaspar et al., 2017) on driver drowsiness.

Table 3.1: Countermeasures on a strategic, tactical, and operative level taken from Anund and Kecklund (2011)

Strategic	**Tactical**	**Operative**
Fatigue management systems	Driver support system (feedback - warning)	Rumble strips
Hours of service regulations	Road signs	Driver support systems (warning and intervention)
Information/education	Parking areas	Campaigns targeted to increased awareness of alertness-enhancing activities that can be carried out during rest
Strategies for planning	Route guidance to parking areas	
Fit for duty test		
Enforcement/control		

Several researchers provided various reviews of drowsiness countermeasures for manual driving (for the reviews see, e.g., J. A. Horne and Reyner (1995), Cummings, Koepsell, Moffat, and Rivara (2001), and Davidsson (2012)). In light of these reviews and study results, it is clear that the options to minimize drowsiness during a less automated drive are limited and the effectiveness of most of the countermeasures - except for a nap or caffeine - are often not clear and/or consistent and are usually not effective beyond the original activation.

However, when an L3 ADS is active, users are not required to monitor the system and can perform more motivating tasks. This could help prevent users from falling asleep or at least prolong the period during which the drivers' drowsiness state is acceptable. One concept that could potentially prevent further increase of driver drowsiness was presented at the CES 2019 (AUDI AG, 2019). Based on film content, various additional components such as climate, vibration, and active chassis are controlled. As this involves different senses, this method aims to avoid or to reduce driver drowsiness during an automated drive (Heinze & Weinbeer, 2018). These and other NDRTs can have a positive influence on the secondary wake drive over a longer period (see section 2.3). Two studies exist that support this assumption (Jarosch et al., 2017; Schömig, Hargutt, Neukum, Petermann-Stock, & Othersen, 2015). For example, KSS ratings are less when a quiz was performed compared to a monitoring task during a drive with an L3 ADS (Jarosch et al., 2017). Also, no further increase in participants' drowsiness level was found when participants performed a quiz task in another study (Schömig et al., 2015).

Also, it is relevant to keep in mind that a DSRS should be designed so that users are not forced to perform only a few specific NDRTs permanently during the entire drive. This is because *"systems that restrict the driver's behaviour or policing systems that force a behavioural change are likely to be less accepted than nonrestricting, informative systems [...]"* (van der Laan, Heino, & de Waard, 1997). Consequently, a DSRS should only be executed when needed, and hence when users already experience a certain level of drowsiness. Thus, the reactivation potential and effectiveness of further NDRTs (especially when drivers are already experiencing drowsiness) must be sufficiently investigated. Besides this, it must be considered that measures against sleepiness often (if at all) only have a brief alarming effect (Brown, 1994; Anund et al., 2008) and that the best solution is to stop driving as soon as possible (J. Horne & Reyner, 1999).

Thus, the developed Drowsiness Management Concept (DMC) allows a drowsiness level (DL_x) to be exceeded only once, accompanied by the offer of the DSRS. If this strategy fails and DL_x is exceeded on one more occasion, a driver's drowsiness level is considered a system limit.

3.3.2 System-based strategy

For lower automation levels, Trivedi, Gandhi, and McCall (2007) pointed out that it is useful to adapt various warning or intervention system parameters depending on the interaction between environmental, vehicle, and driver state metrics. This can be transferred to L3 ADSs. Hence, a system-based strategy (SBS) aims to ensure vehicle safety when there is any uncertainty as to whether or not a driver has an adequate drowsiness level and therefore might not be able to take over control safely. One option of this strategy would be that the system no longer makes lane changes to be prepared when a minimal risk condition needs to be reached.[2]

[2] An L3 ADS according to the SAE (2016) does not have to be able to perform a minimum risk maneuver. However, also an L3 ADS can (though not required per definition) be equipped with such a functionality.

Depending on the type of system failure, the options in which a minimum risk condition can be reached may differ (SAE, 2018, p. 11):

> "*It may entail automatically bringing the vehicle to a stop within its current travel path, or it may entail a more extensive maneuver designed to remove the vehicle from an active lane of traffic and/or to automatically return the vehicle to a dispatching facility.*"

Other options of a SBS may also be performed. For example, a speed reduction could increase the time budget and decrease the intensity of the needed deceleration if it is necessary to stop the vehicle. The adjustment of speed considering a constant deceleration was calculated and illustrated by Bahram et al. (2015). Besides this, an L3 ADS can return the driving task to the driver to avoid a further increase in driver drowsiness. However, as with the most DSRS options, the effect of such a transition was found to be only temporarily useful (Schömig & Kaussner, 2010). Besides this, requests to intervene might lead to startle responses when users are drowsy. In this case, a preparation strategy might be an appropriate strategy to reduce unwanted driver reactions.

3.3.3 Preparation strategy

A preparation strategy aims at reactivating the driver within a short period and avoiding startle reactions. Hence, suitable driver state-related and system-based options are performed simultaneously. This strategy should be executed if the system limit is driver drowsiness and no further system limit (e.g., sensor failure) exists. Drivers can obtain specific information on the current situation (such as speed limits) to prepare them for the following less automated drive. Also, additional system-based options should be performed to enhance overall safety.

3.4 DMC: L3 ADS and DMS interaction

The findings and considerations reported were grouped into the following drowsiness management concept (see figure 3.1). Part A represents the technical system. Part B shows the developed state machine. In this DMC, the different strategies do not represent alternatives to each other. Instead, the different strategies are triggered based on logical operations.

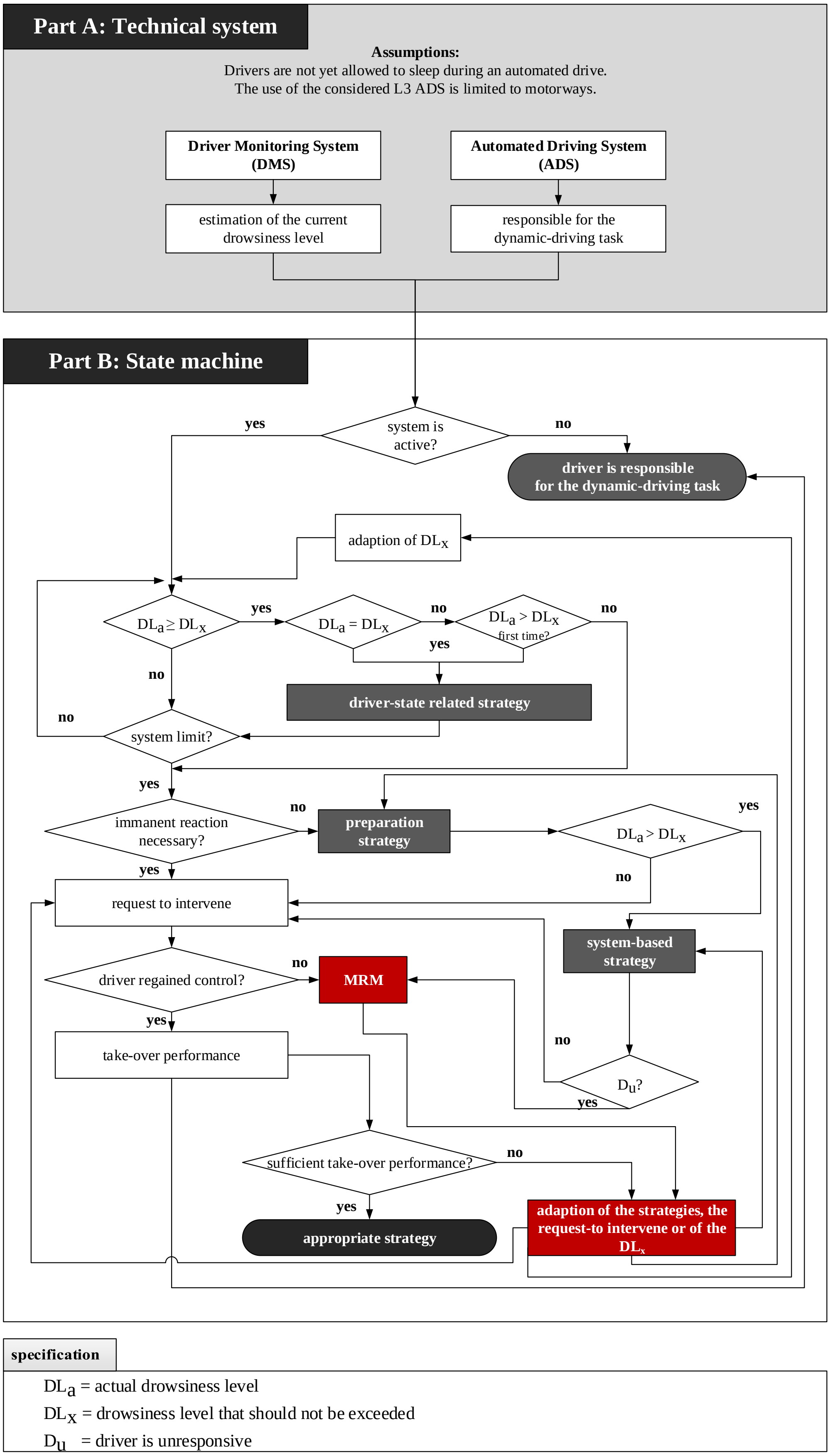


Figure 3.1: Drowsiness Management Concept (DMC) taken from Weinbeer et al. (2018)

Besides assuming that driver drowsiness represents a system limit, this DMC is made to assume that the considered L3 ADS is limited to motorways. The DMC is based on the interaction between an L3 ADS and a DMS (both together are seen here as the technical system). This interaction is required to trigger various strategies that depend on a user's drowsiness state. The DMS considered here is assumed to assess different drowsiness levels based on the observation of a user's direct behavior by a driver monitoring camera. Also, the approach that a DMS requires frequent specific driver inputs would be possible. However, this approach has not yet been able to identify different drowsiness levels. In addition, it would change a monitoring task (associated with lower automation levels) into another monitoring task that would likely represent a less-accepted approach than a non-intrusive, non-mandatory camera-based DMS approach. This can also be argued by the concern that systems that are more restrictive are likely to be less accepted than non-restrictive systems (van der Laan et al., 1997).

The derived DMC assumes that the camera-based DMS can assess different drowsiness levels consistently and reliably. Previous research, however, has shown that a user's drowsiness state cannot always be reliably estimated due to various circumstances (e.g., due to unfavorable light conditions (Friedrichs & Yang, 2010), see also section 2.4). Therefore, it is particularly important to be able to deal with DMS-uncertainty periods, as incorrect timing of the different strategies could lower their acceptance, effectiveness and could even become safety critical. The latter would be the case when a driver's drowsiness state is erroneously not considered the system limit driver drowsiness, although in reality, a user has even fallen asleep.

CHAPTER 4

Research questions and study overview

As described above, the developed DMC faces several challenges. Therefore, this thesis aims to investigate various crucial issues related to this concept. These are:

(1) Which driver state-related or system-based strategy options would be accepted by users?

(2) Are non-driving-related tasks effective in reducing driver drowsiness?

(3) How to design the DMC tolerant towards uncertainty periods of a DMS?

Overall, three studies were conducted to answer these research questions.

***Study 1* (N = 31)**
For the investigation of the issues mentioned above, it is essential to identify an appropriate method to induce driver drowsiness and to simulate an automated drive in real traffic without compromising road safety. Thus, Study 1 aimed *(a)* to provide a methodical basis to investigate drowsiness effects in this specific context. Besides, *(b)* the influence of different drowsiness levels on take-over time metrics was evaluated. Further, *(c)* fundamental knowledge on likely accepted driver state-related or system-based options has been acquired from a user's perspective through subjective assessments.

***Study 2* (N = 71)**

This study assessed the reactivation potential and the effectiveness of non-driving-related tasks. As a targeted offer of non-driving-related tasks (NDRTs) was rated the most accepted driver-state related option in Study 1 (users' perspective), the potential of this option was investigated.

***Study 3* (N = 24)**

According to the DMC, strategies are executed depending on a user's drowsiness level. Hence, DMS uncertainty periods in which a user's drowsiness level cannot be reliably detected (e.g., due to unfavorable light conditions) could lead to a wrong implementation of the different strategies. Thus, such a DMC is tolerant when it prevents the different strategies from being triggered as a result of DMS uncertainty periods. Various approaches were identified to deal with this issue. In total, four concepts were derived and evaluated with regard to users' automation trust and acceptance.

CHAPTER 5

Methodical basis and subjective assessment of different strategies[1]

Abstract: The literature review (see section 2.6) has shown that there is a lack of appropriate test methods for the investigation of drowsiness effects when L3 ADSs are used in particular in real traffic. Therefore, a Wizard-of-Oz approach was chosen to investigate the effectiveness of inducing drowsiness in the real driving environment. For this, a right-hand-drive vehicle (Audi Q7) was modified. Drowsiness levels (DLs) were assessed about every two minutes by two investigators sitting on the back-seat. In total, 31 participants took part in this experiment. It was found that participants could get a feeling of automated driving. Further, it became apparent that drowsiness develops quite individually. Consequently, it can be concluded that a state-dependent approach is required to examine the effectiveness of different strategies and the influence of drowsiness on take-over performance. As 76.67 percent of the sample reached drowsiness level 4 and 63.33 percent experienced the highest drowsiness level 6, this study clearly demonstrates that it is possible to induce drowsiness without sleep deprivation by controlling several factors (e.g., caffeine). Nevertheless, no statistically significant differences in ToT metrics were found between the different drowsiness levels. However, some

[1] this chapter mainly bases on the following publications: Weinbeer et al. (2017) and Weinbeer et al. (2018)

participants were startled when a request to intervene was triggered in the drowsy but not in the non-drowsy condition. Subjective assessments revealed that higher escalations such as a Minimum Risk Maneuver (MRM) should be avoided from the users' perspective and that of the driver state-related options a targeted offer of non-driving-related tasks got the most support. This study provides a basis for evaluating drowsiness effects and different strategies for managing driver drowsiness when an L3 ADS is used.

5.1 Introduction

First, induction of driver drowsiness is an enormous challenge not only for investigating drowsiness effects but also for evaluating the effectiveness and acceptance of possible strategies. So far, different approaches have been used to induce driver drowsiness or fatigue in the context of automated driving (for an overview see table 2.5). Several measures which were taken in these studies, such as performing experiments at certain times of day, longer periods of automated driving, sleep-deprivation, having lunch and avoiding caffeinated beverages appear to be useful for the induction of driver drowsiness. The aim to induce sleepiness (or rather fatigue) by means of a monitoring/vigilance task (Jarosch et al., 2017) was not considered optimal in the context of automated driving, as monitoring per definition is not required during the use of an L3 ADS (SAE, 2018). When taking into account the affect circumplex model, fatigue is a low-activation negative affect (in the quadrant of depression), whereas drowsiness is a low-activation positive affect (in the quadrant of comfort) (Warr, 2013). As L3 ADSs aim to enhance comfort (Hoeger et al., 2011), drowsiness and not fatigue appears to be the relevant driver state in the context of L3 ADSs.

To determine whether a fixed-time or state-dependent approach is best suited to the research question, knowledge of the development of driver drowsiness while using an L3 ADS is needed. So far, the development of driver drowsiness and its influence on take-over performance have only been assessed in driving simulator studies. To gain knowledge about the progress of driver drowsiness in real traffic, this thesis uses a Wizard-of-Oz approach. For this specific research field, a rather simple Wizard-of-Oz approach similar to the RRADS concept (Baltodano, Sibi, Martelaro, Gowda, & Ju, 2015), which does not allow participants to intervene in the real driving process, seems to be more appropriate than the rear-seat Wizard-of-Oz approach introduced by Kiss, Schmidt, and Babbel (2006), because this concept would allow participants to intervene in the real driving process.

5.2 Method

Participants

The sample consisted of 31 employees of the AUDI AG (females: n = 12 and males: n = 19). On average, participants were 31 years (SD = 8) old and have held their driving license for 13.45 years (SD = 8).

Test vehicle and test route

A right-hand-drive vehicle was modified to investigate drowsiness in a simulated L3 drive in real traffic (see figure 5.2 (a)). Participants were never able to intervene in the real driving process. This setting was chosen because it cannot be assumed that every participant will take over control safely when experiencing higher drowsiness levels. Therefore, an additional steering wheel and pedal dummies were installed in the vehicle (see figure 5.2 (d) and (e)). The steering wheel can be turned to the right and left until stop positions are reached (see figure 5.2 (f)). Integrated micro-switches enabled the recording of different signals from the pedals and the steering wheel (e.g., reaching the left or right stop position). Also, driving-school mirrors and an additional rear-view mirror were installed so that participants were able to observe the surrounding traffic. A pilot button (see figure 5.2 (c)) and an LCD screen on the steering wheel provided information about the current pilot state (pilot not available, pilot available, pilot active and please take over). The request to intervene was also presented acoustically (using the take-over sound of the series Adaptive Cruise Control (ACC)). Three LCD displays (see figure 5.2 (e)) were attached to the dashboard to display abstract take-over situations. In this case, stylized brake lights represented the end of a traffic jam. Thus, to handle these take-over situations, participants should react as realistically as possible by using the steering wheel dummy and the pedal dummies. Various cameras were installed to analyze participants' hands-on times. Using these video recordings, hands-on times were analyzed afterward. The experiment started at the motorway entrance Lenting (Germany). The ADS was simulated from Lenting to the interchange Nuernberg-Ost and back again (see figure 5.1). The maximum speed was 130km/h, and lane changes were performed very cautiously. No assistant systems such as ACC or the Lane Keeping Assist (LKA) were used in this study.

