Mediator

Guidelines for monitoring degraded driver performance

Deliverable D4.3 – WP4 – Public

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

Guidelines for monitoring degraded driver performance

Work package 4, Deliverable D4.3

Please refer to this report as follows:
Ahlström, C., De Winkel, K.N., Larsson, K., De Goede, M., Cleij, D., Borowsky, A., Thalya, P., Karlsson, J., Fiorentino, A., Rauh, N. (2023). Guidelines for monitoring degraded driver performance, Deliverable D4.3 of the H2020 project MEDIATOR.

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

Deliverable details:	
Version: **Dissemination level:** **Due date:** **Submission date:**	Final PU (Public) 30/04/2023 30/04/2023

Lead contractor for this deliverable:
Christer Ahlström – Swedish National Road and Transport Research Institute

Report Author(s):	Ahlström, C., Larsson K. – National Road and Transport Research Institute, Sweden De Winkel, K.N., De Goede, M., Cleij, D. – SWOV Institute for Road Safety Research, The Netherlands Borowsky, A. – Ben-Gurion University of the Negev, Israel Thalya, P. – Veoneer Sweden AB, Sweden Karlsson, J. – Autoliv Research, Sweden Fiorentino, A. – Fiat Chrysler Automobiles, Italy Rauh, N. – Chemnitz University of Technology, Germany

Revision history

Date	Version	Reviewer	Description
04/03/2023	Preliminary draft	Christer Ahlström – VTI	Task internal review
27/03/2023	Draft for review	Alessia Knauss – Zenseact	Internal QA review
18/04/2023	Draft for review	Linda Boyle – University of Washington	External QA review
30/04/2023	Final draft	Elmer van Grondelle (WP4 leader) – Delft University of Technology Ernst Verschragen (QA officer) – SWOV Institute for Road Safety Research	Final editing and checks
30/04/2023	Final deliverable	Michiel Christoph (Technical coordinator) – SWOV Institute for Road Safety Research	Approved

Table of contents

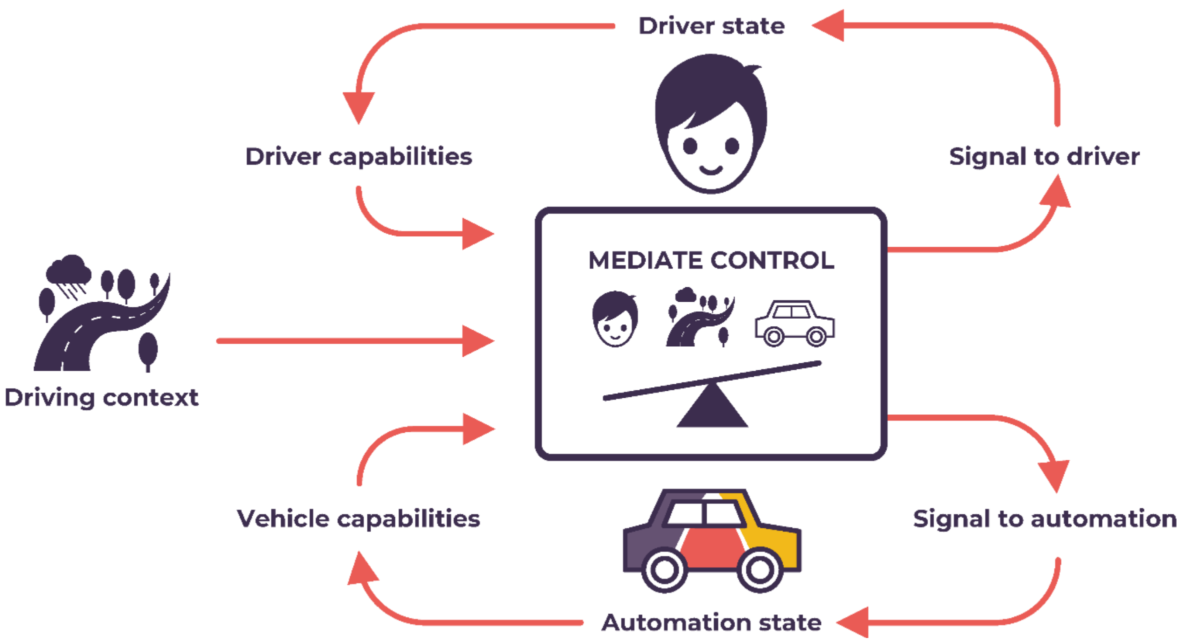
About MEDIATOR

MEDIATOR, a 4-year project coordinated by SWOV Institute for Road Safety Research, has come to an end after four years of hard work. The project has been carried out by a consortium of highly qualified research and industry experts, representing a balanced mix of top universities and research organisations as well as several OEMs and suppliers.

The consortium, supported by an international Industrial Advisory Board and a Scientific Advisory Board, represented all transport modes, maximising input from, and transferring results to aviation, maritime and rail (with mode-specific adaptations).

Vision

Automated transport technology is developing rapidly for all transport modes, with huge safety potential. The transition to full automation, however, brings new risks, such as mode confusion, overreliance, reduced situational awareness and misuse. The driving task changes to a more supervisory role, reducing the task load and potentially leading to degraded human performance. Similarly, the automated system may not (yet) function in all situations.



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

The MEDIATOR project aimed to develop an in-vehicle system, the Mediator system, that intelligently assesses the strengths and weaknesses of both the driver and the automation and mediates between them, while also taking into account the driving context. It assists the timely take-over between driver and automation and vice versa, based on who is fittest to drive. This Mediator system optimises the safety potential of vehicle automation during the transition to full (level 5) automation. It would reduce risks, such as those caused by driver fatigue or inattention, or on the automation side by imperfect automated driving technology. MEDIATOR has facilitated market exploitation by actively involving the automotive industry during the development process.

To accomplish the development of this support system MEDIATOR integrated and enhanced existing knowledge of human factors and HMI, taking advantage of the expertise in other transport modes (aviation, rail and maritime). It further developed and adapted available technologies for real-time data collection, storage and analysis and incorporated the latest artificial intelligence techniques. MEDIATOR has developed working prototypes, and validated the system in a number of studies, including computer simulation, virtual reality, driving simulator and on-road studies.

With MEDIATOR we further paved the way towards safe and reliable future vehicle automation that takes into account who is most fit to drive: the human or the system.

https://mediatorproject.eu/

Mediator

Executive summary

The objective of this work is to describe guidelines for measuring degraded human performance based on driver state and competences from a real-time driver monitoring perspective. The guidelines integrate state-of-the-art knowledge from the literature with knowhow from the industry and practical results from the Mediator project. The formulated guidelines are defined based on functionality, technological possibilities, safety relevance and feasibility.

In summary, an ideal driver monitoring system should have the following general features:

- *Minimally obtrusive sensors*. Camera-based systems have several advantages here, since they have the potential to capture rich information about humans, objects, and their interaction. Unobtrusive sensing is needed to facilitate high adoption rates, to avoid deactivation, and to avoid interfering with drivers' operation of the vehicle.
- *Real-time operation and timely detections*. Impairment detection, and subsequent interventions, have different demands on acceptable latencies. Detection of early signs of fatigue is not time critical (order of minutes) while severe fatigue, microsleep, and long off-road glances are time critical (order of seconds or less). In some situations, discomfort can be detected offline several minutes in advance, for example when approaching harsh weather or a traffic jam. Proactive impairment interventions, in contrast to reactive detection/intervention, is favourable. This requires forecasts of drivers' future readiness levels.
- *Robustness to environmental conditions.* System performance should not be significantly influenced by environmental conditions such as traffic, landscape, weather, and darkness.
- *Automation level dependent*. The drivers' responsibilities change with the level of vehicle automation, which in turn affects the requirements for a driver monitoring system. As an example, continuous distraction detection is highly relevant in manual and assisted driving. In higher levels of automation, where non-driving related task engagement is allowed, it is sufficient to ensure that the driver is attentive in relation to transitions of control.
- *Situational awareness*. A driver/vehicle-unit should have sufficient situational awareness to be able to drive safely. With higher levels of automation, the responsibilities for situational awareness are gradually shifted from the human to the vehicle. Similarly, to be able to provide relevant impairment detections, a driver monitoring system should also be situationally aware and take contextual factors into account. For example, fatigue warning systems would benefit from knowledge about sleep history and driving time, and distraction detection systems would benefit from knowledge about which areas in the surroundings that needs to be sampled to gain sufficient situational awareness.
- *Ecological validity*. Final evaluations/testing of driver monitoring systems should be conducted in ecologically valid settings with naturalistically induced impairments. Lab testing can and should be used in earlier evaluation stages, for example, when testing if an eye tracking system provides high quality tracking throughout a broad range of the population.
- *Minimal intrusion on privacy*. Driver monitoring systems should avoid privacy intrusions. For example, video data should be deleted continuously and should not be stored beyond what is needed for impairment detections.

Since MEDIATOR has focused on driver distraction, driver fatigue, and driver comfort, the guidelines are restricted to these three states. In addition to comfort, distraction, and fatigue, many researchers, legislators, developers, and users also mention sudden sickness and intoxication as important impairments. These impairments have therefore been covered as well in the state-of-the-art review and in the interviews, but not in the actual guidelines.

The goal of driver monitoring is to increase road safety. Achieving this goal depends not only on the performance of the driver monitoring system, but also on the intervention strategy and how the intervention is communicated to the driver. Guidelines on intervention strategies established in the Mediator project are described in van Grondelle (2023).

Though the formulation of definitive operational guidelines for driver monitoring systems still suffer from a lack of knowledge, this should not prevent or delay the introduction and implementation of such systems. Instead, available technologies should be used to address and mitigate impairments to the extent possible, starting with severe behaviours such as incapacitation, alcohol intoxication, microsleeps, and long glances away from the road.

1. Introduction and general considerations

Driving consists of a complex interaction of situational anticipation, information sampling, decision making, and action, where the driver interacts with the environment, with other road users and with the vehicle. The driving process is coordinated by complex interactions encompassing operational, tactical, and strategical abilities. If one or more of these elements are degraded in their functioning, the driver may be less safe and, thus, be unfit to drive. Driver state related factors that negatively impact performance are for example distraction, fatigue, intoxication, and sickness. Mood, emotions and (dis-)comfort can also have a negative safety impact, where anger, distress and discomfort can trigger unnecessary or unexpected actions by the driver. Since human errors and driver impairments are the leading contributors to vehicle collisions (Singh, 2015), active safety systems and vehicle automation has been put forth as the final fix to the shortcomings of human drivers.

Bainbridge (1983) explained that “the more advanced a control system is, so the more crucial may be the contribution of the human operator”. Similarly, Parasuraman and Riley (1997) explained how humans often misuse, disuse, and abuse automation technology, and also argued that humans tend to be poor supervisors of automation. Indeed, we now see how (partial) automation can lead to driver fatigue, disengagement, reduced active participation, increased engagement with non-driving related tasks, increased workload, and deskilling (Dunn et al., 2021; Feldhütter et al., 2019; Greenlee et al., 2019; Llaneras et al., 2013; Noble et al., 2021; Noy et al., 2018).

Fully automated vehicles that can drive from A to B with no human intervention are yet not ready to be deployed on public roads. Meanwhile, a responsible driver must always be present, to monitor the performance of partial automation, or be ready to operate the vehicle in conditions not supported by the automation. Since humans are poor supervisors of automation (Parasuraman & Riley, 1997), another technical safety layer has been suggested, where the human driver is continuously monitored to ensure a sufficient level of fitness (Hayley et al., 2021; Hecht et al., 2018). An issue here is that driver monitoring is difficult and available systems are facing challenges that are not easily overcome (Doudou et al., 2019; Koay et al., 2022; Koesdwiady et al., 2017; Ortega et al., 2022; Perkins et al., 2022). Examples of why driver monitoring is difficult include inter- and intra-individual differences and context dependence, and especially in an automated driving setting, the need to predict the driver’s readiness to re-engage within the automated systems warning timeframe.

The early work of Bainbridge (1983) clearly shows that new human factors related issues arise when automation is introduced. At the heart of the problem lies the fact that the things that can be automated are not necessarily the things that should be automated. It is important that we do not end up in an analogous situation with driver monitoring systems. By setting up requirements and guidelines based on actual needs rather than what is currently possible to measure, it might be possible to steer driver monitoring development in a direction where it makes a true difference for traffic safety.

2. Objective and framework

The objective of this deliverable is to describe guidelines for measuring degraded human performance from a driver monitoring perspective. The main input comes from the state-of-the-art review that was carried out early in the project (Borowsky et al., 2020), from results on practical validity derived from MEDIATOR experiments (Athmer et al., 2023; Borowsky et al., 2023; Fiorentino et al., 2023), and from interviews with representatives involved in driver monitoring system development. These findings are then integrated to guide future driver-state and driver-competence detection.

A key component of the Mediator system is the driver monitoring system that is used to continuously monitor the abilities of the driver. Along with the automation monitoring system, these two components provide the information needed for the decision logic component to mediate control between the driver and the vehicle. In this deliverable, the proposed Safe by Design heuristic by Jannusch et al. (2021) will be adapted and used as a framework to identify abilities and limitations of driver monitoring systems, considering various types of driver impairments as well as different levels of automation.

Driver fitness assessment in MEDIATOR focus on monitoring of *Distraction* and *Fatigue*. The project has also focused on *Comfort* since driver comfort is, next to safety, efficiency, social inclusion and accessibility, one of the main drivers of higher driving automation levels (ERTRAC, 2022). The guidelines in chapter 4 encompass these three states. The preparatory work and the background information in chapters 2 and 3 have a wider scope, also covering *sudden sickness* and *intoxication*, but to a lesser extent.

2.1. Driver impairment definitions

Fatigue is here defined as a biological drive for recuperative rest (Williamson et al., 2011). Within MEDIATOR, the scope of the term encompass fatigue due to extended periods of high or low workload (task-related fatigue), and accumulated sleep debt, prolonged wakefulness, or troughs in the circadian rhythm (sleep-related fatigue). From a fatigue detection perspective, all forms of fatigue can be measured using the same methods. However, there are large differences in which countermeasures to deploy if fatigue is detected (May & Baldwin, 2009). Task related fatigue caused by high task load will benefit from increased automation and in-vehicle technologies that offset driver workload. Conversely, underload caused by monotonous conditions and highly familiar roadways should be countered by increasing the novelty and demand of the driving task. Sleep related fatigue can only be countered by recuperative sleep.

Driver inattention can be defined as insufficient or no attention to activities critical for safe driving (Regan et al., 2011) or as when insufficient information is sampled to be able to form and maintain a mental representation of the situation (Kircher & Ahlstrom, 2016). Whereas all forms of inattention are potentially detrimental to driver fitness, the focus within MEDIATOR is primarily on visual distraction, operationalized as looking away from the road for too often or for too long.

Ensuring a comfortable and positive driving experience is considered a fundamental prerequisite for the acceptance and usage of automated functions (Bellem et al., 2018). In MEDIATOR, the term comfort encompasses traditional comfort aspects such as noise, vibrations or sitting comfort, as well as additional factors such as apparent safety, motion sickness, trust in the system, controllability, familiarity of vehicle operations, and transparency of system states and actions (Beggiato, 2015; Elbanhawi et al., 2015).