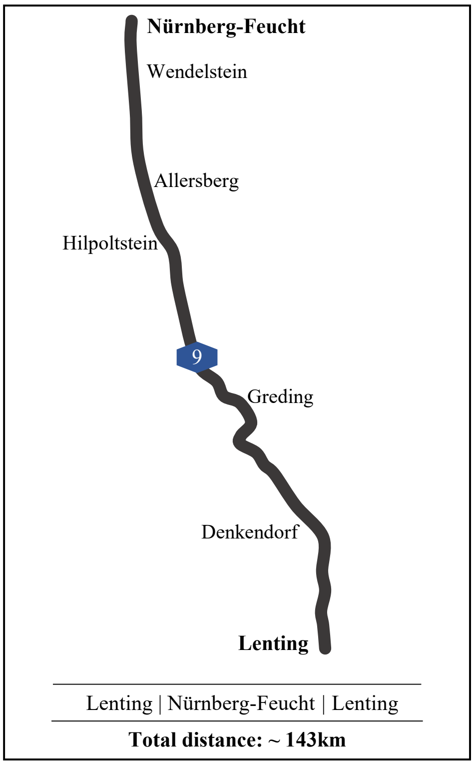
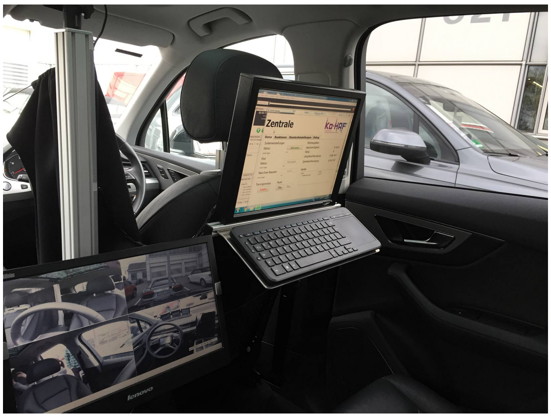


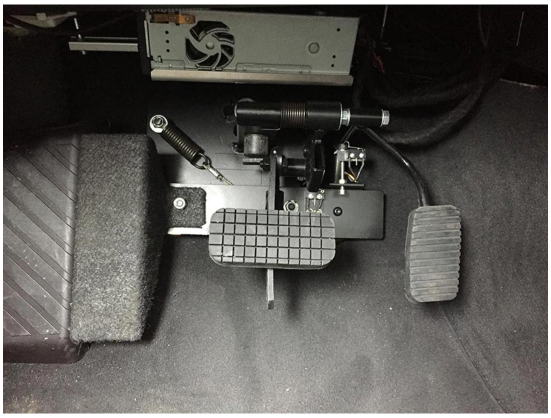
Figure 5.1: Study 1: Test route

(a) Modified right-hand-drive vehicle Audi Q7



(b) Rear seat view

(c) Integrated pilot button



(d) Integrated pedal dummies

(e) steering wheel dummy and abstract take-over situation

(f) Stop position of the steering wheel dummy

Figure 5.2: Wizard-of-Oz approach: Modified right-hand-drive vehicle

Drowsiness assessment

In this study, two raters assessed drowsiness similarly to the procedure proposed by Wierwille and Ellsworth (1994). Based on the considerations of Karrer-Gauß (2012), one level was added to this original scale, and based on the considerations of Wiegand et al. (2009), some indicators were added to drowsiness level 2. Table 5.1 provides an overview of the used scale[2]. As described in chapter 2.4, there are different pros and cons of each drowsiness assessment method. As this study aims to provide a methodical basis for the evaluation of drowsiness and of strategies for managing this driver state, it is essential to find out if high levels of drowsiness can be induced (even without sleep deprivation). Besides, knowledge is required as to whether manipulation of automation duration or a state-dependent analysis of drowsiness effects is suitable. As subjective assessments of drowsiness get less reliable when extreme levels of drowsiness are experienced (E. A. Schmidt et al., 2009) and can have a reactivating effect (Brown, 1994), expert ratings of drowsiness are an appropriate method for the above-mentioned research goals.

In this study, drowsiness was assessed by two raters (sitting in the back seat). The video images were displayed in real time on a screen (see figure 5.2 (b)). This made it possible for the two raters to observe and rate participants' DL. An application was developed for this assessment, which requested raters to evaluate a one-minute interval of participants' drowsiness levels every two minutes. This procedure was chosen because it made it possible to observe the development of drowsiness as a function of time. Preliminary tests showed that even during one-minute intervals higher drowsiness levels could be reached. This is in line with the findings of Anund et al. (2013), who observed that sudden changes in driver drowsiness could not be detected by expert ratings when five-minute intervals were used. However, as this procedure should allow triggering requests to intervene depending on specific drowsiness levels, investigators could communicate (by nodding) about whether to trigger the request to intervene or not. Thus, an analysis of the inter-rater reliability cannot be performed afterward. This limitation was accepted, as the estimation of specified drowsiness levels is crucial for investigating their influence on take-over time metrics.

[2] In Appendix A the German version of the used drowsiness scale is available.

Table 5.1: Levels of drowsiness and indicators

Drowsiness level (DL)	Indicators
1 - *not drowsy*	appearance of alertness present; normal facial tone; normal fast eye blinks; short ordinary glances; occasional body movements / gestures; (Wierwille & Ellsworth, 1994)
2 - *slightly drowsy*	still sufficiently alert; less sharp / alert looks; longer glances; slower eye blinks; first mannerisms as: rubbing face / eyes, scratching, facial contortions, moving restlessly in the seat; (Wierwille & Ellsworth, 1994) and (Wiegand et al., 2009)
3 - *moderately drowsy*	mannerisms; slower eyelid closures; decreasing facial tone; glassy eyes; staring at fixed position (Wierwille & Ellsworth, 1994)
4 - *drowsy*	eyelid closures (1-2s); eyes rolling sideways; rarer blinks; no proper focused eyes; decreased facial tone; lack of apparent activity; large isolated or punctuating movements; (Karrer-Gauß, 2012)
5 - *very drowsy*	eyelid closures (2-3s); eyes rolling upward /sideways; no proper focused eyes; decreased facial tone; lack of apparent activity; large isolated or punctuating movements; (Wierwille & Ellsworth, 1994)
6 - *extremely drowsy*	eyelid closures (4s or more); falling asleep; longer periods of lack of activity; movements when transition in and out of dozing; (Wierwille & Ellsworth, 1994)

Subjective assessment of different strategies

Table 5.2: Selection of options of a driver-state related strategy (DSRS)

Options	*Description*
DSRS-O1:	The vehicle opens the window slightly in order to allow fresh air into the vehicle.
DSRS-O2:	The vehicle emits a scent to stimulate you.
DSRS-O3:	The vehicle increases the volume of the radio.
DSRS-O4:	The vehicle moves the seat into an upright position.
DSRS-O5:	The vehicle adjusts the interior lighting.
DSRS-O6:	The vehicle offers a specific selection of non-driving related tasks (for example a quiz) during the automated drive.

Table 5.3: Selection of options of a system-based strategy (SBS)

Options	*Description*
SBS-O1:	The vehicle ceases to change lanes and drives on the right lane so that the vehicle can come to a safe stop on the hard shoulder should you fall asleep.
SBS-O2:	The vehicle hands the driving task back to you. After that, the system will no longer be available. You take full responsibility for the subsequent drive without the system.
SBS-O3:	The vehicle drives to the next rest area. The system will be available again after a break, depending on your level of drowsiness.
SBS-O4:	The vehicle reduces the maximum speed to give you more time to take control in case of a request to intervene.
SBS-O5:	The vehicle drives without any adjustment. When it recognizes that you have fallen asleep, it brakes, coming to a stop on the hard shoulder.
SBS-O6:	The vehicle drives without any adjustment. When it recognizes that you have fallen asleep, it brakes, coming to a stop on the lane you are in.

This study aimed to gain initial insights into likely accepted options of a driver-state or of a system-based strategy from a user's perspective. Table 5.2 and table 5.3 display these options. Participants were asked to declare the option they would accept most and believe to be the most effective. The latter was only assessed for the DSRS. Besides this, they were asked to evaluate whether they support or reject the different options on a 5-point Likert scale. Options of a preparation strategy were not evaluated in this study as these represent a combination of the driver state-related and system-based strategy.

Experimental design

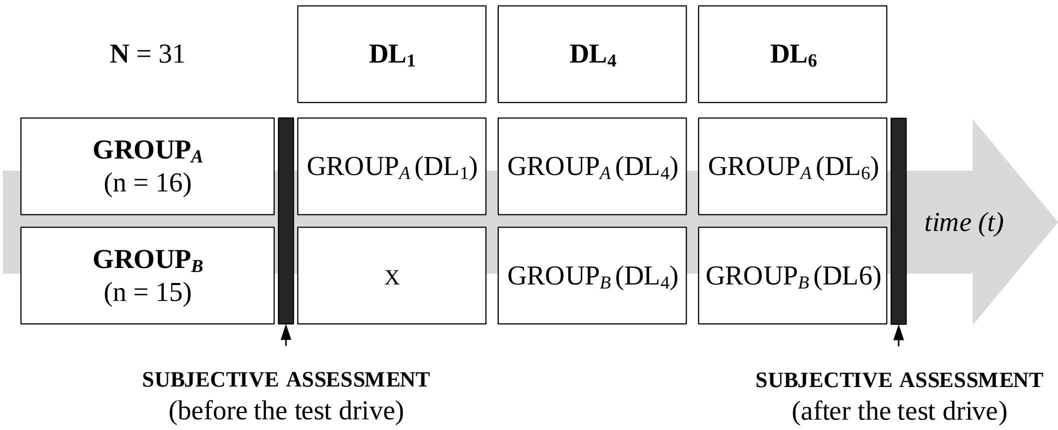


Figure 5.3: Experimental design

Figure 5.3 presents the experimental design used to investigate the above-described research questions. This design was chosen as it allows assessing the time until participants reach DL_4 or DL_6 for the first time and to determine the effect of drowsiness on take-over-time aspects. Participants subjectively assessed the different driver state-related and system-based options before and after the test drive.

Procedure and drowsiness induction

As this study aimed to induce drowsiness, several measures were applied to induce this driver state. However, some useful but extremely laborious measures, such as sleep deprivation, were not used in this study to investigate whether this setting could induce drowsiness without sleep deprivation during normal working hours. If the drowsiness induction is possible even without these measures, this will represent a significant methodical finding. It would allow for a greater sample size and a reduction in the time and cost typically associated with sleep-deprivation studies.

The trials were carried out starting at 9 a.m. or 1 p.m. One test could last a maximum of three hours. Participants were asked not to drink any caffeinated beverages five hours before the study started and to ensure that they would not become hungry during the experiment. Also, participants were informed that no passenger and steering wheel airbag is available in the test-vehicle due to vehicle modification. While participating in the study participants were asked to turn off their mobile phone. They were further asked not to drink or eat anything during the study. Participants were told that they could cancel the study at any time. Also, participants were informed about the drive wizard and about how the simulated motorway pilot can be handled. After activating the motorway pilot, participants should try to relax as much as possible. For this, gentle, relaxing music (volume could not be adjusted) was played. Further, participants were informed that they should avoid talking to the instructors and refrain from closing their eyes and falling asleep during the test drive. Then the take-over scenario, the request to intervene and a subsequent visual-motor-cognitive task were explained and exercised by the participants for the first time. Also, it was explained that a steering impulse to the left or right is required for handling the take-over situation successfully in this experiment. On the way to the motorway entrance, participants exercised the take-over scenario (to prevent training effects, steering to the right was required) and the visual-motor-cognitive task one more time. When the motorway entrance was reached, participants were asked to draw the curtain and to activate the motorway pilot as soon as the status changed to available. When participants reached their group-specific drowsiness levels (see figure 5.3) the take-over situations accompanied by the request to intervene were triggered by the investigators.

5.3 Results

Statistical analysis

The data were analyzed at a significance level of .05 percent. In the case of multiple comparisons, a Bonferroni correction was performed. Permanently narrowed eyes made an assessment of the drowsiness level of one participant impossible. Hence, this data set was excluded from the analysis. One further participant missed assessing the system-based strategy. The influence of driver drowsiness on take-over time metrics was analyzed in a within-subjects comparison of the three different drowsiness levels of $Group_A$ (DL_1, DL_4, and DL_6). Consequently, participants, who had not reached DL_6 had to be excluded from the analysis of take-over time metrics. To control for training effects, $Group_B$ served as a control group. Hence, take-over-time metrics of $Group_A(DL_4)$ and $Group_B(DL_4)$ were compared.

Drowsiness induction

Table 5.4: Time until participants initially reached DL_4/DL_6 (N = 30)

time (minutes)	DL_4 (cumulative percentage)	DL_6 (cumulative percentage)
0	0.00 %	0.00 %
5	3.33 %	0.00 %
10	10.00 %	0.00 %
15	20.00 %	0.00 %
20	23.33 %	3.33 %
25	30.00 %	10.00 %
30	46.67 %	16.67 %
45	60.00 %	40.00 %
60	73.33 %	56.67 %
75	76.67 %	60.00 %
>75	76.67 %	63.33 %
never reached:	23.33 %	36.67 %

Table 5.4 shows the cumulative percentage when participants initially reached DL_4 or DL_6, regardless of the study group. Overall, 76.67 percent of the sample reached DL_4 and 63.33 percent experienced DL_6. On average, participants reached DL_4 after 30.49 (SD = 18.71) minutes and DL_6 after 42.04 (SD = 15.44) minutes for the first time. The number of participants who reached DL_6 in the afternoon (n = 8) was greater than in mid-morning (n = 3).

Influence of different drowsiness levels on take-over time

The effects of different drowsiness levels on take-over time metrics (hands-on and driver intervention time) was analyzed. The driver-intervention time was defined here as the time from the start of the request to intervene until when the stop position of the steering wheel was reached. This differs from known experiments because of the vehicle modification, which does not allow intervention in the real driving process.

Overall, twelve data sets were available for the hands-on and driver-intervention time in $Group_A(DL_4)$. Ten data sets were available for the same metrics in $Group_B(DL_4)$. Take-over times of $Group_A(DL_4)$ and $Group_B(DL_4)$ were normally distributed (Shapiro-Wilks-test: hands-on time $Group_A(DL_4)$ $p = .071$, hands-on time $Group_B(DL_4)$ $p = .118$, driver-intervention time $Group_A(DL_4)$ $p = .557$ and driver-intervention time $Group_B(DL_4)$ $p = .820$). The assumption of variance homogeneity was not violated (Levene's test p > .05). No training effects were found for the hands-on and the driver-intervention times as assessed by a t-test for independent samples (hands-on time: t(20) = -0.286, $p = .778$, and driver-intervention time: t(20) = -0.471, $p =.643$).

In one take-over situation, one participant put both hands on the steering wheel and asked whether steering is required now. Thus, only the hands-on time of this participant could be assessed. Also, only the take-over times of participants who had experienced all specified drowsiness levels (DL_1, DL_4, and DL_6) could be analyzed. In total, nine participants of $Group_A$ experienced DL_1, DL_4, and DL_6. Hands-on times for $Group_A(DL_4)$ were not normally distributed (Shapiro Wilk's test: $p = .021$). Thus, the influence of drowsiness on the hands-on time was analyzed by a Friedman test. Drowsiness did not significantly influence the hands-on time ($\chi^2(2) = 2.00$, $p = .368$). On average, the hands-on time was 1.74s ($SD = 0.27$) when participants were not drowsy and became slightly shorter when participants were drowsy (M = 1.49s, $SD = 0.29$) or extremely drowsy (M = 1.49s, $SD = 0.35$).

Results for the driver-intervention time were similar (n = 8). Driver-intervention times were normally distributed, as assessed by Shapiro-Wilk's test (DL_1: $p = .252$, DL_4: $p = .343$, and DL_6: $p = .942$). The influence of drowsiness on the driver-intervention time was analyzed by an ANOVA with repeated measures (sphericity was not violated, as assessed by a Mauchly's test: $\chi^2(2) = 0.749$, $p = .688$). The mean driver-intervention time of $Group_A(DL_1)$ was 2.09s ($SD = 0.48$). Driver-intervention times became slightly faster with an increase in drowsiness. On average, participants of $Group_A(DL_4)$ completed the take-over situation after M = 1.74s ($SD = 0.70$) and after M = 1.72s ($SD = 0.52$) when they experienced DL_6. Drowsiness also did not significantly influence driver-intervention time in this case ($F(2,14) = 2.64$, $p = .107$).

Figure 5.4 provides an overview of these results.

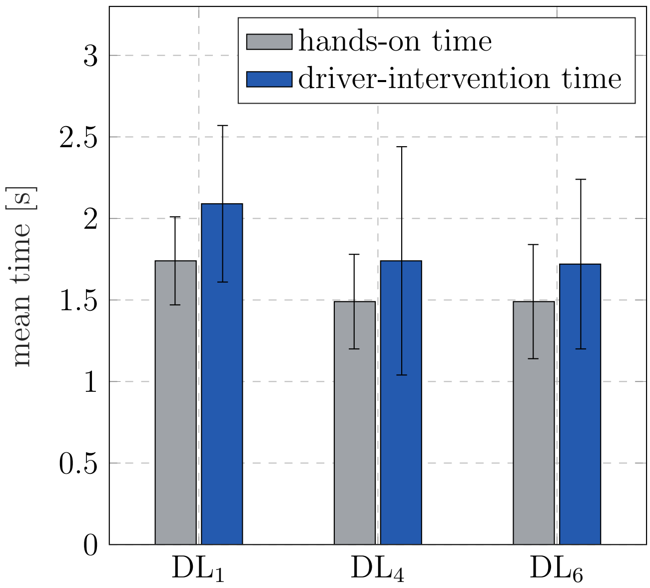


Figure 5.4: Hands-on time (n = 9) and driver-intervention time (n = 8) of $Group_A$ depending on the drowsiness level (DL_1, DL_4, and DL_6) (M ± 1*SD*).

Apart from the quantitative analysis of the influence of drowsiness, it was observed that some participants gave a startled sound when higher drowsiness levels were experienced. In the non-drowsy condition, such reactions were not observed.

Wizard-of Oz - Immersion of automated driving

It was assessed whether the participants could get a feeling of automated driving ("I had felt to be driven automated") on a scale ranging from 1 (no impression of automated driving) to 10 (strong impression of automated driving). On average, the rating was 6.89 (*SD* = 1.52).

Subjective assessments of different strategies

This section reports the subjectively most and least accepted options of the driver state-related and system-based strategies after the test-drive. No significant difference was found between pre- and post-test assessments (for further information and results see Weinbeer et al. (2018)).

Driver-state-related options: When asked which type of option would be the most accepted, DSRS-O6 (targeted offer of non-driving related tasks) was seen to be most popular with 30 percent mentions after the test drive. Participants also assessed DSRS-O6 as the most effective (most reactivating) option with 40 percent mentions after the test drive. Least accepted was DSRS-O6 (vehicle emits a scent) with 6.7 percent mentions.

System-based options: SBS-O1 (no further lane changes and a move to the slow lane) was most widely accepted with 31.9 percent mentions after the test drive. Support for SBS-O4 (reduction in maximum speed) and SBS-O3 (rest area) was the same at 24.1 percent. The options SBS-O5 (vehicle comes to a stop on the hard shoulder if the driver falls asleep) and SBS-O6 (vehicle comes to a stop on the current lane if the driver falls asleep) were rejected by the majority of participants.

5.4 Discussion

It was found that it is possible to induce higher levels of drowsiness without sleep deprivation by controlling several factors (e.g., caffeine and atmosphere). Overall, 76.67 percent of the sample reached DL_4 and 63.33 percent of the sample reached DL_6 without sleep deprivation during regular working hours. Further, it became apparent that drowsiness develops very individually and that higher drowsiness levels can be reached after a very short period. In this study, one participant reached DL_4 within the first five minutes of the test drive, whereas others (23.33 percent of the sample) never reached this drowsiness level.

Thus, it can be assumed that a state-dependent evaluation of drowsiness effects and of different strategies for managing driver drowsiness is required. This is supported by recent studies that also used a state-dependent approach (Vogelpohl, Kühn, Hummel, & Vollrath, 2018; Feldhütter, Kroll, & Bengler, 2018).

Further, the cumulative percentage until participants reached DL_4 or DL_6 for the first time can be helpful for future studies, as it allows an a priori estimation of the time needed to bring a certain number of participants into higher levels of drowsiness. Also, it shows that a test drive should ideally not exceed one hour. A further increase in automation duration only slightly increased the number of participants who reached DL_4 or DL_6 for the first time (see table 5.4).

The different drowsiness levels did not significantly influence ToT metrics in this study, which is in line with the findings of Gonçalves et al. (2016); Vogelpohl et al. (2018) and Feldhütter et al. (2018). In this experiment, the assessed ToT metrics decreased slightly but not significantly. Additionally, it was observed that some participants were startled by a request to intervene. This supports the considerations of Knipling and Wierwille (1994), who emphasized the relevance of appropriate feedback to avoid startled responses. While this consideration referred to lower automation levels, it shows that it remains relevant to proper design L3 ADSs requests to intervene. As the lateral acceleration (Gonçalves et al., 2016) and the longitudinal acceleration (Feldhütter et al., 2018) were significantly influenced by driver drowsiness, it appears that some drowsy participants rather tend to overreact when take-over situations occur.