2.2. Driver monitoring systems

The purpose of driver monitoring systems is to assess driver fitness. This is a complex construct that depends on various psychological and physiological states. Major contributing factors, as indicated by research, policy makers and the industry alike, are inattention, fatigue, intoxication, and sudden sickness. These factors are *latent constructs*, referring to the fact that they reflect theoretical aspects of the human psychological/physiological state that are not directly observable, but must be inferred from manifest, observable variables. The operationalization of each construct, as well as their interrelations and relation to safety, are matters of active investigation. Nevertheless, to visualize the relations addressed in this deliverable and to facilitate the formulation of guidelines, a simplified outline of these relations is made in Figure 2.1. Variables represented in circles are latent constructs and variables in squares are observable manifestations of these constructs that can be exploited by a driver monitoring system.

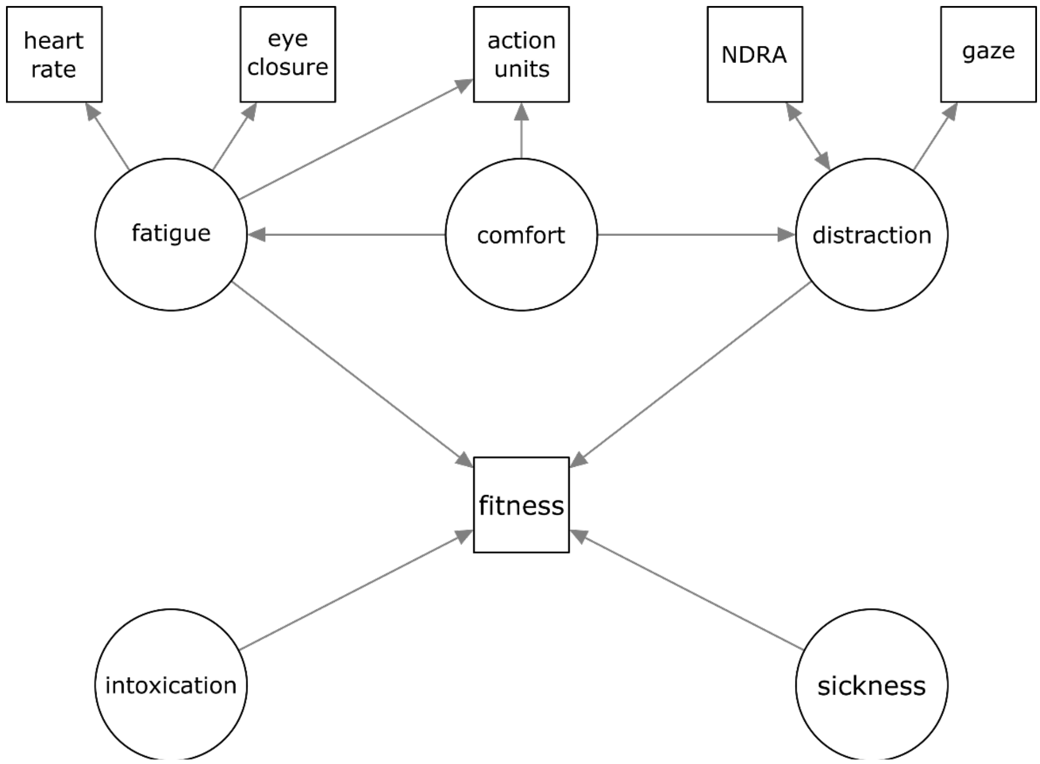


Figure 2.1: Relations between latent factors (circles) and manifest, observable factors (squares), as they are considered in this deliverable. Fitness depends on various latent constructs that reflect certain driver states. Within MEDIATOR, the factors fatigue, distraction, and comfort were monitored. In the project, fatigue was inferred from heart rate and camera data, distraction was inferred from engagement in non-driving related activities (NDRA) and gaze direction and comfort from facial action units. Intoxication and (sudden) sickness are only considered in the literature review.

2.3. Driver monitoring system ratings and regulations

Official regulations and guidelines for driver monitoring systems have started to emerge but are typically limited to specific automation systems. *Regulations* are system design and performance evaluation criteria that are mandatory for a vehicle to gain type approval. Some regulations are already in place, but these are specific to certain assistance functions and note that driver monitoring is required, without specifying details. For instance, UNECE regulation 157 on Automated Lane Keeping Systems (addendum to 1958 Agreement on harmonization) states that “The system shall comprise a driver availability recognition system”, and that “The system shall detect if the driver is attentive.” It specifies that this shall be achieved, for instance, by detecting if the driver is present in a driving position, if the safety belt of the driver is fastened, and if the driver is available to take over the driving task. However, details on the actual implementation of the system and evaluation of its performance are not provided. Similarly, the EU General Safety Regulation 2019/2144 requires driver availability monitoring systems in automated vehicles on all new type approvals from July 2022. Further, the EU General Safety Regulation 2019/2144 states that "Motor vehicles shall be equipped with … driver drowsiness and attention warning systems, and advanced driver distraction warning systems”, but does not specify how this should be implemented. Some more specific criteria are provided in related regulation EU 2021/1341 on Driver Drowsiness and Attention Warning systems, where it is suggested that drowsiness may be detected from control inputs and computer vision systems, although manufacturers are left free to choose any specific implementation. Proof must be provided that the implemented system meets minimum requirements in terms of correct classifications (i.e., *sensitivity*) of drowsiness, on the basis of empirical data collected from human participants while following a proposed test-protocol. An update to the EU general safety regulations that targets distraction (Advanced Driver Distraction Warnings) is currently planned for 2024.

Euro NCAP will develop an assessment protocol (foreseen for 2023) that considers driver monitoring (Occupant Status Monitoring systems) for fatigue and distraction. Because the Euro NCAP rating system is indicative of the safety features augmenting minimal requirements through regulations, these may be considered *guidelines*. The assessment by Euro NCAP will revolve around (i) how reliably and accurately the status of the driver is detected and (ii) what action the vehicle takes based on the information.

2.4. Automation aspects of driver monitoring

There are different ways to define the capabilities and responsibilities of an automated vehicle. The commonly referred to standard J3016 suggests six levels of driver assistance technology (SAE, 2021). To understand their structure, it is important to know that automated vehicles are assumed to operate only in a pre-defined situation/environment. This environment is called the systems’ Operational Design Domain. Level 0 equals unassisted *manual* driving. Levels 1–2 are *assisted* driving where the human driver still is responsible. Levels 3 – 4 represents *piloted* driving where the automated system is responsible within a specific domain and a human driver is responsible for all driving outside this domain. Level 5 is robot taxi; no driver involvement is needed at any point.

MEDIATOR addresses automation on SAE levels 0 – 4, using the terminology defined in Table 2.1. A key point within MEDIATOR has been to adopt a user perspective on automation. Where SAE automation levels align with technical possibilities of automation, MEDIATOR automation levels are based on the driver’s responsibilities and affordances. To illustrate, whereas SAE level 4 represents a level of automation that allows a driver to be out of the loop and that also ensures safe handling of situations where the automation cannot adequately perform the driving task, it does not consider how long one can be out of the loop. In MEDIATOR, the Time-to-Sleep mode is defined from a user perspective: it considers whether the driver can stay out of the loop for a short while or for a long time.

Table 2.1: Automation levels addressed in MEDIATOR (OEDR: Object and Event Detection and Response)

	driver supported			automated driving		
SAE	0	1	2	3	4	5
Automation responsibilities	warnings and momentary assistance	lateral or longitudinal support	lateral and longitudinal support	automated functions drive the vehicle within the defined operational design domain		automated driving under all conditions
Human responsibilities	driver must constantly supervise			driver is not required to drive, but must take over upon request		driver is a passenger
Euro NCAP		Assisted (shared control)		Automated (vehicle in control)		Autonomous
Automation responsibilities		OEDR and other supportive tasks		OEDR and driving. Vehicle has full responsibility		full control
Human responsibilities		OEDR and driving. Driver is fully responsible. No safe transfers		Driver can do non-driving related tasks, but must take over upon request		driver is a passenger
Mediator		Continuous mediation		Driver standby	Time-to-Sleep	
		drivers supported by automation but are responsible and must monitor surroundings and automation.		driver must take back control upon request (order of seconds)	driver must take back control upon request (order of minutes)	
HMI	Manual	Assisted		Piloted		
	non-automated, driver is in full control	drivers are not fully disengaged and must maintain certain responsibilities. This can be steered towards a monitoring task.		drivers monitor while automation performs driving tasks		

2.4.1. Implications of automation on distraction

The purpose of automation is to improve safety and comfort for a vehicle's driver and passengers. One of the ways vehicle automation may contribute to this goal is by reducing driver workload. However, improvident minimization of workload may paradoxically induce a state of cognitive underload, which causes both the mind to wander (Körber et al., 2015), and increases the tendency to engage in non-driving related tasks (Solís-Marcos et al., 2018); thereby actually provoking distraction and reducing situational awareness (Saxby et al., 2013). Moreover, vigilance tasks are known to be particularly demanding and may therefore also promote fatigue (Miller et al., 2015).

Drivers experienced with Continuous mediation engage in non-driving related tasks more frequently (Dunn et al., 2021). They also spend more time with their eyes off the forward roadway when driving automation systems are active, with more frequent and longer duration non-driving-related task glances (Noble et al., 2021). These results suggest that drivers trust Continuous mediation systems to compensate for their distracted driving behaviours.

2.4.2. Attention requirements on the driver per automation level

For consistency within MEDIATOR, we here consider automation modes as defined within the project (see Table 2.1). However, in cases where these modes cover different aspects of the driving task as per SAE levels, we will consider whether and how these differences affect attentional requirements and affordances for the driver.

2.4.2.1. Continuous mediation

As per Table 2.1, the MEDIATOR Continuous mediation automation mode covers SAE levels 1 and 2. These levels differ in terms of the extent to which the driving task is transferred to the vehicle, and thereby also differ in the attentional requirements for the driver. In absence of automation (i.e., manual driving, or SAE level 0), a driver is in active control of the vehicle. The driver is required to provide lateral and longitudinal control of the vehicle, and to continuously monitor for potential hazards. This involves perception and recognition of road conditions as well as the density and behaviours of other road users, and projection, i.e., prediction of situations that may unfold on the basis of these traffic states and knowing how to properly react to them (Fisher et al., 2016; Horswill & McKenna, 2004; Yamani et al., 2016). Occasional monitoring of vehicle state variables via information systems is allowed and required to maintain situational awareness. Drivers are not allowed to engage in activities other than driving, except for minimally obtrusive acts such as hands-free calls and adjustments of vehicle settings. For low levels of automation, the vehicle takes care of the lateral or longitudinal control (SAE level 1) or both lateral and longitudinal control (SAE level 2) for the driver. Under these conditions the drivers are alleviated from performing these control tasks themselves, but they are still required to continuously monitor the performance of the automation, along with performing their main task of monitoring the road environment for potential hazards as well as monitoring vehicle state. Thus, (part of) the control task is transferred to the vehicle, whereas the driver is required to remain vigilant and situationally aware, and must always be able to take immediate control (de Winter et al., 2016). Because advanced automation functionalities within the Continuous Mediation mode increasingly convert the driving task into an attentionally demanding vigilance task, this mode may set the most stringent requirements in terms of driver monitoring system, as automation should never be detrimental to safety.

2.4.2.2. Driver standby

The Standby mode is the first level of automation where the driver is completely alleviated from the driving task, albeit only for unforeseeable amounts of time – drivers must therefore maintain sufficient situation awareness and be ready to take over the driving task from the automation within a few seconds' notice. Challenges here are related to regaining driver fitness and balancing the time until either the automation or the driver becomes (un)fit, making sure always one is fit enough for the driving task.

2.4.2.3. Time-to-Sleep

In the Time-to-Sleep mode, the driver is completely alleviated from the driving task given that the vehicle is within the automated systems operational design domain (such as highway driving). This implies that the amount of time the driver is not required to intervene can be confidently estimated. In other words, drivers can be safely out of the loop for long periods of time and truly immerse themselves in non-driving related tasks. Drivers are thus free to engage in any activity the environment allows and are even free to fall asleep. Consequently, driver distraction monitoring systems are not required, except to make sure that the driver is attentive when control is transferred back to the driver. Challenges in this level of automation are thus to bring the driver back into the loop after complete disengagement and to predict when this will be required sufficiently long in advance for drivers to be ready to take over control when they are required to do so. This is primary related to fatigue, but also to the type of activity that the driver may be immersed in. Transfers of control should allow enough time to safely put away computers, books, food etc.

2.4.3. Implications of automation on fatigue

The level of automation has different implications on fatigue (Ahlström et al., 2023). In manual driving (SAE level 0), fatigue commonly arises during night-time or in the early morning hours. It can also appear after too many uninterrupted hours behind the wheel or after extended periods of high or low workload (Williamson et al., 2011). Assisted driving (SAE levels 1–2) has the potential to reduce fatigue caused by high workload, at least to the extent the driving task itself is causing the overload. However, since the driver in that case shifts from driving to monitoring, levels of fatigue may then instead increase due to boredom and/or exhaustive attentive monitoring without an active task. With more sophisticated and reliable driving automation, it thus becomes harder for a human driver to maintain the vigilance needed to monitor both automation and roadway (Bainbridge, 1983; Carsten & Martens, 2019; Noy et al., 2018).

An effective countermeasure for task-related fatigue is to do something else for a while. In driving, this becomes feasible during piloted driving (SAE levels 3–5) and can be used as a countermeasure to overload (relieve the driver from the driving task) as well as to underload (allow the drier to rest for a while). However, in SAE level 3, the human must remain alert enough to be able to resume manual driving with short notice. The real game changer from a fatigue point of view is therefore when SAE levels 4 and 5 systems are introduced. These will allow a travelling human to sleep, recover and recuperate while on the move. As such, high-level automation will facilitate a truly effective countermeasure for fatigue, given that sleep inertia issues can be managed.

2.4.4. Alertness requirements on the driver per automation level

2.4.4.1. Continuous mediation

In Continuous mediation, the requirements on the driver are the same as when driving manually, i.e., to always be alert.

2.4.4.1. Driver standby

In Driver standby, a takeover request may come with short notice, so the requirements on the driver are like when driving manually. The driver has no responsibilities of the driving while in automated mode. The short take over time (several seconds) allows rest and recovery, which can be beneficial in terms of task related fatigue. However, it does not allow for sleep and recuperation since it takes too long to get back in the loop.

2.4.4.1. Time-to-Sleep

In Time to Sleep, where the take over time horizon is longer (several minutes), the requirements on the driver are relaxed. Here, the only constraint is that the driver must be able to resume control of the vehicle with a few minutes notice. This allows for true recuperation via actual sleep. Note that the requirements on the vehicle increase as the requirements on the driver are relaxed. The vehicle must not only ensure automation fitness within predefined operation design domain, but it must also verify that the driver is attentive and physically positioned to drive before the actual transfer of control takes place. The latter entails a wakeup procedure, if needed, and a safe stop procedure, if the wakeup call is not enough.