To ensure that transitions are executed in a controlled manner, adjusting the request to intervene might help deal with drowsiness (see section 3.3.3). Further, it is possible that state-of-the-art requests to intervene might lead to rapid but also startled reactions in some cases. Thus, future research should address this critical driver reaction and also the idea of a preparation strategy. Humans react faster with increasing expertise (Rasmussen, 1983). Therefore, such effects might explain the decrease in ToT metrics. This study introduced Group_B, which aims to control such training effects. However, such effects can still not be completely ruled out due to the study design and the rather small sample size.

When asked which of the driver state-related options (for an overview see table 5.2) would be accepted most, DSRS-O6 (targeted offer of non-driving-related tasks) was mentioned most frequently (30.9 percent). Further, DSRS-O6 was believed to be the most effective option by 40 percent of the sample. However, further research is needed to investigate various NDRTs and their effectiveness to reduce driver drowsiness. SBS-O1 (no further lane changes and a move to the slow lane) was the most supported system-based option (mentioned by 31 percent), followed by SBS-O4 (speed reduction) (24.1 percent). SBS-O3 (rest area and break) was also mentioned by 24.1 percent of the sample. Further, SBS-O5 and SBS-O6 were rejected by the majority. Thus, it can be considered that higher levels of escalation should be avoided from the users' perspective. This in line with the considerations of van der Laan et al. (1997) who pointed out that restrictive systems *"are likely to be less accepted than nonrestricting"*.

It must be taken into account that the evaluation may be dependent on the point of view. For instance, the perspective of other road users might differ from the users' perspective. If this is the case, system developers face a dilemma, as they must develop systems that are safe and accepted not only by users but also by other road users. Thus, a holistic view is needed for developing safe and accepted systems from different perspectives.

Conclusion

On average, participants could get a feeling of automated driving in the modified right-hand-drive vehicle. Further, it was found that drowsiness develops very individually. Thus, it is concluded that a state-dependent evaluation is required to investigate the influence of drowsiness on take-over performance. 76.67 percent of the sample reached DL_4 and 63.33 percent reached the highest DL_6. This shows that it is possible to induce drowsiness without sleep deprivation in normal working hours by controlling several factors (e.g., caffeine). Further, requests to intervene were triggered depending on specific drowsiness levels. No statistically significant differences in take-over time metrics were found between the different drowsiness levels. However, some participants were startled when a request to intervene happened in the drowsy but not in the non-drowsy condition. Subjective assessments revealed that higher escalations, such as a minimal risk maneuver, should be avoided from the users' perspective. To sum up, this study provides a methodological basis for evaluating of drowsiness effects and different strategies for managing driver drowsiness during an L3 automated drive.

CHAPTER 6

The potential of non-driving-related tasks to manage drowsiness[1]

Abstract: This study investigated the reactivation potential of NDRTs during a simulated automated drive. A total of 71 participants took part in this experiment. After a relaxation phase, the sample was divided into three groups that were given different NDRTs (a dictation, a sports activity, and a relaxation task). In this study, a rating greater than seven on the Karolinska-Sleepiness Scale (KSS) was considered the system limit driver drowsiness. It was found that the targeted use of NDRTs has potential as a suitable option for managing driver drowsiness. No participant of the dictation or sports activity group exceeded level 7 on the KSS after the reactivation phase. Even after the effectiveness phase, there was still a major difference between the number of participants who exceeded level 7 between the dictation and sports activity groups compared to the relaxation group.

6.1 Introduction

In the "driver availability" concept, the current driver state is determined by a driver's arousal level, by the type of currently conducted NDRTs, and by moti-

[1] The following chapter mainly bases on the publication of Weinbeer, Muhr, and Bengler (2019)

vational aspects. If a driver needs to take over control from an ADS, a target driver state must be achieved within a given time budget (Marberger et al., 2018). Various studies evaluated the influence of standardized and of naturalistic NDRTs (for an overview see Naujoks, Befelein, Wiedemann, and Neukum (2018)). However, the type of NDRTs had either none (e.g., Radlmayr et al., 2014) or only a minor impact on take-over performance (Gold, Happee, & Bengler, 2018). Hence, NDRTs might be a suitable approach to avoid or at least to postpone the system limit driver drowsiness during an automated drive. So far, few studies exist that investigated the influence of NDRTs on drowsiness development (or fatigue). One study found that the use of NDRTs can reduce driver fatigue when an L2 ADS was used (Neubauer, Matthews, & Saxby, 2014). Also, studies showed that participants' drowsiness developed more slowly when they executed a NDRT compared to being inactive (Schömig et al., 2015), or when participants performed a motivational compared to a tiring NDRT (Jarosch et al., 2017). In addition, it is known that measures against driver drowsiness were intensively studied for manual driving (for an overview see Hashemi Nazari, Moradi, and Rahmani (2017)). However, the negative influence of distraction, for example, due to cell phone use, has been demonstrated in the context of manual driving (e.g., Strayer, Drews, & Crouch, 2006). As described in section 3.3.1, drivers might be able to perform very different types of NDRTs during an automated drive, even for a more extended period. Based on the present state of research, this study aims to investigate the reactivation potential and effectiveness of different NDRTs to avoid or to postpone the system limit driver drowsiness.

6.2 Method

Participants

Seventy-one employees of the AUDI AG took part in this experiment. The sample consisted of 24 women and 47 men. On average, participants were 31.90 ($SD = 8$) years old and have held their driver's licenses for 14.03 ($SD = 8$) years. Participants were asked to register for this study only if they are usually able to read as a passenger without feeling sick. Further, participants were asked to abstain from all caffeinated beverages for one hour before the experiment.

Test vehicle and test route

A Wizard-of-Oz approach was used; in this case, a modified right-hand-drive vehicle (Audi A4 sedan). This test vehicle was equipped with additional driving school mirrors so that the participants were also able to observe the surrounding traffic. Further, the vehicle was equipped with a 6-inch tablet, which was attached in front of the passenger seat (see figure 6.1). This tablet showed the pilot status (pilot active), speed in km/h and indicators. A 12-inch tablet was integrated into the center console. On this tablet, the applications of the various NDRTs were presented.

(a) Interior view (b) Exterior view

Figure 6.1: Wizard-of-Oz approach: Test vehicle

These applications guided the participants through the entire experiment. As long as an L3 ADS was simulated, a curtain was placed between the participant and the investigator. Similar to Study 1, the investigator simulated the system behavior of a possible future motorway pilot. The maximum speed was 130 km/h and lane changes were performed conservatively. The assistant systems ACC and LKA were not used in this study, as this would represent a state-of-the-art system rather than a future motorway pilot.

The study was conducted on the A9 autobahn in Germany. The experiment started at the motorway service station Koeschinger-Forst (see figure 6.2).

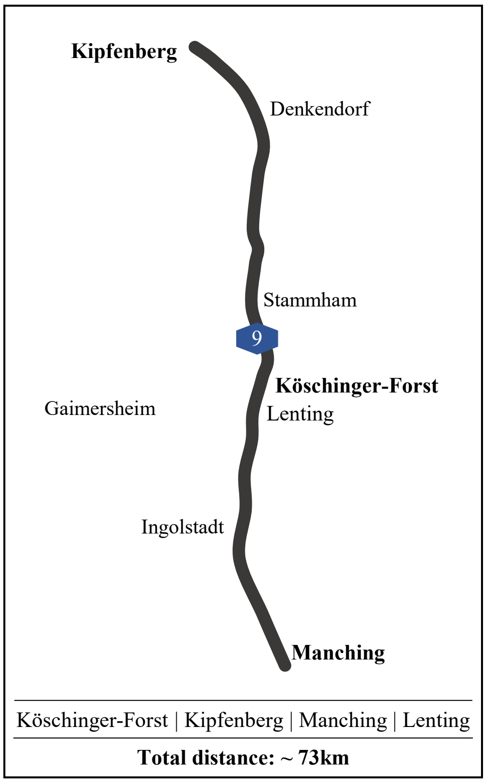


Figure 6.2: Study 2: Test route

Experimental design

The study consists of three parts. Part A used a within-subject design to assess the effectiveness of drowsiness generation. Part B consisted of a reactivation and an effectiveness phase. These two phases allow an investigation of the reactivation potential and effectiveness (even after the actual reactivation) of NDRTs. In Part C, a follow-up survey was conducted regarding the experience when executing the NDRTs of Part B. As a dependent variable, the KSS was used (Åkerstedt & Gillberg, 1990).

Figure 6.3 shows the experimental design and the timing of questioning (KSS_1, KSS_2, KSS_3, KSS_4, and KSS_5).

Further, to check whether the tasks were perceived differently, participants assessed the "In-game GEQ" (Game Experience Questionnaire (GEQ)) at the end

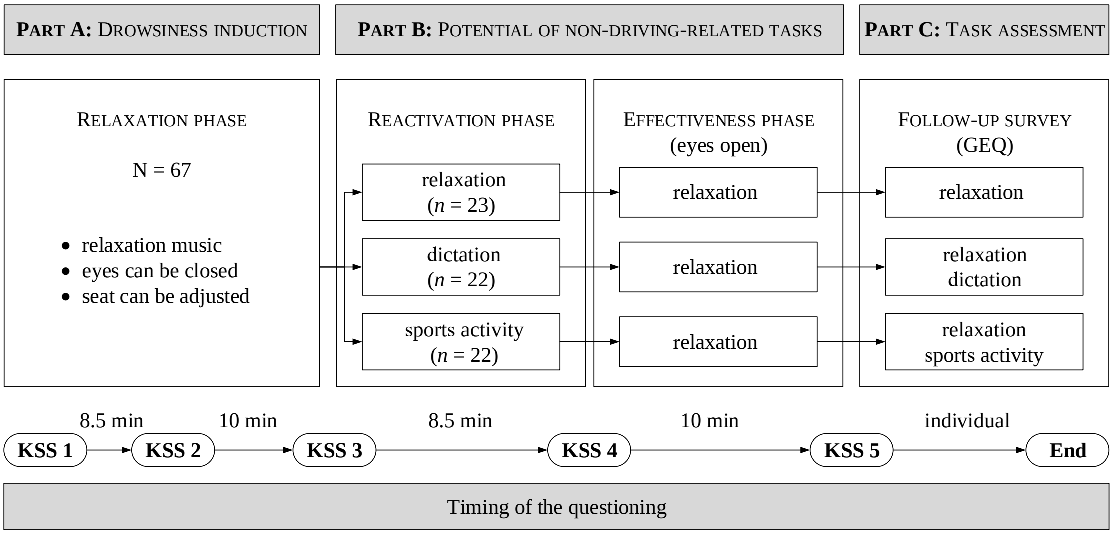


Figure 6.3: Experimental design and timing of the drowsiness assessment

of this study. Based on 14 items, seven components can be calculated (IJsselsteijn, de Kort, Y. A. W., & Poels, 2013). These components are competence, sensory and imaginative immersion (abbreviated as immersion in this thesis), flow, tension, challenge, negative affect, and positive affect. The German version of the GEQ was used in this study (see Engl (n.d.)). However, the items “I was interested in the game’s story” and “I had to put a lot of effort into it” were adjusted for this study to “I was interested in the content of the task” and “I had to put a lot of effort into the task”.

Procedure and non-driving-related tasks

Participants were randomly assigned to one of the three different groups. Four experimental sessions were conducted per day. These started at 8:00 a.m., 10:15 a.m., 1:00 p.m. and 3:15 p.m. The start times were permuted among groups to distribute the effect of circadian rhythm. Each trial was scheduled to last for a maximum of two hours. Participants were informed that a motorway pilot will be simulated and that the investigator will drive the vehicle all the time. During part A, which was identical for all groups, relaxation music was played. Further, participants were informed that they should adjust the volume and/or seat position during the relaxation phases in such a way that they could relax as much as possible. During part A, participants were also allowed to close their eyes. Also, they were informed that, if possible, they should avoid falling asleep during the entire test drive.

Three applications were developed to provide the different NDRTs and to ask participants to rate their current drowsiness level at certain points (see the timing of the questioning in figure 6.3). In part B, the participants first experienced their group-specific task. These are presented in the following.

- *Relaxation:* In this case, the relaxation group can be considered as a control group. Participants of this group were asked to continue relaxing. However, from that moment, they were asked to keep their eyes open.

- *Dictation:* Different studies showed that a large number of users would use the driving time to conduct tasks, such as "texting" (e.g., Pfleging et al., 2016). Hence, it was decided to use dictation as a non-driving-related task, as this requires typing different words for a limited period.

- *Sports activity:* Further, a sports activity (Handytrim fitness device) was tested in this experiment for two reasons. First, it is a task that cannot be executed during a manual drive. Second, using the travel time to improve physical fitness could increase users and the societal benefit generated by ADSs.

In the subsequent effectiveness phase, all participants were asked to relax while keeping their eyes open. This phase aimed to assess the effectiveness of the reactivation phase.

6.3 Results

The significance level of the statistical analysis was .05. Data from four participants were not recorded due to a system crash during the experiment. Data from 67 participants were analyzed.

Part A: Sleepiness induction

The KSS ratings increased significantly within the three times of measurement, as assessed by a Friedman test ($\chi^2 = 49.22$, $p < .001$). At the beginning of the test drive (at KSS_1), the mean KSS rating was 4.48 ($SD = 1.59$), further increased to 5.54 ($SD = 1.41$) at KSS_2 and reached 6.15 ($SD = 1.63$) at the end of *part A* (KSS_3).

Part B: Distribution functions of the KSS ratings after the reactivation and effectiveness phase

KSS scores greater than seven can be allocated to the driver state drowsiness (Johns, 2009). Besides this, Ingre et al. (2006) found that the probability of incidents and accidents (for lower automation levels) increased dramatically when KSS ratings were greater than 7. For the following analysis, it is assumed that KSS ratings greater than level 7 would lead to the system limit driver drowsiness. Thus, the number of participants exceeding a KSS level of 7 is of great importance, as this would represent the number of users who could no longer use the L3 ADS. The reactivation potential is considered high, when the number of participants exceeding a critical drowsiness state, in this case, a KSS rating higher than 7, is small. A reactivation is considered to be effective if the number of participants exceeding a critical drowsiness state remains small even after the actual reactivation (in this case after ten further minutes of relaxation). This represents a kind of worst-case consideration, as users were asked to relax again after the reactivation phase, and thus did not further engage in a reactivating NDRT. Study 1 showed that a state-dependent analysis should be ensured. Therefore, the distribution functions of the KSS ratings for KSS_4 and KSS_5 were determined.

$$KSS\ rating = \begin{cases} x \leq 7 & \text{drowsiness is not considered a system limit} \\ x > 7 & \text{drowsiness is considered a system limit} \end{cases} \tag{6.1}$$

Participants whose KSS rating was greater than 7 at KSS_3 were excluded from the calculation of distribution functions (n=10), as their drowsiness state would already represent the system limit driver drowsiness (see formula 6.1). A total of 57 datasets could be analyzed (relaxation: $n = 18$, dictation: $n = 19$, and sports activity group: $n = 20$). The distribution functions of these groups were calculated at KSS_4 and KSS_5 (see figure 6.4).

After the reactivation phase, no participant of the dictation and sports activity groups exceeded KSS level 7. In contrast, the number of participants who reached level 8 or level 9 on the KSS was 33.34 percent in the relaxation group. At KSS_5 (after the effectiveness phase), 38.89 percent of the relaxation group exceeded level 7 on the KSS.

During the effectiveness phase, the number of KSS ratings greater than 7 increased to 10.52 percent in the dictation group and 15 percent in the sports activity group.

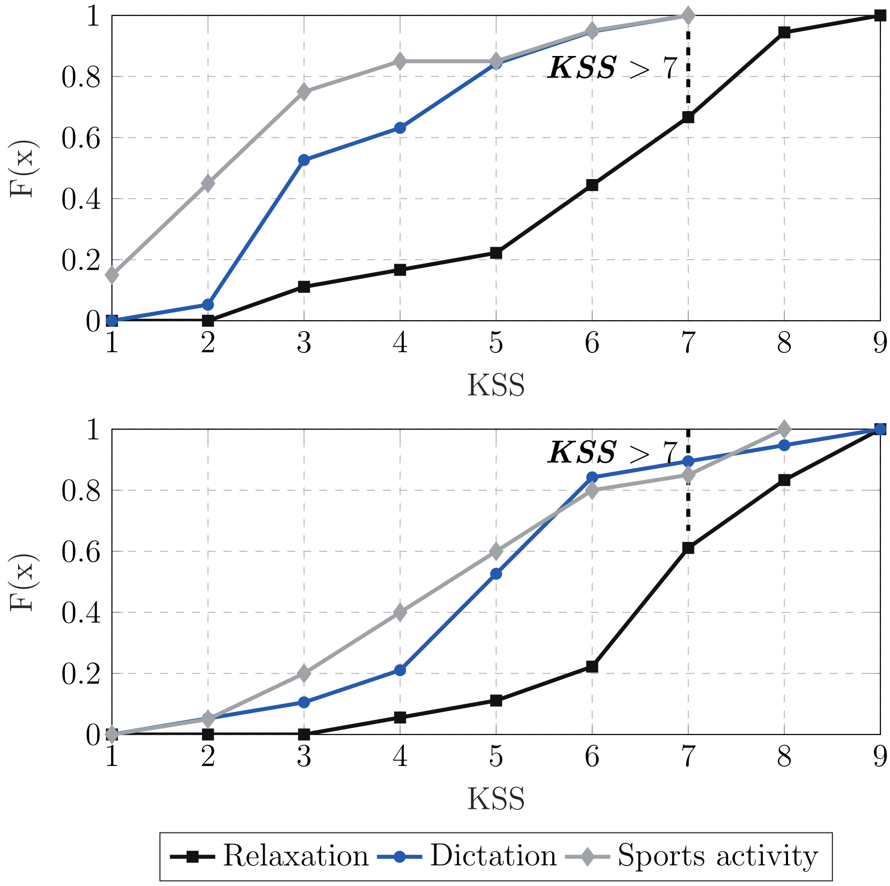


Figure 6.4: The cumulative distribution function of KSS ratings at KSS_4 (top figure) and KSS_5 (bottom figure)

Part C: Assessment of the different NDRTs

The KSS distribution functions clearly show that there is a difference between the sports activity and the dictation groups compared to the relaxation group at KSS_4 and KSS_5. However, the differences between the KSS distributions of the dictation and sports activity groups seem rather small. This raises the question of whether the dictation and sports activity tasks were perceived differently. To check this, participants rated 7 categories according to the In-game GEQ. Figure 6.5 provides an overview of the results.