In full automation, no fatigue detection is needed. Here the driver is never in control, and becomes, in effect, a passenger of the vehicle.

2.4.5. Implications of automation on driver comfort

Ensuring a comfortable and positive driving experience is a prerequisite for the acceptance and usage of automated functions. Driver comfort is also a selling point for vehicle manufacturers as it improves well-being and user satisfaction. By having less to no control of the driving task, automated vehicle users will be less able to predict upcoming driving manoeuvres and associated comfort relevant parameters (e.g., distances to other vehicles, acceleration rates), which might result in in an unpleasant automated driving experience. Thus, next to traditional comfort aspects such as noise, vibrations and sitting comfort, additional comfort aspects are discussed in automated driving settings, such as apparent safety, motion sickness, system trust, controllability, familiarity with vehicle operations as well as mode awareness (Beggiato, 2015; Domeyer et al., 2019; Elbanhawi et al., 2015). As such, maintained driver comfort is an important ingredient when setting up a functional driver/vehicle unit collaboration, where both are aware of each other's limitations, strengths and current states and can act accordingly (ERTRAC, 2022; Klien et al., 2004).

It is assumed that automated driving has the potential to be more efficient, more environmentally friendly, and safer than manual driving. Specifically, it is expected to decrease traffic congestions, fuel consumption and emissions as well as the number of crashes by compensating for human errors (ERTRAC, 2022). However, to exploit these potentials, users need to accept and activate driving automation as often as possible. Based on driving fitness, comfort and preferences in a certain context, the vehicle could be set up to actively propose automation features as a service to the driver. An example use case that has been investigated in MEDIATOR is when approaching and driving in traffic jams, a scenario which is associated with frequent rear-end crashes, and where drivers would generally prefer automation instead of driving manually (see Mediator Deliverable D1.2; Borowsky et al., 2020). To exploit these potential safety and comfort benefits of automation in (and even before) reaching traffic jams, the driver/user needs to be aware that such a situation is upcoming and whether automation is available at this specific moment. This is a feature that Mediator could offer by actively proposing automation features “at the right time”.

Next to comfort and safety aspects, the active proposal of vehicle automation features could tackle the problems that a great proportion of drivers are unaware of having advanced driver assistance systems in their vehicles, and that they disuse/misuse or demonstrate misperceptions about what the system can and cannot do (McDonald et al., 2018). Therefore, potential safety effects of assistance features are diminished. In addition, unnecessary interventions by the driver due to uncomfortable or unexpected vehicle operations (e.g., if apparent safety is perceived as compromised) could lead to safety-critical and unnecessary takeover situations (Hergeth et al., 2017; Techer et al., 2019). Detection of comfort issues could thus allow for adapting automation features such as driving style aspects (such as distance to vehicle ahead and lateral position) and/or information presentation with the overall aim to prevent disengagement of automation or dangerous and unnecessary takeover situations. Because driving comfort is primarily related to dynamic driving situations, constant comfort evaluation is necessary to prevent discomfort.

3. Capabilities and limitations of driver monitoring systems

This chapter describes capabilities and limitations of driver monitoring systems. The content is based on the state-of-the-art review conducted early in the project and results from the project experiments. To make sure that the information is up to date, the state-of-the-art has been summarized and updated. A series of interviews/discussions with representatives from the industry (driver/occupant monitoring companies, tier 1 suppliers, original equipment manufacturers, and test organisations) was also conducted to account for some of the hidden knowledge that is sometimes unaccounted for in the scientific literature.

3.1. Results from the literature

Driver monitoring systems are being deployed at an increasing rate worldwide, with installation rates estimated to increase from 1% in 2019 to 71% by 2026 (Barnden, 2019). This rapid increase is driven by regulatory bodies and testing organisations, and by the increasing automation of new vehicles and the need to ensure the safety of both driver and surrounding road users. Driver monitoring has an importand role to play for managing the human-machine interface and to ensuring operator engagement and safety during (automated) driving, but such applications areas require systems that are robust, reliable, and accurate.

Hermens (2020) reviewed about 100 fatigue and distraction detection systems. Each system was rated on eight criteria: validity, intrusiveness, availability, robustness, sustainability, acceptability, cost, and compatibility with other devices in the vehicle or used by the driver. Her conclusion was that no single system stands out on all criteria, and for sufficient monitoring of fatigue and distraction, a combination of systems or system features is needed. Adding alcohol intoxication would likely add yet another system.

Of the eight criteria, validity and intrusiveness/acceptability are essential for meaningful driver monitoring. The system must be used by the drivers, and it must be capable of measuring the impaired state sufficiently well. In this respect, it is not possible to rely on a system requiring the driver to wear for example a pair of glasses or a head band with dry electrodes. Neither is it future proof to rely on vehicle control measures such as lateral variability since modern assistance systems and/or automated functions often affect or even take over lateral and longitudinal control of the vehicle. For the application at hand, these constraints effectively limit the range of plausible driver monitoring systems to those based on direct and unobtrusive sensors.

There are some general limitations to minimally intrusive systems such as camera-based solutions:

- Lighting conditions: There are frequent and sudden variations of lighting in real-life driving. These changes happen quickly and depend on daytime (day/night), weather, driving environment (streets lined with trees, driving under a bridge) and artificial light (headlights, street lighting). Infrared cameras with corresponding illumination mitigate some of the issues, but not all. For example, squinting in strong sunlight causes the lower eyelid and eyelashes to occlude the pupil partially or fully, making it difficult to track the eyes.
- Camera view occlusions and hand deformation: Various facial occlusions can occur due to face masks, hoods, hand activities, glasses, and phones or laptops. The latter is especially relevant in vehicles with automated driving functionalities. Also, detection and tracking of

hands for recognition of driver distraction actions is a challenging problem. By including such garments/appearances in the training set when developing the systems, driver monitoring suppliers now manage such situations quite well, but certain lenses and IR-blocking sunglasses are still problematic.

- Variety of people: The system should work on all individuals throughout several ethnicities, genders, and age ranges. Developers are trying to take bias out of face and eye tracking systems, and it is working quite well, but nonconformity to stereotypes makes it difficult to cover all corner cases.

Driver monitoring research can essentially be subdivided in three branches – a sensor branch, a feature extraction branch, and a driver state assessment branch.

The sensor branch aims to provide high quality data for the latter stages. This includes minimally obtrusive physiological measurement sensors based on optical, capacitive sensors, magnetic induction, radar, laser, pressure sensors and strain-gauge sensors (Leonhardt et al., 2018), and behavioural measurement sensors typically based on cameras in the visible light, near-infrared light or far-infrared light frequency range, including depth sensing and stereo-cameras. The wavelength of a camera-based driver monitoring system is typically either 850nm or 940nm. While both can be used in darkness, the 940-nm wavelength is often preferred due to its invisibility to the human eye and its fewer interferences from the natural environment. Solar IR levels at 940nm are less than half compared to 850nm due to atmospheric absorption (Li et al., 2022). It is also common to predict the driver's state using vehicle measures based on especially lateral control (Liu et al., 2009), but since MEDIATOR focus on vehicle automation, this aspect will not be included in this report.

The feature extraction branch aims to extract vital signs, physiological indictors, facial features, head pose, gaze direction, eyelid opening or non-driving related activities (NDRA) from data/video streams. This branch of research is prospering thanks to recent achievements in machine learning, including deep learning, and the availability of open-source algorithms and pre-trained neural networks for facial feature detection and gaze estimation (Wang et al., 2018). This has led to improved eye/head tracking performance and higher detection rates of non-driving related activities such as mobile phone usage, with mean accuracies well above 90% (Kashevnik et al., 2021).