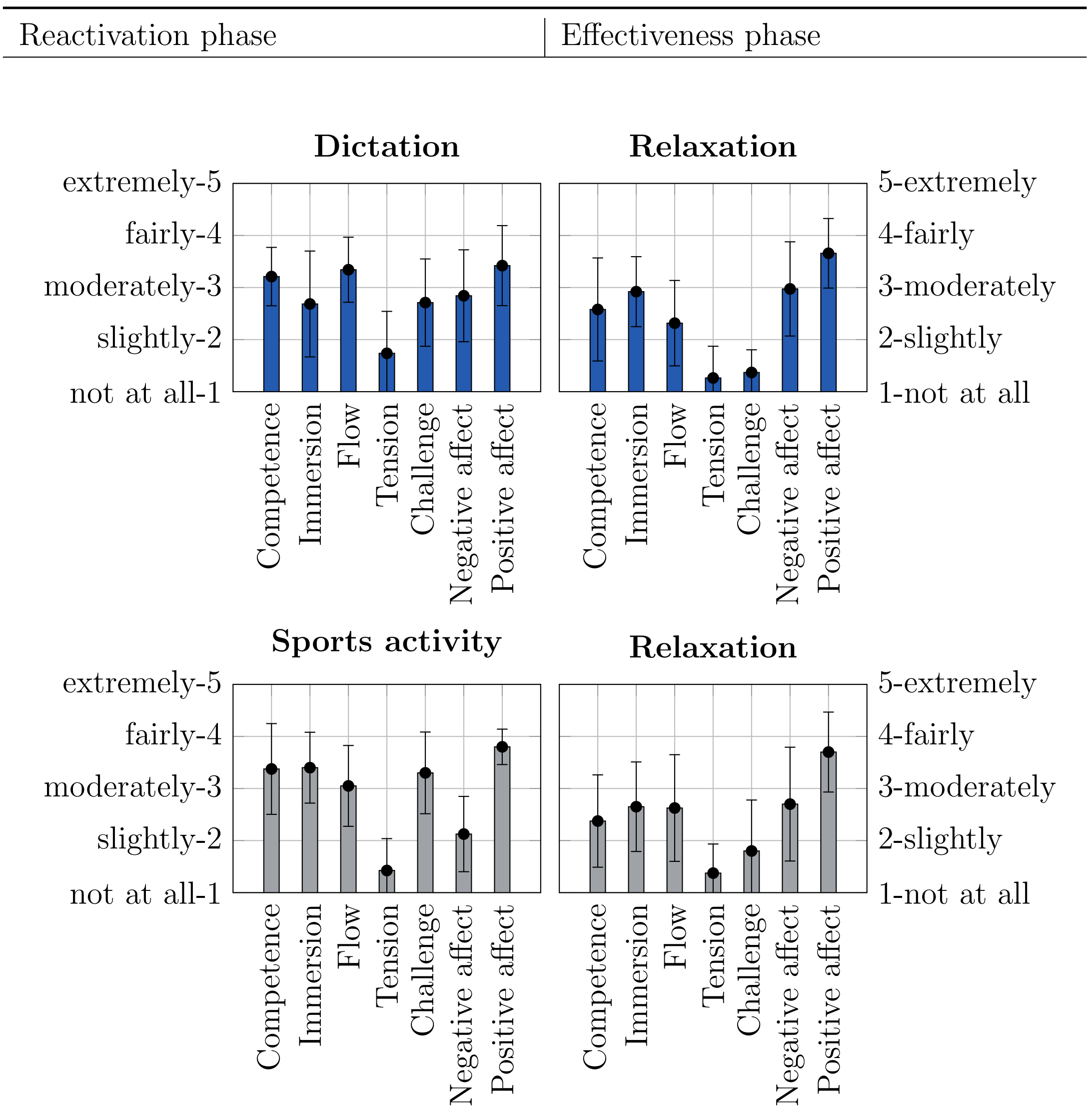


Figure 6.5: In-game GEQ assessment of the different tasks (dictation: n = 19 and sports activity task: n = 20) (M ± SD)

For this analysis, the significance level was adjusted to p = .007 due to the multiple comparisons (Bonferroni correction). Overall, the assessment of the dictation and sports activity tasks did not differ significantly: competence (U = 158.5, z = -0.910, p = .363), immersion (U = 107.5, z = -2.383, p = .017), flow (U = 153.0, z = -1.070, p = .285), tension (U = 144.5, z = -1.357, p = .175), challenge (U = 122.0, z = -2.222, p = .026), negative affect (U = 99.0, z = -2.597, p = .009), and positive affect (U = 134.0, z = -1.670, p = .120). In addition, it was assessed in which components the tasks of the reactivation and effectiveness phase were perceived differently (dictation compared to the relaxation task, and sports activity compared to the relaxation task). The assessment of the dictation task compared to the relaxation task differed significantly in the following categories: flow (z = -3.185, p = .001), and challenge (z = -3.517, p<.001). The sports activity task differed significantly compared to the relaxation task in the following categories: competence (z = -2.975, p = .003), immersion (z = -3.280, p = .001), and challenge (z = -3.517, p < .001), as assessed by a Wilcoxon test.

Wizard-of Oz - Immersion of automated driving

It was assessed whether participants could get a feeling of automated driving (I had felt to be driven automated) on a scale ranging from 1 (no impression of automated driving) to 10 (strong impression of automated driving). On average, the rating was 6.33 (SD = 2.27).

6.4 Discussion

The drowsiness generation by relaxing music within part A of the study can be considered successful. Ten participants already reached a KSS score greater than 7 at KSS_3 (in about 20 minutes). Hence, the drowsiness generation phase should not be longer for this study purpose. KSS scores greater than 7 are considered a system limit in the further analysis (see formula 6.1). Thus, the data of participants who already exceeded this level were not used for the calculation of the KSS distribution functions. Besides this, the driver state drowsiness can be reached rather quickly. This finding is in line with other studies that also showed that drowsiness occurs quickly during an automated drive (Weinbeer et al., 2017; Vogelpohl et al., 2018; Feldhütter et al., 2018). Further, this study proved the reactivation potential of NDRTs. No participant of the dictation and sports activity groups exceeded KSS level 7 at the end of the reactivation phase. In addition, after the subsequent effectiveness phase, the number of participants who exceeded a KSS level of 7 was considerably smaller (with 10.52 percent when participants did the dictation and 15.00 percent when participants did the sports activity task before the relaxation task) compared to the group who had to relax during the entire study (38.89 percent). Hence, the results of this study support the observations made by Schömig et al. (2015) and Jarosch et al. (2017). As there were only small differences between the dictation and sports activity groups regarding the reactivation potential and the effectiveness, it can be concluded that in this study the fact that the participants performed a reactivating NDRT was more important than the type of task. Also, it became clear that the dictation and sports activity tasks were not perceived differently, as the in-game GEQ assessments of these two tasks did not differ significantly. For the dictation and the sports activity groups, there was a significant difference in the experience of "challenge" when comparing the assessment of the subsequent relaxation task with the task of the reactivation phase. This might indicate that an increase in drowsiness is likely to occur when a task is perceived as not challenging. Further, the fact that the category "challenge" has not been exhausted in the dictation and sports activity groups (see figure 6.5) might indicate that other or freely chosen tasks can even be more reactivating. However, as about 80 percent of the dictation group and the sports activity group reached a KSS score lower than 5 at KSS_4, other tasks may only lead to a marginally greater decrease in drowsiness. However, a possible positive effect of other tasks might last longer even after the actual reactivation phase. Further studies might address this issue. Although the reactivation potential was demonstrated, it needs to be considered that the use of measures against sleepiness is no longer possible at very high drowsiness states (see section 3.3.1). Further research is required to identify whether a KSS score greater than 7 is a suitable system limit during an automated drive.

Conclusion

This study found that NDRTs have the potential to be a suitable option for managing driver drowsiness. This is because no participant of the dictation or sports activity group exceeded level 7 on the KSS after the reactivation phase. Even after the effectiveness phase, there was still a major difference between the number of participants exceeding level 7 between the dictation and sports activity groups compared to the relaxation group. Future studies might also evaluate the potential of other naturalistic NDRTs and especially their link to the reactivation effectiveness.

CHAPTER 7

Dealing with uncertainty periods of a Driver Monitoring System

Abstract: According to the Drowsiness Management Concept (DMC), identified strategies are executed depending on a user's drowsiness level. Hence, a Driver Monitoring System (DMS) failure could lead to a false activation of the different strategies, which could lower the user's acceptance of an L3 ADS or even become safety-critical (see section 3.1). Hence, a DMC should be designed to be tolerant of uncertainty periods of a DMS. Regardless of the type of system adaption, the question must be answered as to whether users should be informed about a DMS uncertainty and an accompanying system adaption. Hence, this study investigated how various approaches to dealing with a DMS uncertainty period impact user acceptance and trust in automation. Four different concepts were identified. The influence of these on user acceptance and trust in automation was evaluated by a counterbalanced within-subjects design (N = 24). Offering a compensation task was found to have the potential to be an accepted and trusted approach to deal with DMS-uncertainty periods. It also became apparent that users would like to be informed about the cause of a system adaption, emphasizing their need and desire for feedback and transparency ADSs.

7.1 Introduction

The interaction between an ADS and a DMS

For automation levels lower than level 3, drivers decide whether or not to trust a DMS feedback (e.g., break recommendation) and whether or not to adjust their behavior subsequently. In case of an inaccurate detection, resulting in an incorrect DMS feedback, drivers can ignore this information and keep on driving while being fully responsible for vehicle safety. However, an increase in automation brings with it a paradigm shift. The ADS in collaboration with the DMS (see section 3.1) are responsible for ensuring vehicle safety. Therefore, in addition to managing a critical driver state such as drowsiness, DMS-uncertainty periods must be managed. So far it is unclear how DMS-uncertainty periods should be managed and to what extent different solutions influence users' automation trust and acceptance.

From a technical perspective, managing the uncertainty periods of a DMS is a challenge as the performance of a DMS needs to be assessed to trigger different approaches. The funded EU project RobustSENSE describes an architecture that addresses system performance assessment in the context of automated driving. In this concept, the interior camera system, which aims to detect driver distraction and drowsiness, provides information about whether the stereo camera is up and running correctly (Saccagno et al., 2016). Based on such an assessment, different approaches for dealing with DMS uncertainty periods may be triggered. The information on whether a DMS-uncertainty period exists or not is considered prerequisite in this study.

Acceptance, trust, and system information

The introduction of ADSs requires not only a technological improvement of the sensors but also the consideration of human factors such as user acceptance. Consequently, sufficient HMI concepts need to be developed (Bengler et al., 2014). A procedure for assessing users' system acceptance was provided by van der Laan et al. (1997). This procedure allows an assessment of a practical (usefulness score) and pleasant (satisfying score) dimension on a nine-item scale. The terms acceptability and acceptance can be differentiated by the time of the assessment. Acceptability refers to assessments before using a system, whereas acceptance refers to assessments after a system has been used (Verberne, Ham, & Midden, 2012).

Perceived usefulness and ease of use are relevant influencing factors on system acceptance according to the Technology Acceptance Model (TAM) (Davis, Bagozzi, & Warshaw, 1989). The Automation Acceptance Model (AAM) is an advancement of the TAM for automated systems. This model takes into account the construct *trust*, which can influence the *perceived ease of use*, the *perceived usefulness* and the *behavioral intention to use* a system (Ghazizadeh, Lee, & Boyle, 2012). From this relationship, it can be deduced that trust as a mediating factor for acceptance must be considered for a successful implementation of ADSs. Further, to ensure safety, it is particularly important to create a corresponding trust in automation. This means that inadequate calibration, in which users' trust exceeds (overtrust) or is lower (distrust) than a system's capability, should be prevented. It must be taken into account that overtrust can result in misuse and that distrust can result in disuse of an automated system (Lee & See, 2004). Overtrust is considered a negative consequence of the out-of-the-loop performance issues (Kaber & Endsley, 1997). However, Norman (1990) emphasized that:

> *"[...] although the human operators are indeed no longer 'in the loop', the culprit is not automation, it is the lack of continual feedback and interaction."*

Principles such as *"the human operator must be able to monitor the automated system"*, *"automated systems must be predictable"*, and *"the automated system must also be able to monitor the human operator"* for human-centered automation were derived in the field of aviation. Although these principles were developed under the assumption that *"the pilot bears the ultimate responsibility for the safety of any flight operation"* (Billings, 1991), which can be allocated to L2 systems, some principles might still be valid for L3 ADSs. Future systems will be able to perform the driving task under certain conditions; however, they still require a driver to take control (e.g., in case of a construction site) or due to legal requirements (see section 2.1).

Hence, it can be assumed that there will still be a great need for human-centered automation design in the future. This consideration is supported by the guidelines provided by Endsley (2017), who defined automation transparency as follows:

"A high degree of transparency and observability of system behavior and functioning is needed, making it clearly apparent not only what the system is currently doing but also why it is doing it and what it will do next."

Further, it should be noted that interfaces should provide information about a system's performance and its ability to handle future tasks (Endsley, 2017). Also, a transparent information presentation could help to avoid situations that cannot be managed by the user and thus can represent an alternative to unclear adaptive systems (Bengler, 2011).

In summary, it can be concluded that taking into account users' trust and acceptance is crucial for the successful implementation of ADSs. The relevance of the system transparency and thus the clear information presentation for a human-centered automation design became apparent. However, it remains unclear how uncertainty periods of a DMS should be managed during an automated drive.

Dealing with uncertainties - human factors approaches

Two approaches may be suitable for managing DMS-uncertainty periods. These are presented below.

One approach would be forced user action in the case of a DMS uncertainty. This would be similar to the "potential-trigger" approach suggested by Wimmer (2014) or to the alertness-request approach proposed by J. Schmidt, Braunagel, et al. (2016). However, unlike these approaches, inputs would only be required when DMS-uncertainty periods occur and not during the entire drive with an L3 ADS.

Another option could be a system adaption, as an increase in vehicle safety can be achieved by adapting various parameters (e.g., timing of the warning) based on the interaction between an environmental, a vehicle, and a driver model (Trivedi et al., 2007). When translated to automated driving, this means that if a DMS cannot reliably detect the driver or the driver state, an ADS can adapt its behavior to enhance vehicle safety and to be prepared if an MRM must be performed. However, it should be remembered that high levels of escalation, including more pronounced system adaptions, should be avoided from a user's perspective (see chapter 5.3 and Weinbeer et al. (2018)).

It should also be noted that feedback as an alternative to system adaption in complex automated systems has the potential to enhance the human-machine interaction and increase system transparency (Bengler, 2011). The impact of system performance information on users' trust has been the subject of several driving simulator studies, with partially contradictory results. It was found that trust and acceptance of automation increased when an uncertainty face was displayed if a system limit was reached. Furthermore, the generalized symbol was considered useful, particularly in complex situations, as it can avoid displaying too much information (Beller, Heesen, & Vollrath, 2013). A slight but not significant increase in trust in automation and in usability assessments was found when system-performance feedback was provided on a 7-point scale, compared to a baseline condition that did not contain additional ADS performance information (Hergeth, 2016). Another driving simulator study found that participants took over faster when information about system uncertainty was provided, as opposed to participants who did not get any uncertainty feedback (Helldin, Falkman, Riveiro, & Davidsson, 2013). Contrary to the results of Beller et al. (2013), this study found that participants trusted the automated system less when they received uncertainty information compared to a condition with no additional feedback. A possible explanation for this might be a lack of information about the cause of the uncertainty (Helldin et al., 2013).

Concepts that require driver input, such as the "potential-trigger" (Wimmer, 2014), could be a helpful approach for managing DMS-uncertainty periods. It became clear that there are partially contradictory results regarding the need to provide performance information about an ADS. The focus of system-uncertainty feedback still was always limited to the performance of outward-directed sensors of an ADS, although the prediction of these sensors cannot normally be influenced by an adaption of users' behavior or by bypasses such as requested user interactions. In contrast to this, uncertainty periods of a DMS are strongly influenced by a user's behavior, e.g., if the driver turns away, reliable detection of a driver's drowsiness state could be impaired or not possible (see section 2.5). Hence, it is of interest whether the feedback about a DMS-uncertainty period influences driver behavior and how different feedback concepts influence user trust and acceptance of automation. This information is necessary for designing the interaction between an ADS and a DMS with tolerance towards uncertainty periods of a DMS. So far, to the author's knowledge, no study exists that addresses these specific issues. The different identified approaches are summarized in figure 7.1.

Based on the present state of research, this study aims to gain knowledge about which concept should be provided when it comes to a DMS-uncertainty period concerning users' trust and acceptance in automation.

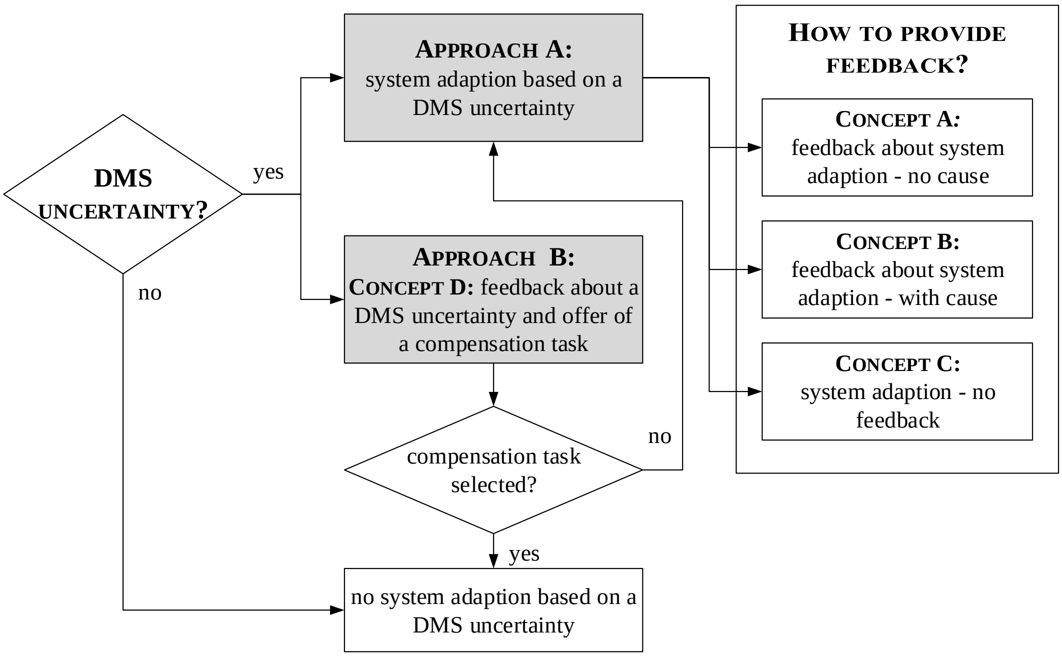


Figure 7.1: Managing DMS-uncertainty periods: approaches and feedback options

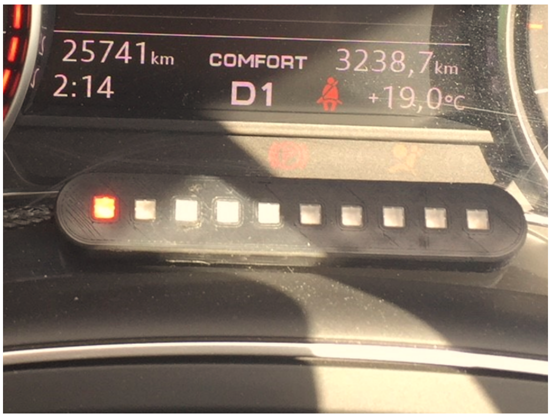
7.2 Method

Participants

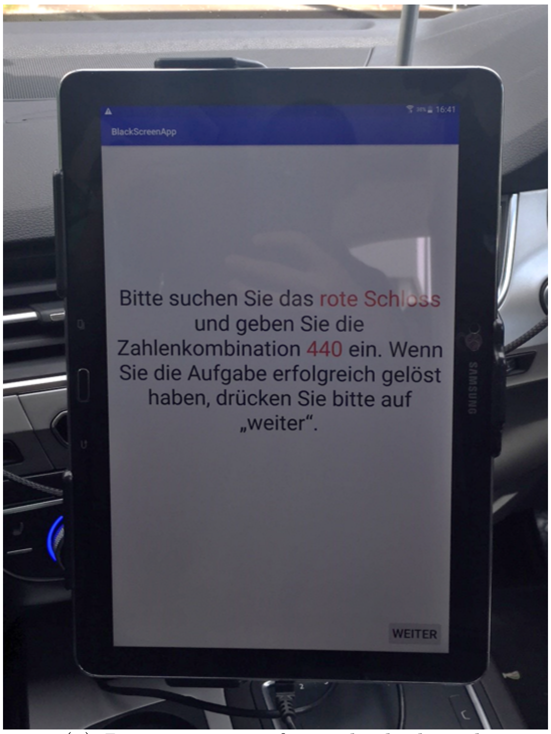
Twenty-four participants took part in this experiment. All participants were employees of AUDI AG. The sample consisted of 6 women and 18 men. On average, participants were 28 years ($SD \pm 8$) old and have held their driving license for 11 ($SD \pm 8$) years.