Temporal context can improve detection performance compared to frame-based methods, by exploiting temporal information and frame content from several sequential frames (Moslemi et al., 2021). This step typically encompasses a few frames and should not be confused with longer-term dependencies in the data, such as the process of slowly becoming more fatigued, or of engaging and disengaging from the driving task while performing a non-driving related activity (Lee, 2014).

The third branch, driver state assessment, will be covered in the upcoming sections.

3.1.1. Driver distraction monitoring

Technically, classifications of visual distraction can be obtained via the following process: (1) obtaining sensor readings, i.e., camera images, (2) a feature-extraction stage where image processing is performed to obtain estimates of gaze direction and recognition of NDRAs, and subsequently (3) a classification step where the obtained information is passed to the algorithm that uses the information to generate actual classifications of driver distraction. The first attention monitoring or distraction detection algorithms were based on the notion that as soon as the gaze is directed away from forward or the driver is engaged in a non-driving related task, this immediately leads to “distraction”. This goes hand in hand with the driver distraction definitions stating that a shift of attention to anything not relevant for driving immediately equals distraction, regardless of the outcome of the situation (Foley et al., 2013). As many developers have realised, these demands on the driver are unreasonably strict and lead to many false distraction detection events. Real-world real-time algorithms therefore give the driver some leeway by granting a certain amount of looking-away time. In practise, this is implemented by measuring the time spent looking away from the forward roadway. Common thresholds for the time that it is allowed to look away are for example 3 seconds as suggested in Euro NCAP’s safe driving assessment protocol (Euro NCAP, 2022). Given that 3 seconds can be a very long time, depending on the driving context, these thresholds are clearly set to strike a balance between false alarm rate and detection accuracy. There is no explicit theoretical motivation why 3 seconds is a good choice. A theoretically more appealing threshold would be to let it vary with situational complexity.

The fact that attention monitoring or distraction detection algorithms do not take situational complexity into account is a general concern. And it does not only apply to the temporal threshold. It also applies to the definition of “away from forward”. Typically, this region is defined as a static area of interest defined in a coordinate system that is fixed within the interior of the vehicle. This fixed definition of straight ahead is problematic since it does not allow the driver to look sideways when going through intersections, or to monitor the far end of sharp curves. Minor workarounds have been suggested, such as in the AttenD algorithm (Ahlstrom et al., 2013), which has a built-in mechanism for acknowledging the necessity of mirror and speedometer glances, or in the modified percent road centre algorithm, where the road centre region is expanded to the left or right depending on the curvature of the road (Ahlstrom et al., 2011).

It should also be noted that inferences of distraction based on gaze direction are inherently uncertain (Ahlström, Kircher, et al., 2021). First, eye trackers only measure where and for how long we look in a certain direction or at a certain target. It is not a direct overt measure of visual attention, and neither does it measure the purpose of the glance or what information that reaches the brain. Second, driving relies heavily on peripheral vision to acquire visual information, and all this information is unaccounted for when only considering gaze directions. Third, it has been shown that not all foveated information is cognitively processed. And finally, eye movement data do not provide an easy way to determine whether the sampled information was relevant, necessary, and sufficient for the driver in the current situation. Considering these limitations, it is clear that driver distraction monitoring should not be based on single foveations, without also considering glance history and the present traffic situation.

Given the advancements in environmental sensing in combination with theoretical developments in the definition of road user attention, context-aware algorithms that requires glances towards pre-defined target areas, which are identified by a combination of infrastructure and priority rules, have started to appear. Indeed, in a revised version of the above-mentioned AttenD algorithm (Ahlström, Georgoulas, et al., 2021), distraction is considered a multidimensional problem, represented by a number of separate buffers that may each deplete or restore at different rates, depending on situational context. However, context-aware distraction detection is a new field and currently in the development/research phase.

Another approach to distraction detection is based on driver activity recognition, targeting for example phone use by explicitly detecting mobile phones in the camera image (Moslemi et al., 2021). However, engagement in a non-driving related activity does not necessarily mean that the driver is inattentive, so there might be issues related to acceptance and compliance.

In addition to visual distraction, there are also other forms of inattention such as cognitive distraction and mind wandering. A typical indicator of cognitive distraction is a narrowing of visual scanning (Victor et al., 2005), or a lack of situational awareness, which can be measured via frequent misses of relevant objects in the context-aware visual distraction detection algorithms mentioned above. Cognitive distraction and mind wandering are difficult to induce, detect, and verify (Kotseruba & Tsotsos, 2022).

3.1.2. Driver fatigue monitoring

Non-obtrusive fatigue detection is usually based on heart rate metrics or on metrics extracted from the driver's appearance such as frequent blinking, closed eyes, yawning, and nodding. Eye, mouth, and head features are considered to be the most effective for estimating fatigue (Kotseruba & Tsotsos, 2022).

The parasympathetic influence when falling asleep slows down the heart and make its beating less regular. At the same time, the sympathetic nervous system is activated to resist falling asleep while driving (Vicente et al., 2016). These phenomena can be quantified using various heart rate variability metrics. However, both environmental factors, intra-individual time-varying differences as well as inter-individual differences obscure the relationship between heart rate variability and fatigue, which leads to inconsistent findings (Lu et al., 2022).

Eye strain, difficulty focusing, heavy eyelids and difficulties keeping the eyes open are other signs of fatigue. These features can be extracted from video data via a series of processing steps including face detection and tracking, facial landmarks detection, eyelid detection, and head dropping, mouth drooping and yawning detection. Eye features (blinks and closures) are further processed into standard fatigue measures such as percentage of eye closure and blink frequency (Albadawi et al., 2022). The long-term temporal aspect is crucial for fatigue detection why indicators such as slow blinking and yawning are often aggregated over several minutes across time (Bakker et al., 2021).

Many research papers show promising accuracies well above 90% (Albadawi et al., 2022). However, commercial systems seem to have difficulties achieving high sensitivity and specificity at the same time (Cori et al., 2021). For example, one of the most well-validated fatigue detection systems on the market has a sensitivity of 23% and a specificity of 96% when relating fatigue detections to lane departures (Shekari Soleimanloo et al., 2019). Thus, while the cautionary alarm produces few false positives, the proportion of missed events is concerning. Similarly, despite being a strong indicator of driver fatigue for highly averaged data, the commonly used percentage of eyelid closure fails to detect fatigue at finer temporal resolutions as well as on an inter-individual level (Golz et al., 2010; Trutschel et al., 2011). Combining multi-modal metrics and a global context, with additional features for continuous driving time, temperature, current time, and sleep duration may be a way to improve fatigue detection performance (Qian et al., 2021). Another approach to improve the performance of fatigue detection systems is to use personalised algorithms. The potential gain is large since between-individual phenotypic factors account for 50–95% of the variance (Yamazaki & Goel, 2020), and performance typically increase with about 20% when using personalised algorithms (Bakker et al., 2021).

3.1.3. Comfort monitoring

In empirical research, subjective measures of (dis)comfort are still the gold standard, typically by means of questionnaires (Anjani et al., 2021) or by continuous monitoring using measurement devices actuated by experiment participants when they experience discomfort. However, in everyday use of automated vehicles these methods are subjective and obtrusive, and there is desire to get more objective human comfort measures that can be assessed by sensors in real-time. Current approaches typically estimate comfort based on emotion or stress indicators. Examples include various physiological measurements such as heart rate, heart rate variability, galvanic skin response, pupillometry, blood oxygen level saturation, electromyography, electroencephalography, seat pressure distribution, and postural analysis (Ayaz et al., 2012; Beggiato et al., 2018; Ikeda et al., 2018; Tan et al., 2008).

Vehicle accelerations and behaviours (de Winkel et al., 2023; He et al., 2022) may impose physical discomfort and cause motion sickness, which may be inferred from expressions such as yawning, from pallor and from skin conductance. A sensitivity profile could be established, and predictions could be made for individual drivers, possibly incorporating knowledge of the route and vehicle accelerations (Irmak et al., 2022).

Facial expression recognition is a promising technique for estimating emotional states. Real-time facial expression recognition is typically based on automatic facial action unit analysis (Zhi et al., 2020). Even though action unit analysis only reflects changes in facial appearance, and not the expression/emotion per se, combinations of specific action unit changes can be used to infer particular emotional states (Ekman & Friesen, 2003). Automatic camera-based approaches have been presented (Bryant & Howard, 2019; Ko, 2018), but one must be aware that a person's intent goes beyond their facial expression. Individual differences in the quality and quantity of facial expressions obfuscates the relationship between action units patterns and distinct emotional states, and so does ambiguities such as contextual clauses (irony/sarcasm), and socio-cultural context (Barrett et al., 2019). Importantly, facial emotion recognition does not explain the trigger of the emotion, which would require information fusion with systems for environmental sensing, where the triggering event may even have occurred in the past.

Discomfort may also be inferred from situations that drivers find uncomfortable. Mediator Deliverable D1.2 (Borowsky et al., 2020) presents an overview of potentially uncomfortable driving situations, including the expected probability for a decrease in comfort, the probability to get timely information about these situations, and the possible time span for detection in advance. Using a lookup table to predict discomfort has its limits. Generalisability to situations not included in the table is limited, unexpected changes in drivers' comfort cannot be covered, and the expected probabilities listed in a lookup table may not work well for the individual driver.

3.1.4. Intoxication monitoring

There are five major categories of alcohol intoxication sensing approaches (Paprocki et al., 2022): breath alcohol devices, bodily fluid testing, transdermal sensors, optical techniques, and indirect estimates of intoxication based on physiological parameters or behaviour. The gold-standard for measuring alcohol is through gas chromatography, however, breath alcohol sensors are also reliable and are used in ignition interlock devices and for law enforcement. A disadvantage is that these sensors require regular maintenance, and that they only make one measurement.

Indirect estimates based on camera feeds has the potential to continuously monitor the driver's intoxication level. The estimate is based on the finding that alcohol impairment affects oculomotor control by decreasing the velocity and accuracy of glance behaviours while increasing the number and duration of fixations (Garrisson et al., 2021; Maurage et al., 2020; Silva et al., 2017). Koch et al. (2023) puts together these results by fusing various eye and head movement data in an attempt to classify drunk driving based on driver monitoring cameras. While the results are promising, there is still some way to go before the accuracy and reliability is good enough, and the technique must also be validated on a larger population.

3.1.5. Sudden sickness

Sudden sickness is an umbrella term that covers a variety of conditions (diabetic shock, cardiac events, seizures, etc.). Common for these severe states is driver incapacitation. Vital signs can be monitored with wearables, and to some extent also with no-contact sensors, but the sensors are sensitive to motion artifacts (Leonhardt et al., 2018). Instead of monitoring vital signs, it is reasonable to regard sudden sickness as a period of lack of response (Fredriksson et al., 2021). This may be detected indirectly, for instance from uncorrected fatigue or distraction warnings.

Whereas some conditions that cause a driver to be suddenly unavailable, such as an epileptic seizure, other forms of sickness may develop more gradually. An example is motion sickness. Mathematical models are currently under development that can predict future motion sickness with a reasonable accuracy. These models can make predictions using present observations on, for example, pallor, gasping behaviour, and manifestations of physiological discomfort such as postural changes, while also taking into account knowledge of individual sensitivities and route planning information (Irmak et al., 2022). If a risk of, or early stages of, motion sickness is/are detected, this may be put to use in automation. Decision logic may for instance propose to the driver to take over the driving task, as active involvement in driving is known to mitigate this form of sickness (de Winkel et al., 2021; Rolnick & Lubow, 1991).

3.1.6. Driver monitoring in automated vehicles

Driver monitoring plays a critical role in automated driving as long as the automation allows the driver to have some control over the vehicle (Halin et al., 2021). In Continuous mediation, the driver is responsible for the driving task, and the driver monitoring system should therefore monitor the driver continuously. In Driver standby, the driver is no longer in charge of the driving task and does not need to supervise them. The driver must, however, be fallback-ready to be able to take over the control of the vehicle upon request from the vehicle. A driver monitoring system should therefore be capable of (i) assessing whether the current and near future state of the driver allows him/her to take over the control of the vehicle if requested now or in the near future, and (ii) monitoring the driver’s state continuously if the automated function is disengaged.

UNECE regulation 157 and EU General Safety Regulation 2019/2144 mandates driver availability monitoring for automated functions (see section 2.3). This includes, for example, hands on wheel detection. In addition, El Khatib et al. (2019) discuss the need for driver monitoring while in automated driving mode to let the vehicle know if the driver is fit enough to regain control when asked to do so. If the driver decides to respond to a take-over request, a driver monitoring system would be useful to check whether the driver’s recent and current state allows for this.

3.2. Results on practical validity from Mediator

This section summarized the driver monitoring related findings from the experiments carried out in MEDIATOR.

3.2.1. Fatigue monitoring

A field study was conducted to investigate the transition from alert to sleepy while driving on real roads in real traffic, both during manual driving and when driving with Continuous mediation (Volvo's Pilot Assist 2), and both during daytime (supposedly alert) and night-time (sleep deprived). Participants indicated their level of sleepiness by using the Karolinska sleepiness scale (Åkerstedt et al., 2014). As expected, the results showed that night-time driving led to markedly increased levels of sleepiness, whereas partially automated driving led to slightly higher levels, especially in the night-time drives when the sleep pressure was high. During daytime, when the drivers were alert, partially automated driving had little or no detrimental effects on driver fatigue. More details about the study can be found in Ahlström, Zemblys, et al. (2021).

Two separate fatigue detection systems were developed based on the collected data, one based on physiological data, and one based on video data. The outcome of these measurements was validated based on the participants' scores on the Karolinska sleepiness scale. The performance of the two systems is summarized below.

3.2.1.1. Camera-based fatigue detection

The camera-based real-time fatigue detection system is based on an algorithm developed by Bakker et al. (2021). The system was set up as a two-stage model with a generic deep feature extraction module combined with a personalised fatigue detection module. The system can operate in two modes: binary classification ("sleepy" versus. "alert"); and in continuous-output regression-like mode, estimating directly the Karolinska sleepiness scale values. For binary classification we use the criterion that the "sleepy" class corresponds to reported values >=7, and "alert" corresponds to values <7. The system can also operate in a generic mode and in a personalised mode, where the latter adapts to a certain individual.

The generic model has a binary classification accuracy of 76%. This means that 24% of the test data are misclassified, either as false alarms (false positives, indicating fatigue when the driver is alert) or "misses" (false negatives, indicating alert when the driver fatigued). The personalised model, trained on data from one day and tested on data from another day, showed an accuracy of 90% and a mean absolute error of 0.7. Some sort of personalisation seems crucial since different people have different expressions of sleepiness. However, the increased detection performance comes at the cost of reduced usability and acceptance.

The following lessons learnt are worth mentioning:

- Camera positioning is key to acquire high quality data. Here, the camera was positioned on the steering column, meaning that the view from the camera was blocked during some steering manoeuvres, that the captures image cropped the face on very short or tall participants, etc. This had practical implications. For example, not all desired facial features could be used for sleepiness detection since they were outside the view from the camera.
- Personalised algorithms outperform generic algorithms.
- Truly independent train, validation and test datasets should be used when developing machine learning algorithms. To avoid generalisability issues when developing personalised algorithms, train and validation data should come from recordings from the same person but on different days and preferably under slightly different conditions.