Test vehicle and test route

(a) Right-hand-drive vehicle with instrument cluster



(b) Display of the system adaption countdown for the drive-wizard



(c) Request to perform the lock-task

(d) Locks installed into the central armrest

Figure 7.2: Modifications of the Audi Q7 for the evaluation of HMI-feedback concepts

The test vehicle of study 1 (Audi Q7) was used again. However, as the study at hand aimed to investigate the influence of different types of feedback on users' trust and acceptance of automation, an instrument cluster was integrated (see figure 7.2 (a)). Locks were installed into the central armrest (see figure 7.2 (d)). An application on the tablet prompted participants to unlock a certain lock at specific times (see figure 7.2 (c)). A countdown was integrated at the original driver's seat to inform the drive wizard about whether a system adaption needs to be simulated (see figure 7.2 (b)). The L3 ADS was simulated twice from the motorway entrance Lenting to Denkendorf and back again, giving a total of four sections (see figure 7.3).

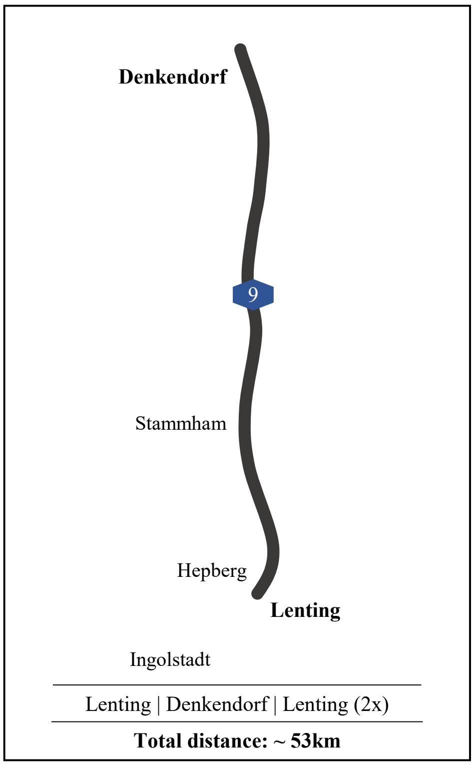


Figure 7.3: Study 3: Test route

Experimental design

A within-subjects design was used to investigate the influence of the different concepts (Concept$_A$, Concept$_B$, Concept$_C$, and Concept$_D$) on users' trust and acceptance of automation. To counteract sequence effects, a complete counterbalanced design was used, leading to 24 different conditions. Participants were randomly assigned to one of these conditions.

Independent variables

In total, based on the two different approaches for managing uncertainty periods of a DMS (see figure 7.1), four concepts were derived and tested in this study.

Concept$_A$ (see figure 7.4), Concept$_B$ (see figure 7.5), Concept$_C$ (see figure 7.6) differ in the type of feedback about a system adaption (in this case, that the vehicle ceases to change lanes).

Concept$_A$ provides feedback on the system adaption but does not explicitly explain the cause of this adaption, representing an implicit feedback option. In contrast, Concept$_B$ contains this information, representing an explicit feedback option. Concept$_C$ does not include any feedback. Concept$_D$ (see figure 7.7) takes approach B (offer of a compensation task) into account, and thus provides feedback about a DMS-uncertainty (similar to Concept$_B$) and offers a compensation task to bypass a system restriction.[1]

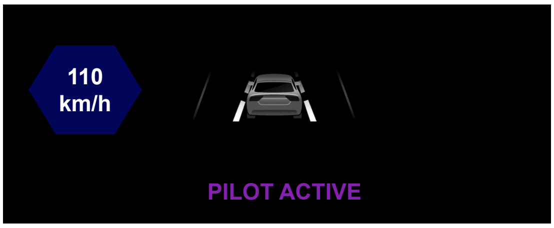
Implicit feedback: Provides feedback about the system adaption but does not explicitly explain the cause of this adaption.

Figure 7.4: *Concept A*

Explicit feedback: Contains information about the cause (driver monitoring is not possible) of the performed system adaption.

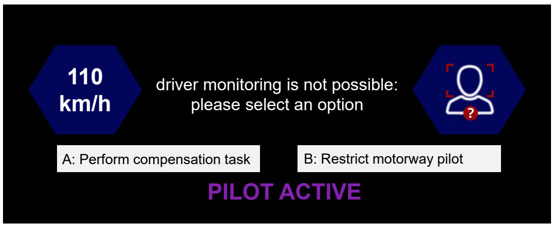
Figure 7.5: *Concept B*

[1] In Appendix B the German versions of the different concepts are available.



No feedback: System adaption is performed, but feedback is not provided

Figure 7.6: *Concept C*



The offer of a compensation task: Users can decide between an alternative task (in this case the SuRT) and a system restriction.

Figure 7.7: *Concept D*

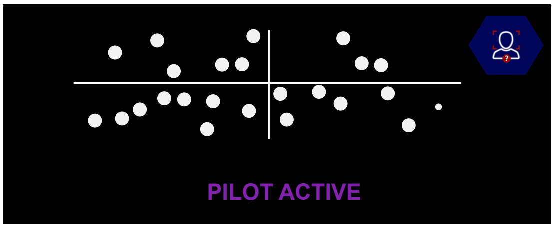


Figure 7.8: *Concept D: Option A*

Figure 7.9: *Concept D: Option B*

Dependent measures

To assess user acceptance of automation, the questionnaire provided by van der Laan et al. (1997) was used in this study. Trust and mistrust of automation were evaluated. The German, validated version (provided by Pöhler, Heine, and Deml (2016)) of the automation trust questionnaire, initially developed by Jian, Bisantz, and Drury (2000), was used. Besides this, participants were asked to rate whether the signal tone is mandatory (when they experienced Concept_A, Concept_B, and Concept_D). The usefulness of the additional acoustic feedback was analyzed by gaze data (change from looking to the locks (central armrest) to looking ahead). Additionally, at the end of the study, participants were asked to rate their preferred concept and to assess different questions concerning their information preferences and the compensation task. For example, participants were asked whether they wish to be permanently informed about an active camera-based DMS. For this, a seven-point Likert scale was used (1 "strongly disagree, 2 "disagree", 3 "slightly disagree", 4 "neither agree nor disagree", 5 "slightly agree", 6 "agree", 7 "strongly agree").

Procedure

Before the test drive, participants were informed that according to the current legal situation drivers must be permanently receptive during an automated drive. It was explained that if a DMS cannot reliably detect the driver state, an L3 ADS could adapt its behavior (e.g., reduce the maximum speed or cease to change lanes) to enhance vehicle safety.

Also, it was explained and shown on a map that the experiment consists of four sections in which the ADS will be simulated. It was also clarified that if a DMS-uncertainty occurs during the test drive, one of four different concepts will be presented, but that the concept does not vary within one section. After that, the compensation task, in this case, a SuRT, was explained and exercised by the participants. It was explained that if a reliable detection is not possible, the driver can avoid a system adaption by performing this task, and thus let the DMS know that the driver is still receptive. Participants were able to decide between *"Option A: Performance of the compensation task"* (see figure 7.8) and *"Option B: Limitation of the ADS"* (see figure 7.9) by pressing specific buttons on the modified multi-function steering wheel dummy. Similar to study 1, participants were asked to draw the curtain when the motorway entrance was reached. In this study, the participants' task was to observe the surrounding traffic throughout the test drive unless they were requested to unlock a combination lock. The task request and code were displayed on the center console tablet after the L3 ADS was simulated for four minutes. This task was integrated into this experiment to avoid an implausible simulation of a DMS-uncertainty period. Therefore, the locks were installed into the central armrest, forcing participants to turn away for entering the code (see figure 7.2 (c) and (d)).

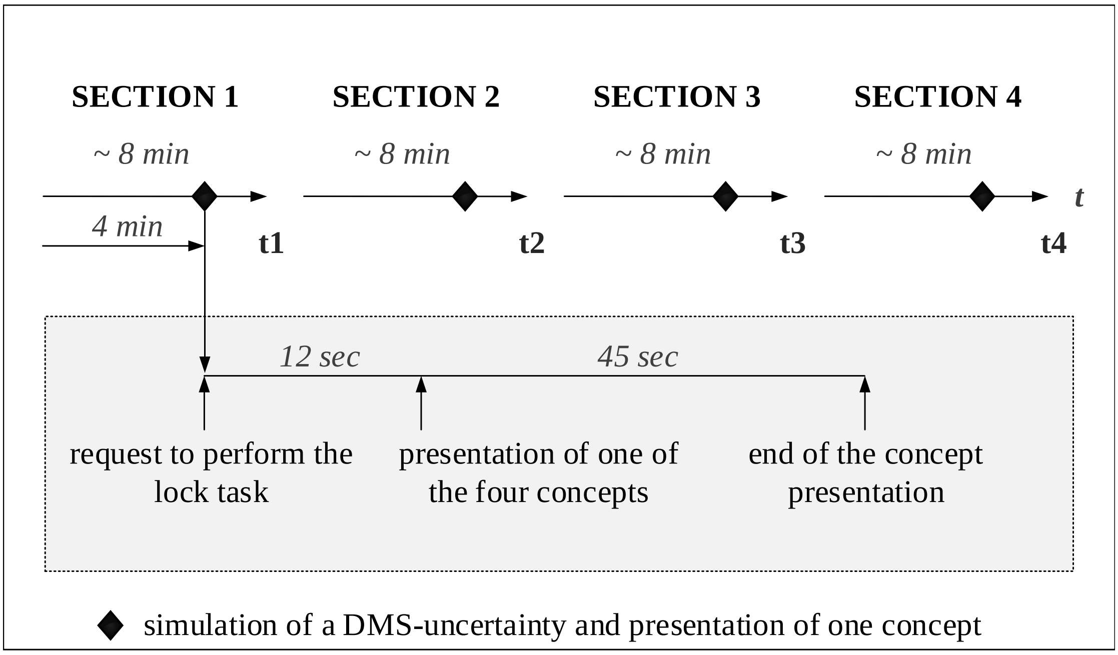


Figure 7.10: Procedure and timing of the questioning (t1 - t4)

Twelve seconds after this task request, one of the four concepts was displayed for 45 seconds (following the suggestion of Wimmer (2014)). During this time span, the drive wizard did perform a passing maneuver only when a participant chose the compensation task while Concept$_D$ was presented. The initiations of Concept$_A$, Concept$_B$ and Concept$_D$ were additionally signaled acoustically. In each section, one DMS-uncertainty period was simulated. After each section, participants were asked to answer different questions regarding their trust in automation, its acceptance, and the experienced concept. Figure 7.10 illustrates this procedure.

7.3 Results

Hergeth et al. (2016) found that trust in automation increases over time. Therefore, it was analyzed whether the sequence of the presented concepts influenced users' trust in the automation and acceptance ratings. To exclude such effects, the data from the first (t1) and last measurement (t4) were compared by a Wilcoxon test. No sequence effect was observed for all of these comparisons (usefulness: $z = -0.604$, $p = .546$, satisfying: $z = -0.859$, $p = .390$, trust: $z = -0.130$, $p = .896$, and mistrust: $z = -0.748$, $p = .454$).

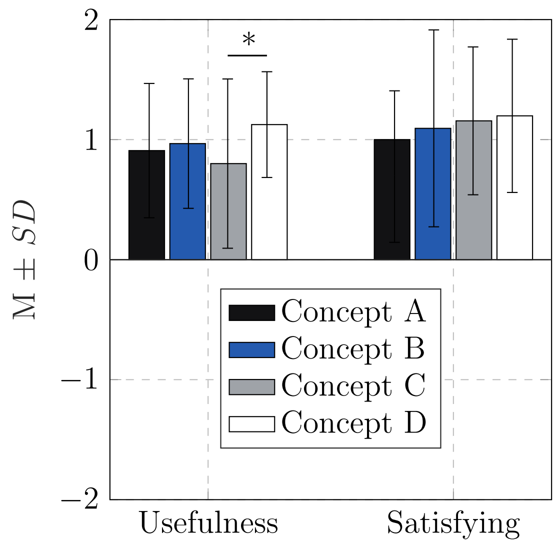


Figure 7.11: Acceptance assessment

It was shown that the type of concept significantly influenced the perceived usefulness of the simulated ADS, as assessed by a Friedman test ($\chi^2(3) = 11.12$, $p = .011$). Due to multiple comparisons, a Bonferroni correction was performed, leading to an adjusted significance level of $p < .008$. Post-hoc tests showed that the usefulness of the ADS was rated significantly higher when Concept_D was presented in case of a DMS uncertainty compared to Concept_C ($p = .007$, $d_{\text{Cohen}} = 0.553$). There was no statistically significant difference between the ADS satisfaction ratings depending on the concept, ($\chi^2(3) = 3.74$, $p = .291$). Figure 7.11 presents the results of the acceptance ratings.

Friedman tests revealed that the type of concept did not influence trust ($\chi^2(3) = 0.260$, $p = .967$) and mistrust ($\chi^2(3) = 3.708$, $p = .295$) in automation. Descriptive analysis demonstrated that the ADS received the highest trust ratings when

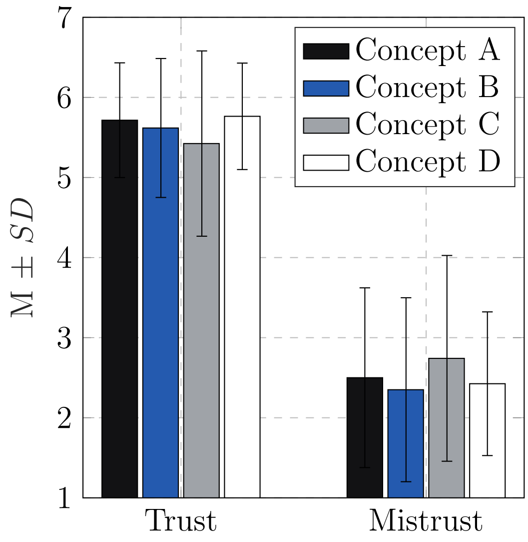


Figure 7.12: Trust and mistrust assessment

Concept_D was presented. ADS-mistrust ratings were lowest for Concept_B and highest for Concept_C. Figure 7.12 summarizes these results. When asked to declare the feedback concept one would prefer most, Concept_D was rated best. In total, 50 percent of the sample rated Concept_D their first choice. Concept_B was chosen by 37.5 percent, Concept_C by 8.34 percent, and Concept_A by 4.16 percent of the sample.

It was analyzed whether participants performed the compensation task to avoid a system adaption (Approach B/Concept_D, see figure 7.1). It was found that 63 percent of the sample selected the SuRT to bypass the system adaption. Also, participants considered the possibility to perform a compensation task to avoid a system adaption useful $M = 5.21$, $SD = 1.87$ (on the seven-point Likert scale: "I find the possibility to perform a compensation task to avoid a system restriction (e.g., the system ceases to change lanes) useful.").

Further, it was found that, on average, participants rated the uncertainty icon presented in Concept_B (see figure 7.5) as easy to understand $M = 5.88$, $SD = 1.23$ ("The icon illustrates a missing driver-state detection in an easy to understand manner."). Additionally, participants declared whether they would mind the signal tone. Participants would not want to miss the signal tone, regardless of the concept (Concept_A $M = 1.88$, Concept_B $M = 1.83$, and Concept_D $M = 1.96$). The additional analysis of gaze data showed that 94.2 percent of the participants looked ahead (likely into the instrument cluster) in response to the activation of the visual and acoustic feedback of Concept_A, Concept_B, and Concept_D. Video data of one participant could not be analyzed due to a recording error. Thus, 69 data sets (23 per concept) were available for this analysis. Participants stated that they would likely adapt their behavior to increase the probability of correct detection to avoid a system restriction M = 5.58, *SD* = 1.38 ("If a reliable detection is not possible (e.g., due to turning away or due to unfavorable lighting conditions), I would likely adapt my behavior to increase the probability of a correct detection and to avoid a system restriction.").

Two different attitudes were observed about the wish to be informed about an active camera-based DMS.

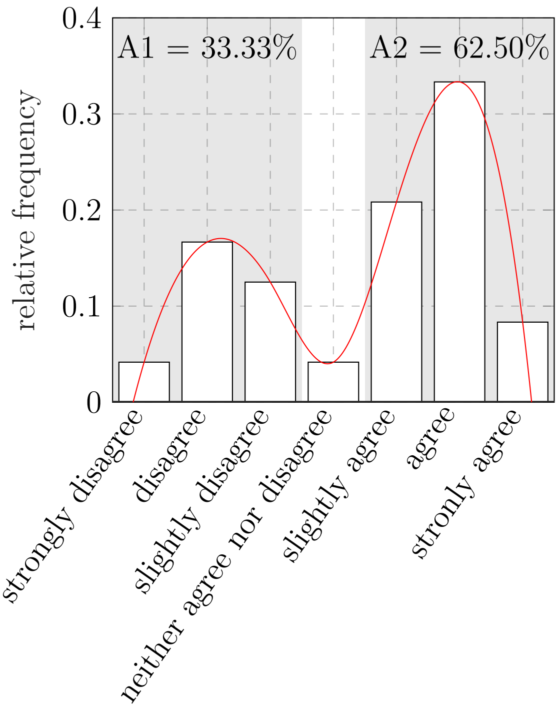


Figure 7.13: "I wish to be permanently informed about an active camera-based DMS."

Of the participants, 62.50 percent want to be permanently informed about an active camera-based system (slightly agree - strongly agree), while 33.33 percent do not wish to be informed about that (slightly disagree - strongly disagree). Figure 7.13 presents the distribution function of the wish to be informed about an active camera-based DMS.

Wizard-of Oz - Immersion of automated driving

On average, participants rated that they could get a feeling of automated driving (M = 7.35) (SD = 1.34). This was assessed on the scale ranging from 1 (no impression of automated driving) to 10 (strong impression of automated driving).

7.4 Discussion

The results showed that the type of concept significantly influenced the perceived usefulness of the simulated L3 ADS. It became clear that Concept_D, which allowed participants to select a compensation task to avoid a system restriction, resulted in the highest usefulness, satisfying and trust ratings of the simulated ADS. Additionally, about two-thirds of the sample decided to perform this task to avoid a system restriction. Thus, *Approach B* represents an accepted and useful approach to deal with a DMS-uncertainty period. This is in line with the considerations of Bengler (2011), who suggested that a presentation of information can increase system transparency and therefore has the potential to be an alternative to unclear adaptive systems. It can be deduced from this that in the case of a DMS-uncertainty period, an offer of a compensation task should be provided before a system adaption is performed.

In this study, the information about the DMS-uncertainty and the accompanying system restriction (Concept_B) led descriptively to slightly higher trust and lower mistrust ratings than no feedback (Concept_C). Thus, similar to the results of Beller et al. (2013) and Hergeth (2016), this study indicates that providing uncertainty information is more likely to increase and not to decrease trust in automation as it was found by Helldin et al. (2013).

A presentation of the cause of a system adaption (Concept_B) (explicit feedback) descriptively led to slightly higher usefulness and satisfaction ratings compared to Concept_A (implicit feedback). Also, Concept_B was preferred by 37.5 percent of the sample, whereas Concept_A was named the preferred option by 4.16 percent of the sample. Thus, it can be concluded that users wish to be informed about the cause of a system adaption if possible. Further, the uncertainty icon of Concept_B was assessed as easy to understand by the majority of the sample and could, therefore, represent an alternative to the uncertainty face suggested by Beller et al. (2013).

Results of this study support the transparency principles provided by Norman

(1990) and Endsley (2017). Representing no feedback at all (Concept$_C$) resulted in highest mistrust and lowest trust and acceptance ratings and thus would not be a suitable concept for dealing with DMS-uncertainty periods from a user's perspective.

Of the sample, 94.2 percent stopped to perform the lock task and looked ahead as a result of the visual and auditory feedback in Concept$_A$, Concept$_B$, and Concept$_D$. Thus, providing visual and acoustic feedback can be considered helpful in dealing with DMS-uncertainty periods, regardless of the concept. This is because the probability of correct driver- or driver-state detection would likely increase as a consequence of this behavior change.

Some limitations had to be accepted for this study. First, the interaction between an ADS and a DMS was only simulated. The DMS was time-based and linked to the lock task. Therefore, feedback concepts might responded differently than if a real DMS had been used. In each section, a concept was triggered only once. Further studies should, therefore, evaluate the influence of DMS-reliability (including frequency and duration of such an uncertainty) on users' trust and acceptance of automation. The system performance assessment concept provided in the RobustSENSE project has been kept rather general for DMSs. Therefore, further research should focus on the development of reliable DMSs and on the performance assessment of DMSs. In this study, Concept$_D$ with the SuRT as a compensation task was the most accepted and preferred concept. More personal or exciting bypass-tasks, however, could have the potential to increase the acceptance of this approach. Future studies should address this.

Conclusion

The DMC triggers different strategies depending on a user's drowsiness level (see chapter 3). Thus, the wrong timing of such strategies could impair safety, but probably also the acceptance and perceived usefulness of an L3 ADSs. Hence, knowledge about how a DMC should be designed tolerant of uncertainty periods of a DMS is required. Overall, four concepts were identified that could help deal with such periods. The impact of these concepts on users' trust and acceptance of automation was assessed in a counterbalanced within-subjects design (N = 24). It was found that offering a compensation task has the potential to be an accepted approach for dealing with DMS-uncertainty periods. It also became clear that the majority wishes to be informed about the cause of a system adaption and about an active camera-based DMS. Thus, this study underpins the strong need and desire of users for transparent automated systems.

CHAPTER 8

General discussion

This chapter provides an overview of the main findings of the different studies. Following that, the methodological and conceptual implications are discussed.

8.1 Summary of the results

As drowsy users of an L3 ADS might need more time to sufficiently understand a current situation, and could consequently exceed the total time budget or over-react when it comes to unexpected take-over situations, this thesis considered driver drowsiness another system limit of an L3 ADS (see chapter 3). Hence, to prevent possible safety impairments caused by driver drowsiness, a DMC was developed that takes the perspective of users and manufacturers into account (see chapter 3). This thesis also investigated key elements of the developed DMC (see chapter 5, 6, and 7).

Study 1 (see chapter 5) provided a methodological basis for the investigation of drowsiness effects in the context of automated driving. The study demonstrated that the development of driver drowsiness varies widely between participants and can be induced to a large extent – without sleep deprivation and during regular working hours – by various measures (e.g., no caffeinated beverages, relaxing music, simulated L3 ADS). In this study, 63.33 percent of the sample reached the highest level of drowsiness. Requests to intervene were triggered depending on specific drowsiness levels. No statistically significant differences in take-over time metrics were found between the different drowsiness levels. However, some participants reacted startled when a request to intervene happened in the drowsy state, which did not occur in the non-drowsy condition. Rivera, Talone, Boesser, Jentsch, and Yeh (2014) distinguished between startle and surprise responses. Startle is usually provoked by intensive stimulation (Rivera et al., 2014), whereas surprise occurs as a result of an unexpected event (Horstmann, 2006) (for further information on the differentiation between startle and surprise see Rivera et al. (2014)). Subjective assessments of the system-based strategy showed that higher escalations like an MRM should be avoided from the users' perspective. Subjective assessments of different options of the DSRS revealed that a targeted offer of NDRTs has the potential to be an accepted option of a DSRS.

Subsequently, Study 2 (see chapter 6) assessed the reactivation potential and the effectiveness of NDRTs. For this, KSS ratings greater than 7 were considered the system limit driver drowsiness. It was found that NDRTs have a high reactivation potential and that the reactivation in most cases remained effective beyond the actual reactivation phase. To trigger different strategies and options (such as offering specific NDRTs) depending on a user's drowsiness level, reliable and continuous information about a user's drowsiness state is needed. The performance of DMSs, however, can be impaired by different circumstances (e.g., by unfavorable light conditions (Friedrichs & Yang, 2010), see also section 2.4). Therefore, a DMC should be designed to be tolerant towards periods in which a DMS cannot reliably detect a user's drowsiness level.

Study 3 (see chapter 7) focused on how DMC-uncertainty periods can be managed. For this, the influence of four different concepts on users' trust and acceptance of automation was evaluated. It was found that offering a compensation task to bypass a DMS-uncertainty period has the potential to be an accepted and trusted approach to deal with DMS-uncertainty periods from the users' perspective. It became apparent that users wish to be informed about the cause of a system adaption and about an active camera-based DMS.

8.2 Methodological implications

Terminology

This thesis recommends distinguishing between the constructs drowsiness/sleepiness and fatigue, especially in the context of L3 ADSs. Fatigue, for example, increases with the duration of a demanding task (Johns, 2007; Åkerstedt, 2011) and can lead to attention and vigilance problems (Brown, 1994). Further, fatigue is considered a low-activation negative affect (in the affect quadrant of depression), whereas drowsiness is regarded as a low-activation positive affect (in the affect quadrant of comfort) (Warr, 2013). L3 ADSs aim to increase comfort (Hoeger et al., 2011), which already suggests that driver drowsiness, and likely not fatigue, is the driver state that needs to be taken into account for L3 ADSs. Thus, it appears unlikely that users will continue to use a system that reinforces a negative feeling such as fatigue. In contrast, drowsiness, which is *"a transitional state between wakefulness and sleep"* (Johns, 1998) appears to be relevant, as it can impair information processing (Mullins et al., 2014) and consequently has the potential to negatively influence take-over performance. Also, users wish to relax during the use of ADSs (e.g., Yang et al., 2019). However, relaxing has high somnificity scores (Johns, 2002) and is therefore likely to increase driver drowsiness. Further, it is doubtful that the fatigue model provided by May and Baldwin (2009) already took L3 ADSs into account, as the authors pointed out, *"Passive fatigue is produced when a driver is mainly monitoring the driving environment over an extended period of time when most or the entire actual driving task is automated."* (May & Baldwin, 2009). However, users of an L3 ADS do not need to monitor such a system (SAE, 2018). Instead, a user of an L3 ADS needs to be receptive, and therefore must be able to *"reliably and appropriately focus his/her attention in response to a stimulus"* (SAE, 2018). Hence, to investigate the influence of driver drowsiness in the context of L3 ADSs, it is suggested that no studies should be made on the fatigue-model provided by May and Baldwin (2009).

Drowsiness induction

The measures used to induce driver drowsiness can be considered appropriate, as a large number of participants reached DL_4 and DL_6 (see chapter 5). Hence, future studies could use such a procedure to induce driver drowsiness in the context of automated driving. As a larger number of participants are likely to get drowsy, this can increase the availability of appropriate data and, consequently, the statistical power of the results. Besides this, it became evident that drowsiness develops very individually. Hence, a state-dependent approach is needed to evaluate the influence of drowsiness on take-over performance instead of a fixed-time approach. However, this requires an online assessment of driver drowsiness. One suitable approach would be to use an existing DMS that can assess different drowsiness levels to trigger state-dependent take-over situations. For example, this procedure was used by Schömig et al. (2015). For this procedure, however, such a DMS must be available. Further, it faces several limitations (e.g., results strongly depend on the reliability of the used DMS and can hardly be replicable if the DMS is not a commercial product). For the investigation of fundamental drowsiness effects, strategies to deal with this driver state or HMI concepts, expert ratings of driver drowsiness appear to be an appropriate method. This method does not rely upon state-of-the-art DMSs, and results can be replicated. Several recently published articles also assessed driver drowsiness (or fatigue) based on expert ratings (e.g., Kreuzmair et al., 2017; Vogelpohl et al., 2018; Feldhütter et al., 2018; Naujoks, Höfling, Purucker, & Zeeb, 2018). In contrast to these, the fixed-time approach has rarely been used since 2017 (e.g., Jarosch & Bengler, 2019). This supports that the expert rating of driver drowsiness is a helpful method to study drowsiness in the context of automated driving.

Wizard-of-Oz approach

The majority of participants was able to get a feel for automated driving. On average, immersion rating was slightly higher when using the Audi Q7 than with the Audi A4, which had less elaborate modifications. However, standard deviations were smaller for the Audi Q7. Hence, it appears that the modifications made in the Audi Q7 might be helpful to generate a more homogeneous impression of an L3 ADS. However, it must be considered that the study purpose differed between the three studies, and therefore other impact factors might have contributed to the smaller standard deviations. Further, in recent years the number of other researchers taking a Wizard-of-Oz approach has increased sharply (Baltodano et al., 2015; Wang, Sibi, Mok, & Ju, 2017; Gold & Meyer, 2017; Marberger, Manstetten, & Korthauer, 2017; Cabrall, Petrovych, & Happee, 2018), indicating that there is still a great need for methods that allow a more realistic investigation of the interaction between humans and automation.

8.3 Conceptual implications

Several conceptual implications can be derived based on the study results. As the occurrence and development of driver drowsiness differ strongly between the various participants (see chapter 5), it will not be possible to manage driver drowsiness just by limiting the pilot availability to a specified duration. Instead, it is necessary to gain information about a user's drowsiness state through a DMS approach. Hence, a manufacturer must decide between the different DMS-approaches (see section 2.5). If the camera-based DMS approach is used, one should also consider that several cameras are likely to be required to reliably detect driver drowsiness, due to the likely performed NDRTs (Hecht et al., 2019).

Different studies found no clear link between driver drowsiness (or fatigue) and a decrement of take-over performance (Feldhütter et al., 2017; Jarosch et al., 2017; Kreuzmair et al., 2017; Weinbeer et al., 2017; J. Schmidt, 2018; Naujoks, Höfling, et al., 2018). However, as the results of Gonçalves et al. (2016) showed that drowsiness led to higher lateral accelerations in the event of a take-over situation, in Study 1 of this work startled sounds were observed among drowsy participants in response to a request to intervene (Weinbeer et al., 2017) (see chapter 5), and the findings of Feldhütter et al. (2018) showed that the longitudinal acceleration significantly increased when participants were drowsy, it seems reasonable to conclude that drowsiness in the context of automated driving rather tends to lead to overreaction than to a meaningful increase of ToT metrics. As the effects of driver drowsiness on take-over performance are not clear and consistent (Radlmayr et al., 2019), the validity of the general formulated Yerkes-Dodson Law (see section 2.2) for take-over performance in the context of automated driving must be questioned.

Since some participants were not able to take over in a controlled manner, a state-dependent adaption of the request to intervene could be helpful to avoid overreactions caused by a request to intervene. When such overreactions are observed during the development process of an L3 ADS functionality, several measures can be taken to avoid potentially hazardous take-over reactions. First, the drowsiness level, which was considered the system limit of a DMC, can be adjusted. Hence, already a lower drowsiness level should be viewed as the system limit. Second, the preparation strategy could be a suitable approach. However, this strategy could only be executed when no other system limit (e.g., sensor failures) has occurred (see section 3.3.3).

In case of a system-based strategy, a reduction in maximum speed, an adjustment of driving behavior (no further lane changes and driving on the slow lane) or a rest at a service station were rated best in terms of acceptance. In contrast, an MRM that would stop the vehicle on the emergency or ego lane was rejected by the majority of participants. Hence, it can be concluded that higher escalations should be avoided from the users' perspective (Weinbeer et al., 2018). This finding should also be taken into account for the specification of L3 ADSs.

Study 2 (see chapter 6) demonstrated that it is possible to reduce driver drowsiness by a targeted offer of NDRTs and that this reactivation in most cases remained effective even beyond the original reactivation phase. Hence, manufacturers can offer a specific task or a selection of specific tasks when a user reaches or exceeds a certain drowsiness level. If it is known which tasks help to reduce a user's drowsiness state, individual suggestions could be made. Thus, manufacturers might consider this option. For the implementation of such a driver-state related option, appropriate HMI concepts need to be developed.

Several recommendations on how DMS-uncertainty periods should be managed from the users' perspective can be derived from the results of Study 3 (see chapter 7). It was found that offering a task to bypass DMS-uncertainty periods has the potential to be an accepted and trusted approach. It became apparent that the majority wishes to be informed about the cause of a system adaption. Also, about two-thirds of the sample wish to be informed when a camera-based DMS is active. Hence, Study 3 demonstrated the users need and wish for feedback and transparent L3 ADSs (see chapter 7).

8.4 Limitations and future research

Besides the limitations described in the different studies (see chapters 5, 6, and 7), it must be kept in mind that the samples used in these studies consisted exclusively of employees of AUDI AG. Thus, the sample differed from the general population in terms of nationality, age, and likely technical affinity. However, customers who will pay for and use an L3 functionality are likely to have an affinity for new technologies. Thus, the results can be considered meaningful for the specification of L3 ADSs. However, further research is needed to clarify whether the development of driver drowsiness and the acceptance and effectiveness of different strategies differ between nationalities and age groups.

It must be noted that, unlike driving simulator studies, Wizard-of-Oz studies face other limitations. An automated system is simulated by a drive wizard and not by a real ADS. Therefore, results might depend on the driving behavior of a drive wizard. Hence, objectivity might be compromised due to this more realistic setting. To reduce the possible impact on participants' assessments, the drive wizard was always the same within one study. ACC and LKA were not used in these studies as this would represent a state-of-the-art system rather than a future motorway pilot. Apart from this, it must be considered that situations occur in which a drive wizard may not behave like a future L3 ADS. Challenging situations appeared to be close cut-in situations (especially by trucks likely due to the rather low maximum velocity of 130km/h) and approach lanes also leading to cut-in situations. Further, other factors such as traffic density vary depending on the time of day. To control for this, every condition of the different studies was always distributed throughout the day. In this thesis, participants were always informed that the investigator simulates an L3 ADS and not a real technical system. Here, also a covered method could be used. The sample, however, consisted solely of employees of the AUDI AG. Therefore, it is likely that some participants would not believe that the used Audi Q7 or Audi A4 are equipped with a real L3 ADS. Hence, some participants could feel deceived, whereas others might be nervous as they believe that they are driving for the first time with an L3 ADS, which might also influence the study results. To avoid such inhomogeneities, an open and not a covert method was used in this case.

As the number of Wizard-of-Oz studies increased in the recent years and as it appears that Wizard-of-Oz studies are a helpful method, especially for fundamental human-factors research questions, future research should focus on standardization of Wizard-of-Oz studies to enhance the objectivity and transferability of the study results.

Also, the development and the influence of driver drowsiness on take-over performance in the context of automated driving should be evaluated in a longitudinal analysis. This could give insights if driver drowsiness should and can be managed more individually.

The DMC developed in this thesis is based on a relation between an L3 ADSs and a DMSs (see chapter 3). Besides this, the relevance of DMSs is likely to increase due to the Euro NCAP Roadmap 2025, which sees DMSs as a prerequisite for automated driving (Euro NCAP, 2017). So far, it is unclear how the reliability of such systems will be assessed. Hence, further research is needed to identify meaningful test scenarios for the assessment of DMSs. However, as various factors will influence the accuracy of non-intrusive camera-based DMSs, especially during an automated drive (e.g., positioning of the camera(s), user posture, seat position, steering wheel adjustment, performing NDRTs, light conditions), the definition of the testing procedure will be extremely challenging. Therefore, this thesis suggests that the focus of assessment should not be only on the reliability of DMSs in specific test scenarios, but also on how DMS-uncertainty periods are managed and to what extent this influences the users' trust and acceptance. This is supported by the results of Study 3 (see chapter 7).

Further, several key elements of the DMC were evaluated in this thesis. Therefore, in a next step, the developed DMC should be evaluated with a real L3 ADS in combination with a DMS. Also, further research must take the perspective of other road-users into account, especially when higher system-based escalations are needed.

CHAPTER 9

Conclusion

Users would like to use ADSs especially if they are in an unfavorable state (e.g., when they feel tired (Payre et al., 2014)). However, while using an L3 ADS, users will not be allowed to sleep (see chapter 2.1). Furthermore, the announced ADSs focus more on comfort in relatively safe traffic situations (Bengler et al., 2017). Hence, there is a risk of a mismatch between the users' wishes and expectations and the actual system limits. In addition, with the introduction of L3 ADSs, the role of DMSs is likely to change from a recommendation to an intrusive, restrictive nature when drowsiness is detected during an L3 drive. However, restrictive systems can lead to lower system acceptance than informative systems (van der Laan et al., 1997). As a result, this may cause an L3 ADS to be used less frequently. However, a regular use of such systems is required to achieve an increase in traffic safety by introducing ADSs.

Thus, the aim for ADSs to increase safety, user comfort, and traffic efficiency proved to be very challenging and cannot be achieved simply by taking the driver out of the control loop. Instead, it is a considerably complex interaction between users, automated systems, and other road users. Therefore, the human factor must be considered very seriously in this context.

Hence, further research is needed to design the complex interaction between ADSs and users safely and acceptably from the perspective of users, manufacturers, and other road users. In particular, further research should focus on longitudinal studies to examine the individual development and impact of driver drowsiness on take-over performance. In addition, test procedures for DMSs must be developed, and studies need to be conducted that use a real L3 ADS and DMS to investigate a DMC, related feedback concepts, and driver states under real conditions. Besides this, further research should focus on how to standardize Wizard-of-Oz studies to increase the objectivity and transferability of results.

References

Ahmed, I. M., & Thorpy, M. J. (2011). Clinical evaluation of the patient with excessive sleepiness. In M. J. Thorpy & M. Billiard (Eds.), *Sleepiness: Causes, consequences, and treatment* (pp. 36–49). Cambridge: Cambridge University Press.

Åkerstedt, T. (2011). Shift work disorder and sleepiness. In M. J. Thorpy & M. Billiard (Eds.), *Sleepiness: Causes, consequences, and treatment* (pp. 186–203). Cambridge: Cambridge University Press.

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, *52*(1-2), 29–37. doi: 10.3109/00207459008994241

Anund, A., Fors, C., Hallvig, D., Åkerstedt, T., & Kecklund, G. (2013). Observer rated sleepiness and real road driving: An explorative study. *PloS ONE*, *8*(5), e64782. doi: 10.1371/journal.pone.0064782

Anund, A., & Kecklund, G. (2011). Motor vehicle driving and excessive sleepiness. In M. J. Thorpy & M. Billiard (Eds.), *Sleepiness: Causes, consequences, and treatment* (pp. 82–91). Cambridge: Cambridge University Press.

Anund, A., Kecklund, G., Vadeby, A., Hjälmdahl, M., & Åkerstedt, T. (2008). The alerting effect of hitting a rumble strip - a simulator study with sleepy drivers. *Accident analysis and prevention*, *40*(6), 1970–1976. doi: 10.1016/j.aap.2008.08.017

AUDI AG. (2019). *Audi immersive in-car entertainment.* Retrieved from `https://www.audi-mediacenter.com/de/audi-auf-der-ces`

-2019-11175/audi-immersive-in-car-entertainment-11180

Bahram, M., Aeberhard, M., & Wollherr, D. (2015). Please take over! An analysis and strategy for a driver take over request during autonomous driving. In *2015 IEEE intelligent vehicles symposium (IV)* (pp. 913–919). doi: 10.1109/IVS.2015.7225801

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. doi: 10.1016/0005-1098(83)90046-8

Baltodano, S., Sibi, S., Martelaro, N., Gowda, N., & Ju, W. (2015). The RRADS platform: A real road autonomous driving simulator. *Proceedings of the 7th International Conference on Automotive User Interfaces and Vehicular Applications*, 281–288. doi: 10.1145/2799250.2799288

Beller, J., Heesen, M., & Vollrath, M. (2013). Improving the driver-automation interaction: An approach using automation uncertainty. *Human factors*, *55*(6), 1130–1141. doi: 10.1177/0018720813482327

Bengler, K. (2011). The art to make an error: The dilemma between prevention, learning and mitigation. In P. C. Cacciabue, M. Hjälmdahl, A. Luedtke, & C. Riccioli (Eds.), *Human modelling in assisted transportation* (pp. 9–13). Milano: Springer Milan.

Bengler, K. (2015). Grundlegende Zusammenhänge von Automatisierung und Fahrerleistung. *Zeitschrift für Verkehrssicherheit*(3), 169–173.

Bengler, K., Dietmayer, K., Färber, B., Maurer, M., Stiller, C., & Winner, H. (2014). Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent Transportation Systems Magazine*, *6*(4), 6–22. doi: 10.1109/MITS.2014.2336271

Bengler, K., Winner, H., & Wachenfeld, W. (2017). No human – no cry? *at - Automatisierungstechnik*, *65*(7). doi: 10.1515/auto-2017-0021

Billings, C. E. (1991). *Human-centered aircraft automation: A concept and guidelines.* Moffett Field, California: National Aeronautics and Space Administration, Ames Research Center.

Borbély, A. A. (1982). A two process model of sleep regulation. *Human Neurobiol*, *1*(3), 195–204.

Brown, I. D. (1994). Driver fatigue. *Human factors*, *36*(2), 298–314. doi: 10.1177/001872089403600210

Bubb, H., Bengler, K., Grünen, R. E., & Vollrath, M. (2015). *Automobilergonomie.* Wiesbaden: Springer Fachmedien Wiesbaden. doi: 10.1007/978-3-8348-2297

-0

Bundestag. (2017a). Achtes Gesetz zur Änderung des Straßenverkehrsgesetzes. *BGBl. Teil 1 Nr. 38*, 1648–1650.

Bundestag. (2017b). Beschlussempfehlung und Bericht des Ausschusses für Verkehr und digitale Infrastruktur (15. Ausschuss): zu dem Gesetzentwurf der Bundesregierung – Drucksachen 18/11300, 18/11534, 18/11683 Nr. 10.

Cabrall, C. D. D., Petrovych, V., & Happee, R. (2018). Looking at drivers and passengers to inform automated driver state monitoring of in and out of the loop. In N. Stanton (Ed.), *Advances in human aspects of transportation* (pp. 695–707). Cham: Springer.

Cadillac. (2017). *Cadillac super cruise.* Retrieved from `https://www.cadillac.com/world-of-cadillac/innovation/super-cruise`

Čolić, A., Marques, O., & Furht, B. (2014). *Driver drowsiness detection: Systems and solutions.* Cham: Springer International Publishing. doi: 10.1007/978-3-319-11535-1

Cummings, P., Koepsell, T. D., Moffat, J. M., & Rivara, F. P. (2001). Drowsiness, counter-measures to drowsiness, and the risk of a motor vehicle crash. *Injury Prevention*, *7*, 194–199. doi: 10.1136/ip.7.3.194

Davidsson, S. (2012). Countermeasure drowsiness by design - using common behaviour. *Work*, *41*(Supplement 1), 5062–5067. doi: 10.3233/WOR-2012-0798-5062

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance.* London and New York: Academic Press.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982–1003. doi: 10.1287/mnsc.35.8.982

De Valck, E., & Cluydts, R. (2001). Slow-release caffeine as a countermeasure to driver sleepiness induced by partial sleep deprivation. *Journal of Sleep Research*, *10*(3), 203–209. doi: 10.1046/j.1365-2869.2001.00260.x

Dinges, D. F., & Mallis, M. M. (1998). Managing fatigue by drowsiness detection: Can technological promises be realized? In *Managing fatigue in transportation* (pp. 209–229). Oxford: Pergamon Press.

Drake, C. L. (2011). Subjective measures of sleepiness. In M. J. Thorpy & M. Billiard (Eds.), *Sleepiness: Causes, consequences, and treatment* (pp. 60–71). Cambridge: Cambridge University Press.

Drucker, P. F. (2009). *The practice of management.* New York, NY: HarperCollins.

DS. (2017). *DS 7 crossback technology.* Retrieved from https://www.dsautomobiles.co.uk/ds-models/ds-7-crossback/design/technology#

Ebrahim, P. (2016). *Driver drowsiness monitoring using eye movement features derived from electrooculography* (Dissertation). Universität Stuttgart, Stuttgart.

Eichinger, A. (2011). Untersuchungskonzepte für die Evaluation von Systemen zur Erkennung des Fahrerzustands. *Berichte der Bundesanstalt für Straßenwesen F 80*.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human factors*, *59*(1), 5–27. doi: 10.1177/0018720816681350

Engl, S. (n.d.). *mobile gaming: Eine empirische Studie zum Spielverhalten und Nutzungserlebnis in mobilen Kontexten* (Magisterarbeit, Universität Regensburg, Regensburg). Retrieved from http://www.stephanengl.com/wp-content/uploads/presentations-publications/Mobile-Gaming-by-Stephan-Engl.pdf

Euro NCAP. (2017). Euro NCAP 2025 roadmap: In pursuit of vision zero.

Feldhütter, A., Gold, C., Schneider, S., & Bengler, K. (2017). How the duration of automated driving influences take-over performance and gaze behavior. In C. M. Schlick et al. (Eds.), *Advances in ergonomic design of systems, products and processes* (pp. 309–318). Berlin, Heidelberg: Springer Berlin Heidelberg.

Feldhütter, A., Kroll, D., & Bengler, K. (2018). Wake up and take over! The effect of fatigue on the take-over performance in conditionally automated driving. *Proceedings of the IEEE 21th International Conference on Intelligent Transportation Systems (ITSC) 2018*, 2080–2085. doi: 10.1109/ITSC.2018.8569545

Filtness, A. J., Armstrong, K. A., Watson, A., & Smith, S. S. (2017). Sleep-related crash characteristics: Implications for applying a fatigue definition to crash reports. *Accident analysis and prevention*, *99*(Part B), 440–444. doi: 10.1016/j.aap.2015.11.024

Friedrichs, F., & Yang, B. (2010). Camera-based drowsiness reference for driver state classification under real driving conditions. In *2010 IEEE intelligent vehicles symposium (IV)* (pp. 101–106). doi: 10.1109/IVS.2010.5548039

Gaspar, J. G., Brown, T. L., Schwarz, C. W., Lee, J. D., Kang, J., & Higgins, J. S. (2017). Evaluating driver drowsiness countermeasures. *Traffic injury prevention*, *18*(sup1), S58-S63. doi: 10.1080/15389588.2017.1303140

Gasser, T. M., Arzt, C., Ayoubi, M., Bartels, A., Bürkle, L., Eier, J., ... Vogt, W. (2012). Rechtsfolgen zunehmender Fahrzeugautomatisierung. *Berichte der Bundesanstalt für Straßenwesen Heft F 83*.

Gasser, T. M., Schmidt, E. A., Bengler, K., Chiellino, U., Diederichs, F., Eckstein, L., ... Zeeb, E. (2015). Bericht zum Forschungsbedarf - Runder Tisch Automatisiertes Fahren – AG Forschung.

George, P. T. (2018). The psycho-sensory wake drive - a power source for power naps and other common sleep-wake phenomena: a hypothesis. *Sleep & breathing = Schlaf & Atmung*, *22*(1), 41–48. doi: 10.1007/s11325-017-1505-6

Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the technology acceptance model to assess automation. *Cognition, Technology & Work*, *14*(1), 39–49. doi: 10.1007/s10111-011-0194-3

Gillberg, M., Kecklund, G., & Åkerstedt, T. (1994). Relations between performance and subjective ratings of sleepiness during a night awake. *Journal of Sleep Research*, *17*(3), 236–241.

Gold, C. (2016). *Modeling of take-over performance in highly automated vehicle guidance* (Dissertation). Technische Universität München, München.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1938–1942. doi: 10.1177/1541931213571433

Gold, C., Happee, R., & Bengler, K. (2018). Modeling take-over performance in level 3 conditionally automated vehicles. *Accident analysis and prevention*, 3–13. doi: 10.1016/j.aap.2017.11.009

Gold, C., Körber, M., Hohenberger, C., Lechner, D., & Bengler, K. (2015). Trust in automation – Before and after the experience of take-over scenarios in a highly automated vehicle. *Procedia Manufacturing*, *3*, 3025–3032. doi: 10.1016/j.promfg.2015.07.847

Gold, C., & Meyer, M.-L. (2017). Take-over performance in a wizard of oz vehicle in level 3 conditionally automated driving. *Active safety and automated driving, 3. interdisciplinary expert dialog*.

Gonçalves, J., & Bengler, K. (2015). Driver state monitoring systems – Transferable knowledge manual driving to HAD. *Procedia Manufacturing*, *3*, 3011–3016. doi: 10.1016/j.promfg.2015.07.845

Gonçalves, J., Happee, R., & Bengler, K. (2016). Drowsiness in conditional automation: Proneness, diagnosis and driving performance effects. In *2016 IEEE 19th international conference on intelligent transportation systems (ITSC)* (pp. 873–878). doi: 10.1109/ITSC.2016.7795658

Graf, O. (1960). *Arbeitsphysiologie*. Wiesbaden: Betriebswirtschaftlicher Verlag Dr. Tb. Gabler GmbH. doi: 10.1007/978-3-663-02317-3

Grandjean, E. (1979). Fatigue in industry. *British Journal of Industrial Medicine*, *36*(3), 175–186.

Hashemi Nazari, S. S., Moradi, A., & Rahmani, K. (2017). A systematic review of the effect of various interventions on reducing fatigue and sleepiness while driving. *Chinese journal of traumatology = Zhonghua chuang shang za zhi*, *20*(5), 249–258. doi: 10.1016/j.cjtee.2017.03.005

Hebb, D. O. (1955). Drives and the C.N.S. (conceptual nervous system). *Psychological Review*, *62*(4), 243–254. doi: 10.1037/h0041823

Hecht, T., Feldhütter, A., Radlmayr, J., Nakano, Y., Miki, Y., Henle, C., & Bengler, K. (2019). A review of driver state monitoring systems in the context of automated driving. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th congress of the international ergonomics association (iea 2018)* (pp. 398–408). Cham: Springer International Publishing.

Heinze, M., & Weinbeer, V. (2018). *A method for excitation of the ascending reticular activating system of a person in a motor vehicle as well as apparatus* (No. DE102017200601A1).

Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013). Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 210–217. doi: 10.1145/2516540.2516554

Hergeth, S. (2016). *Automation trust in conditional automated driving systems: Approaches to operationalization and design* (Dissertation). Technische Universität Chemnitz, Chemnitz.

Hergeth, S., Lorenz, L., Vilimek, R., & Krems, J. F. (2016). Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Human factors*, *58*(3), 509–519. doi: 10.1177/0018720815625744

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: A new approach. *Psychophysiology*, *10*(4), 431–436. doi: 10.1111/j.1469-8986.1973.tb00801.x

Hoeger, R., Zeng, H., Hoess, A., Kranz, T., Boverie, S., Strauss, M., ... Nilsson, A. (2011). The future of driving. Deliverable D61.1 final report. *Highly automated vehicles for intelligent transport*.

Horne, J., & Reyner, L. (1999). Vehicle accidents related to sleep: a review. *Occupational and Environmental Medicine*, *56*(5), 289–294. doi: 10.1136/oem.56.5.289

Horne, J. A., & Reyner, L. A. (1995). Driver sleepiness. *Journal of sleep research*, *4*(Suppl 2), 23–29. doi: 10.1111/j.1365-2869.1995.tb00222.x

Horne, J. A., & Reyner, L. A. (1996). Counteracting driver sleepiness: Effects of napping, caffeine, and placebo. *Psychophysiology*, *33*(3), 306–309. doi: 10.1111/j.1469-8986.1996.tb00428.x

Horstmann, G. (2006). Latency and duration of the action interruption in surprise. *Cognition and Emotion*, *20*(2), 242–273. doi: 10.1080/02699930500262878

Hörwick, M. A. (2011). *Sicherheitskonzept für hochautomatisierte Fahrerassistenzsysteme* (Dissertation). Technische Universität München, München.

Hu, S., & Zheng, G. (2009). Driver drowsiness detection with eyelid related parameters by support vector machine. *Expert Systems with Applications*, *36*(4), 7651–7658. doi: 10.1016/j.eswa.2008.09.030

IJsselsteijn, W. A., de Kort, Y. A. W., & Poels, K. (2013). The game experience questionnaire: Technische Universiteit Eindhoven. Retrieved from `https://pure.tue.nl/ws/portalfiles/portal/21666907`

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. doi: 10.1111/j.1365-2869.2006.00517.x

Jarosch, O., & Bengler, K. (2019). Is it the duration of the ride or the non-driving related task? What affects take-over performance in conditional automated driving? In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th congress of the international ergonomics association (IEA 2018)* (pp. 512–523). Cham: Springer International Publishing.

Jarosch, O., Kuhnt, M., Paradies, S., & Bengler, K. (2017). It's out of our hands now! Effects of non-driving related tasks during highly automated driving on drivers' fatigue. *9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*.

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71.

Johns, M. W. (1998). Rethinking the assessment of sleepiness. *Sleep Medicine Reviews*, *2*(1), 3–15.

Johns, M. W. (2000). A sleep physiologist's view of the drowsy driver. *Transportation Research Part F: Traffic Psychology and Behaviour*, *3*(4), 241–249. doi: 10.1016/S1369-8478(01)00008-0

Johns, M. W. (2002). Sleep propensity varies with behaviour and the situation in which it is measured: the concept of somnificity. *Journal of Sleep Research*, *11*(1), 61–67. doi: 10.1046/j.1365-2869.2002.00274.x

Johns, M. W. (2007). Drowsy driving and the law a submission to the tasmania law reform institute: re: Issues paper no 12.

Johns, M. W. (2009). What is excessive daytime sleepiness. In P. Fulke & S. Vaughan (Eds.), *Sleep deprivation: Causes, effects and treatment* (pp. 1–37). New York: Nova Science Publishers.

Johns, M. W. (2010). A new perspective on sleepiness. *Sleep and Biological Rhythms*, *8*(3), 170–179. doi: 10.1111/j.1479-8425.2010.00450.x

Jung, S.-J., Shin, H.-S., & Chung, W.-Y. (2014). Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel. *IET Intelligent Transport Systems*, *8*(1), 43–50. doi: 10.1049/iet-its.2012.0032

Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, *16*(3), 126–131. doi: 10.1002/prs.680160304

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T.,

& Fukasawa, K. (2006). Validation of the karolinska sleepiness scale against performance and eeg variables. *Clinical neurophysiology*, *117*(7), 1574–1581. doi: 10.1016/j.clinph.2006.03.011

Karrer-Gauß, K. (2012). *Prospektive Bewertung von Systemen zur Müdigkeitserkennung: Ableitung von Gestaltungsempfehlungen zur Vermeidung von Risikokompensation aus empirischen Untersuchungen* (Dissertation). Technische Universität Berlin.

Kerschbaum, P., Lorenz, L., Hergeth, S., & Bengler, K. (2015). Designing the human-machine interface for highly automated cars — challenges, exemplary concepts and studies. In *2015 IEEE international workshop on advanced robotics and its social impacts (arso)* (pp. 1–6). doi: 10.1109/ARSO.2015.7428223

Kircher, A., Uddman, M., & Sandin, J. (2002). Vehicle control and drowsiness. *Swedish National Road and Transport Research Institute*.

Kiss, M., Schmidt, G., & Babbel, E. (2006). Das Wizard of Oz Fahrzeug: Rapid Prototyping und Usability Testing von zukünftigen Fahrerassistenzsystemen. *3. Tagung Aktive Sicherheit durch Fahrerassistenz*. Retrieved from `https://mediatum.ub.tum.de/doc/1145112/1145112.pdf`

Knipling, R. R., & Wang, J.-S. (1994). Crashes and fatalities related to driver drowsiness/fatigue. *National Highway Traffic Safety Administration*.

Knipling, R. R., & Wierwille, W. W. (1994). Vehicle-based drowsy driver detection: Current status and future prospects. *IVHS America Fourth Annual Meeting*.

Körber, M., & Bengler, K. (2014). Potential individual differences regarding automation effects in automated driving. *Interacción '14 Proceedings of the XV International Conference on Human Computer Interaction*. doi: 10.1145/2662253.2662275

Krajewski, J., Sommer, D., Trutschel, U., & Edwards, D., Golz, M. (2009). Steering wheel behavior based estimation of fatigue. *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 118–124.

Kreuzmair, C., Gold, C., & Meyer, M.-L. (2017). The influence of driver fatigue on take-over performance in highly automated vehicles. *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*.

Lee, J. D., & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human factors*, *46*(1), 50–80.

Lexus Europe Newsroom. (2016). *Lexus history.* Retrieved from https://newsroom.lexus.eu/lexus-history/

MacLean, A. W., Fekken, G. C., Saskin, P., & Knowles, J. B. (1992). Psychometric evaluation of the stanford sleepiness scale. *Journal of Sleep Research*, *1*(1), 35–39.

Marberger, C., Manstetten, D., & Korthauer, A. (2017). Hochautomatisiertes Fahren im öffentlichen Verkehr – Ergebnisse einer Realfahrzeugstudie zur Bewertung der User Experience sowie der Fahrerleistung in Übernahmesituationen. *8. Tagung Fahrerassistenz*.

Marberger, C., Mielenz, H., Naujoks, F., Radlmayr, J., Bengler, K., & Wandtner, B. (2018). Understanding and applying the concept of "driver availability" in automated driving. In N. A. Stanton (Ed.), *Advances in human aspects of transportation, advances in intelligent systems and computing 597* (pp. 595–605). Cham: Springer Cham.

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*, *12*(3), 218–224. doi: 10.1016/j.trf.2008.11.005

McDonald, A. D., Schwarz, C., Lee, J. D., & Brown, T. L. (2012). Real-time detection of drowsiness related lane departures using steering wheel angle. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*, *56*(1), 2201–2205. doi: 10.1177/1071181312561464

Mercedes Benz. (n.d.). *Attention assist: Drowsiness-detection system warns drivers to prevent them falling asleep momentarily.* Retrieved from https://media.daimler.com/marsMediaSite/en/instance/ko/ATTENTION-ASSIST-Drowsiness-detection-system-warns-drivers-to-prevent-them-falling-asleep-momentarily.xhtml?oid=9361586

Michon, J. A. (2985). A critical view of driver behavior models: What do we know, what should we do? In L. Evans & R. C. Schwing (Eds.), *Human behavior and traffic safety* (pp. 485–524). Boston, MA: Springer.

Mullins, H. M., Cortina, J. M., Drake, C. L., & Dalal, R. S. (2014). Sleepiness at work: A review and framework of how the physiology of sleepiness impacts the workplace. *The Journal of applied psychology*, *99*(6), 1096–1112. doi: 10.1037/a0037885

Naujoks, F., Befelein, D., Wiedemann, K., & Neukum, A. (2018). A review of non-driving-related tasks used in studies on automated driving. In N. A. Stanton (Ed.), *Advances in human aspects of transportation, advances in intelligent systems and computing 597* (pp. 525–537). Cham: Springer Cham.

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

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. doi: 10.1177/1541931214581432

NHTSA. (2013). Preliminary statement of policy concerning automated vehicles (NHTSA 14-13). *National Highway Traffic Safety Administration.*

Norman, D. A. (1990). The 'problem' with automation: Inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *327*(1241), 585–593. doi: 10.1098/rstb.1990.0101

Oron-Gilad, T., Ronen, A., & Shinar, D. (2008). Alertness maintaining tasks (AMTs) while driving. *Accident analysis and prevention*, *40*(3), 851–860. doi: 10.1016/j.aap.2007.09.026

Östlund, J., Peters, B., Thorslund, B., Engström, J., Markkula, G., Keinath, A., ... Foehl, U. (2005). Driving performance assessment - methods and metrics. *D2.2.5 AIDE.*

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

Patentschrift. (1922). *Sicherheitsvorrichtung für elektrische Fahrzeuge, insbesondere mit einmänniger Bedienung.* (No. 383547).

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*(Part B), 252–263. doi: 10.1016/j.trf.2014.04.009

Pfleging, B., Rang, M., & Broy, N. (2016). Investigating user needs for non-driving-related activities during automated driving. In J. Häkkila & T. Ojala (Eds.), *Proceedings of the 15th international conference on mobile and ubiquitous multimedia* (pp. 91–99). doi: 10.1145/3012709.3012735

Phipps-Nelson, J., Redman, J. R., & Rajaratnam, S. M. W. (2011). Temporal profile of prolonged, night-time driving performance: breaks from driving temporarily reduce time-on-task fatigue but not sleepiness. *Journal of Sleep Research*, *20*(3), 404-115.

Platho, C., Pietrek, A., & Kolrep, H. (2013). Erfassung der Fahrermüdigkeit. *Berichte der Bundesanstalt für Straßenwesen Heft F 89*.

Pöhler, G., Heine, T., & Deml, B. (2016). Itemanalyse und Faktorstruktur eines Fragebogens zur Messung von Vertrauen im Umgang mit automatischen Systemen. *Zeitschrift für Arbeitswissenschaft*, *70*(3), 151–160. doi: 10.1007/s41449-016-0024-9

Radlmayr, J., Feldhütter, A., Frey, A., Jarosch, O., Marberger, C., Naujoks, F., ... Bengler, K. (2019). Drowsiness and fatigue in conditionally automated driving – towards an integrative framework. In D. de Waard et al. (Eds.), *Proceedings of the human factors and ergonomics society europe chapter 2018.*

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 2063–2067. doi: 10.1177/1541931214581434

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-13*(3), 257–266. doi: 10.1109/TSMC.1983.6313160

Rauch, N., Kaussner, A., Krüger H.-P., Boverie, S., & Flemisch, F. (2009). The importance of driver state assessment within highly automated vehicles. *16th ITS World Congress and Exhibition on Intelligent Transport Systems and Services*.

Rivera, J., Talone, A. B., Boesser, C. T., Jentsch, F., & Yeh, M. (2014). Startle and surprise on the flight deck: Similarities, differences, and prevalence. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1047–1051. doi: 10.1177/1541931214581219

Rosenthal, T. C., Majeroni, B. A., Pretorius, R., & Malik, K. (2008). Fatigue: An overview. *Am Fam Physician*, *78*(10), 1173–1179.

Rumelt, R. P. (2013). *Good strategy bad strategy: The difference and why it matters*. London: Profile Books Ltd.

Saccagno, A., Ritter, W., Mejuto, P., Schäufele, B., Tango, F., Schyr, C., ...

Waldschmidt, C. (2016). System performance assessment specification: Deliverable D5.1. *RobustSENSE*.

SAE. (2016). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles (surface vehicle recommended practice: Superseding J3016 Jan 2014).*

SAE. (2018). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles (surface vehicle recommended practice: Superseding J3016 Sep 2016).*

Saffarian, M., Winter, J. C. F. d., & Happee, R. (2012). Automated driving: Human-factors issues and design solutions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 2296–2300. doi: 10.1177/1071181312561483

Sagberg, F. (1999). Road accidents caused by drivers falling asleep. *Accident analysis and prevention*, *31*(6), 639–649. doi: 10.1016/S0001-4575(99)00023-8

Sahayadhas, A., Sundaraj, K., & Murugappan, M. (2012). Detecting driver drowsiness based on sensors: A review. *Sensors*, *12*(12), 16937–16953. doi: 10.3390/s121216937

Saper, C. B., Cano, G., & Scammell, T. E. (2005). Homeostatic, circadian, and emotional regulation of sleep. *The Journal of comparative neurology*, *1*(493), 92–98. doi: 10.1002/cne.20770

Schirmer, J.-E. (2017). Augen auf beim automatisierten Fahren! Die StVG-Novelle ist ein Montagsstück. *NVZ*, 253–257.

Schmidt, E., Decke, R., Rasshofer, R., & Bullinger, A. C. (2017). Psychophysiological responses to short-term cooling during a simulated monotonous driving task. *Applied ergonomics*, *62*, 9–18. doi: 10.1016/j.apergo.2017.01.017

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. doi: 10.1016/j.trf.2011.01.005

Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident analysis and prevention*, *41*(5), 1087–1093. doi: 10.1016/j.aap.2009.06.007

Schmidt, J. (2018). *Detektion der Reaktionsbereitschaft beim hochautomatisierten*

Fahren (Dissertation). Technische Universität Berlin, Berlin.

Schmidt, J., Braunagel, C., Stolzmann, W., & Karrer-Gauß, K. (2016). Driver drowsiness and behavior detection in prolonged conditionally automated drives. In *2016 IEEE intelligent vehicles symposium (IV)* (pp. 400–405). doi: 10.1109/IVS.2016.7535417

Schmidt, J., Dreißig, M., Stolzmann, W., & Rötting, M. (2017). The influence of prolonged conditionally automated driving on the take-over ability of the driver. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 1974–1978. doi: 10.1177/1541931213601972

Schmidt, J., Stolzmann, W., & Karrer-Gauß, K. (2016). Experimental evaluation of different request intervals for a driver alertness device for conditionally automated driving with induced drowsiness. In *AAET 2016 Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel* (pp. 97–117). Braunschweig: ITS Automotive Nord.

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. doi: 10.1016/j.promfg.2015.11.005

Schömig, N., & Kaussner, A. (2010). Evaluation of an attention monitor for drowsy driving. In A. Schieben & F. Flemisch (Eds.), *Validation of concept on optimum task repartition* (pp. 18–34). Deliverable D.33.6.

Shen, J., Barbera, J., & Shapiro, C. M. (2006). Distinguishing sleepiness and fatigue: focus on definition and measurement. *Sleep Medicine Reviews*, *10*(1), 63–76. doi: 10.1016/j.smrv.2005.05.004

Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). A comparison of the cell phone driver and the drunk driver. *Human factors*, *48*(2), 381–391. doi: 10.1518/001872006777724471

Teigen, K. H. (1994). Yerkes-dodson: A law for all seasons. *Theory & Psychology*, *4*(4), 525–547. doi: 10.1177/0959354394044004

Toyota Global Newsroom. (2008). *Toyota enhances pre-crash safety system with eye monitor.* Retrieved 01.10.2018, from `https://newsroom.toyota.co.jp/en/detail/281564`

Trivedi, M. M., Gandhi, T., & McCall, J. (2007). Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety. *IEEE Transactions on Intelligent Transportation Systems*, *8*(1), 108–120. doi: 10.1109/TITS

.2006.889442

van der Laan, J. D., Heino, A., & de Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*(1), 1–10. doi: 10.1016/S0968-090X(96)00025-3

Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2012). Trust in smart systems: Sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. *Human factors*, *54*(5), 799–810. doi: 10.1177/0018720812443825

Vogelpohl, T., Kühn, M., Hummel, T., & Vollrath, M. (2018). Asleep at the automated wheel - sleepiness and fatigue during highly automated driving. *Accident analysis and prevention*. doi: 10.1016/j.aap.2018.03.013

Volkswagen. (n.d.). *Driver alert system: Senses when you start to feel tired and need a break.* Retrieved from `https://www.volkswagen.co.uk/technology/car-saftey/driver-alert-system`

Wandtner, B., Schömig, N., & Schmidt, G. (2018). Effects of non-driving related task modalities on takeover performance in highly automated driving. *Human factors*, *60*(6), 870–881. doi: 10.1177/0018720818768199

Wang, P., Sibi, S., Mok, B., & Ju, W. (2017). Marionette: Enabling on-road wizard-of-oz autonomous driving studies. *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 234–243. doi: 10.1145/2909824.3020256

Warr, P. (2013). How to think about and measure psychological well-being. In R. R. Sinclair, M. Wang, & L. E. Tetrick (Eds.), *Research methods in occupational health psychology* (pp. 733–750). New York: Routledge.

Weinbeer, V., Baur, C., Radlmayr, J., Bill, J.-S., Muhr, T., & Bengler, K. (2017). Highly automated driving: How to get the driver drowsy and how does drowsiness influence various take-over aspects? *8. Tagung Fahrerassistenz*.

Weinbeer, V., Bill, J.-S., Baur, C., & Bengler, K. (2018). Automated driving: Subjective assessment of different strategies to manage drowsiness. In D. de Waard et al. (Eds.), *Proceedings of the human factors and ergonomics society europe chapter 2017 annual conference* (pp. 5–17).

Weinbeer, V., Muhr, T., & Bengler, K. (2019). Automated driving: The potential of non-driving-related tasks to manage driver drowsiness. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of* *the 20th congress of the international ergonomics association (IEA 2018)* (pp. 179–188). Cham: Springer International Publishing.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (Fourth edition ed.). London: Routledge.

Wiegand, D. M., McClafferty, J., McDonald, S. E., & Hanowski, R. J. (2009). Development and evaluation of a naturalistic observer rating of drowsiness protocol: Final report. *The National Surface Transportation Safety Center for Excellence*.

Wierwille, W. W., & Ellsworth, L. A. (1994). Evaluation of driver drowsiness by trained raters. *Accident analysis and prevention*, *26*(5), 571–581.

Wimmer, M. (2014). *Entwicklung und Erprobung von Mensch-Maschine-Systemen zur automatisierten Fahrzeugführung* (1., Aufl. ed.). Göttingen: Cuvillier Verlag.

Yang, Y., Klinker, J. N., & Bengler, K. (2019). How will the driver sit in an automated vehicle? - The qualitative and quantitative descriptions of non-driving postures (NDPs) when non-driving-related-tasks (NDRTs) are conducted. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th congress of the international ergonomics association (IEA 2018)* (pp. 409–420). Cham: Springer International Publishing.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. doi: 10.1002/cne.920180503

Yoshida, J. (2016). *12 in-car devices to watch drivers.* Retrieved from `https://www.eetimes.com/document.asp?doc_id=1329238`

APPENDIX A

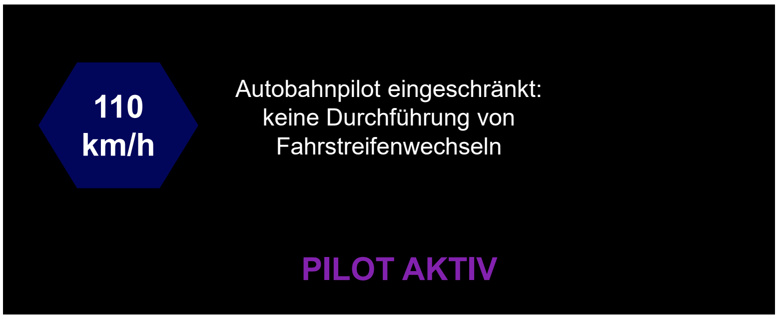
The German version of the drowsiness scale

Schläfrigkeitslevel und Indikatoren

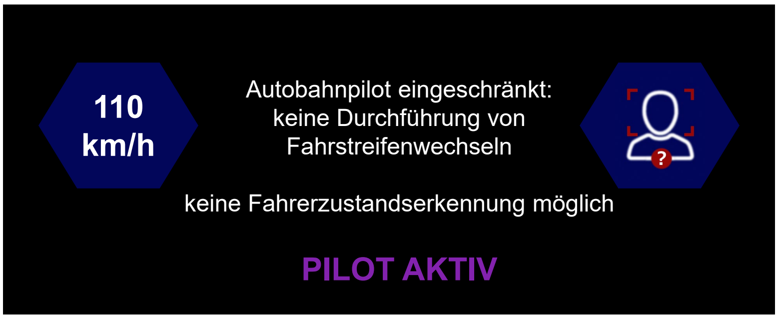
Schläfrigkeitslevel	Indikatoren
1 - *Nicht schläfrig*	Anwesenheit von Aufmerksamkeit; normal schnelle Lidschläge; normal häufige Blickwechsel; normaler Gesichtstonus; normal häufige Körperbewegungen/ Gesten (Wierwille & Ellsworth, 1994)
2 - *Etwas schläfrig*	Ausreichende Aufmerksamkeit vorhanden; weniger scharfe/aufmerksame Blicke; längere Fixationen; langsamere Lidschläge; erstes Auftreten von Manierismen (Augenreiben, Gesichtsreiben, Kratzen, Verziehen des Gesichts, unruhiges Umherrutschen im Sitz) (Wierwille & Ellsworth, 1994) und (Wiegand et al., 2009)
3 - *Mäßig schläfrig*	Auftreten von Manierismen; langsamere Lidschlüsse; starke Fixationen; glasige Augen; abnehmender Gesichtstonus (Wierwille & Ellsworth, 1994)
4 - *Deutlich schläfrig*	Lange Lidschlüsse (1-2s); Seitwärtsrollen der Augen; fehlende Fixation; seltene Lidschläge; geringer Gesichtstonus; nur vereinzelt Körperbewegung; bequeme Position im Sitz (Karrer-Gauß, 2012)
5 - *Sehr schläfrig*	Fehlende Aktivität; längere Lidschlüsse (2-3s); Augenrollen seitwärts und nach oben; keine richtige Fixation; verminderter Gesichtstonus; große isolierte oder punktuelle Bewegungen (Wierwille & Ellsworth, 1994)
6 - *Extrem schläfrig*	Lange Zeitspannen mit fehlender Aktivität; Sekundenschlaf; sehr lange Lidschlüsse (mehr als 4s); ruckartige Bewegungen beim Übergang in und aus dem Sekundenschlaf) (Wierwille & Ellsworth, 1994)

APPENDIX B

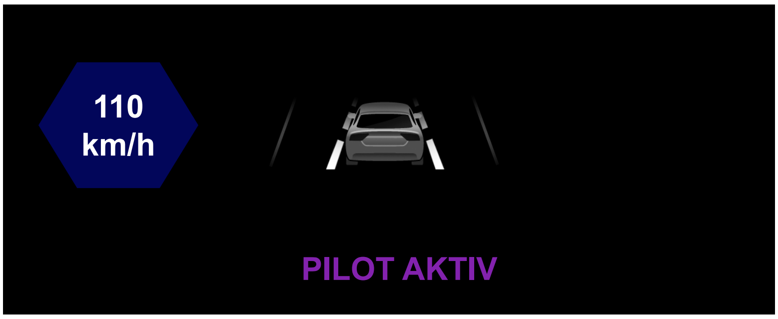
The German versions of the different feedback concepts



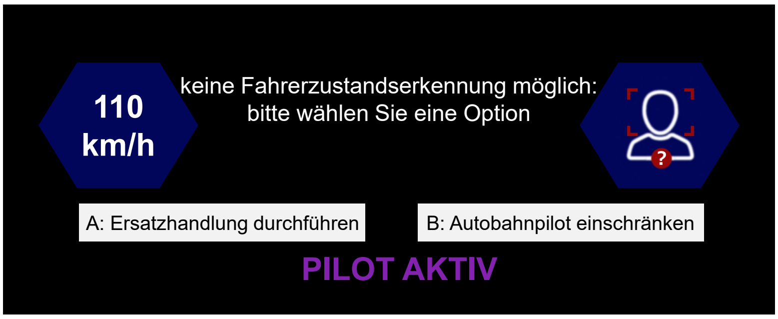
Concept A



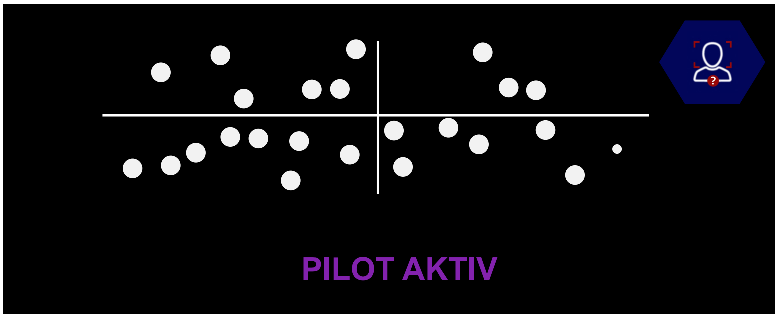
Concept B



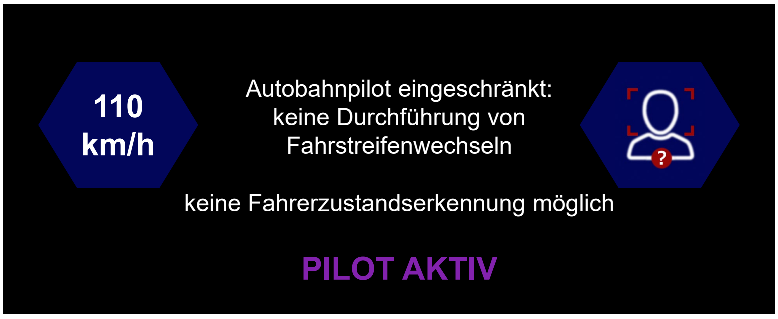
Concept C



Concept D



Concept D: Option A



Concept D: Option B