3.2.1.2. Physiology-based fatigue detection

A binary classifier based on the AdaBoost method was developed to classify alert and sleepy episodes using the same Karolinska sleepiness scale-based criterion as above (Lu et al., 2021).

Different heart rate variability metrics acquired with a consumer wearable device were used as input to the classifier. The system can operate in a generic mode and in a personalised mode, where the latter is tailored to a certain individual, either by normalising the feature values with a baseline recording, or by injecting the training set with data from the individual being evaluated.

Out-of-fold validation (worst case scenarios where the trained model is doing inference on unseen drivers) shows a mean absolute error of 1.0 with good bounds for confidence level. This means that on average, the estimates are about 1 unit different from the subjective Karolinska sleepiness scale ratings. Binary classification accuracy was 78.9%. In the case where the model has seen all the drivers during training, the mean absolute error performance was 0.5 and the accuracy 86.5%, which is an indication of the performance that can be expected from a personalised algorithm.

Increasing robustness by using both camera-based and physiology-based features has not been done in the project, but it is a natural next step. We do not expect a substantial increase in accuracy, but availability would become better. For example, camera data loss is expected in higher levels of automation when the driver is out of position or when the face is obstructed with objects such as laptops.

In the later on-road study with the Mediator Human Factors in-vehicle prototype (cf. Mediator Deliverable 3.4; Fiorentino et al., 2023), the following lessons learnt are worth mentioning:

- Data were collected from sensors integrated in the steering wheel as well as from chest contact electrodes. The steering wheel electrodes provided fair quality data while drivers kept both hands on the wheel and only performed small steering movements. However, quality and availability deteriorated with hand movements and hands-off-wheel automation modes. This means that sensors located in the steering wheel provide data with sufficient quality in a non-obtrusive manner, but the technique is only suitable in manual driving or in Continuous mediation mode. For Driver standby and Time-to-Sleep mode, data will need to be complemented by other sensors, for example from a chest strap.
- For optimal sleepiness detection, fully continuous data would be required, but fair accuracy was achieved with down to 5 minutes of continuous data.
- The tested algorithm lacks in individualization capabilities, and while it performs fair for some participants it leaves room for performance improvement for others.

3.2.2. Distraction monitoring

An on-road study with ten professional drivers was conducted to investigate driver distraction when using the Mediator Technology Integration in-vehicle prototype. The drivers drove a 1-hour route ten times, 4 times with the complete Mediator system including driver state interventions, 4 times with the complete Mediator system but without driver state interventions, 1 time with the Mediator system essentially deactivated, and 1 time with a “misbehaving” Mediator system. Both quantitative and qualitative data were collected in the study. Distraction events were defined based on the AttenD algorithm (Ahlstrom et al., 2013). More details about the study setup and the results can be found in Mediator Deliverable D3.4 (Fiorentino et al., 2023).

The study aimed to investigate three questions:

1. Do distraction warnings reduce the number of distraction events?
2. Is there a difference in visual distraction when the Mediator system is available versus unavailable?
3. Is there a difference in the number of distraction events when driving manually versus when driving with Pilot Assist (SAE level 2)?

Repeated measures ANOVA results on the quantitative data showed that the distraction warnings did not have a significant effect on the number of distraction events. Neither was there a difference in the number of distraction events between manual driving and driving with SAE level 2. However, ANOVA results indicated that there was a significant difference in when the complete Mediator system was available versus unavailable, with a higher likelihood that the driver is looking away from the forward view when the Mediator system is active, Figure 3.1. This result is expected given the extra visual load and the continuously changing time budget information. Although the difference is significant, it is not likely to impose higher risks compared to driving without Mediator.

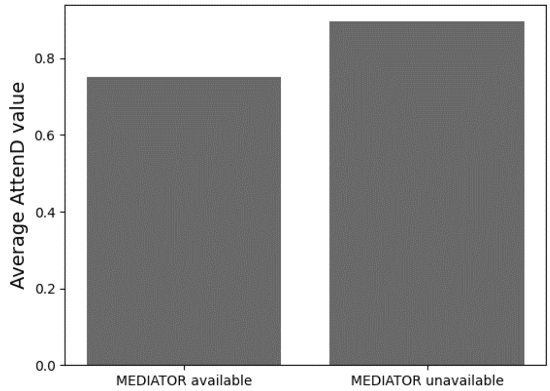


Figure 3.1: Average AttenD score when MEDIATOR was available versus unavailable.

Analysis of qualitative data, which included single-item measures combined with interview data showed that when distraction warnings were part of the Mediator functionality, the reliability of the system was rated low. One explanation might be that the distraction warnings were often perceived to be false alarms, and hence reduced the perceived reliability of the system. However, the overall reliability of the Mediator system was rated on average as good.

Another on road study, performed with 50 naive participants and a Wizard-of-Oz setup, investigated the effect of the full Mediator HMI, including distraction warnings, on driver distraction in Continuous mediation mode. Distraction was defined as looking away from the road for 2 seconds using the AttenD algorithm (Ahlstrom et al., 2013). The driver state algorithms used this AttenD output to determine the urgency level of the distraction. A higher urgency resulted in more urgent warning signals. The Mediator HMI was compared to a baseline HMI that was based on existing HMI designs (i.e., mainly using simple icons and sounds for interaction). In the baseline HMI the distraction warnings were turned off. A significant effect of both proportion of the time the drivers were distracted as well as the maximum duration of a distraction event was found between the Mediator and the baseline HMI. The results imply that the Mediator HMI, including distraction warnings, reduced distraction for these measures.

The differences in results between the two on-road studies could be explained by both the sample type and the study setup. Professional drivers were expected to look at the mediator HMI more often so they could give their feedback which could have resulted in more distraction. Additionally, the triggering and implementation of the distraction warnings differed between the two studies. The first study provided the same warning every time distraction was detected, while the warnings in the second study had a low urgency level when distraction was initially detected but the urgency level could increase when distraction continued (i.e., the severity of the distraction increased). There were many other differences between the studies, such as a different route, different vehicles and different HMI designs. It is therefore difficult to pinpoint exactly what caused the differing results. Regarding driver monitoring systems, it is advised to investigate the effect of adjusting the urgency level of the distraction warning on the compliance to this signal. If compliance is indeed increased with adjustable urgency levels of a distraction warning signal, this calls for driver monitoring systems to provide information on the severity of the distraction (rather than only the binary signal of distracted or not).

Lessons learnt regarding distraction monitoring from the on-road experiments in the Mediator Technology Integration in-vehicle prototype includes:

- The distraction identification was done by two cameras, a face camera placed on the dashboard to the right of the steering wheel and a body-view camera placed below the rear-view mirror. Complementing and combining eye/face tracking data with activity recognition software provided a better understanding of the driver's actions.
  - Information from the activity recognition software, in addition to the gaze tracking data, can be used to trigger more accurate and customized countermeasures.
  - Activity recognition benefits from combining several data sources such as body pose, hands activity, object recognition and gaze direction.
- Interviews with the participants, supported by questionnaire data, revealed that the distraction detection algorithm was too sensitive and gave false warnings, for example when the driver was waiting at an intersection and looking around to cross the intersection safely.
  - The number of false distraction warnings should be reduced by better quality assurance of sensor data and intermediate processed data.
  - False warnings should be mitigated by taking the driving context into account. The latter was done in the project by adapting the eyes-on-road requirements based on proximity sensors and road type.
  - Gaze direction algorithms should use (more) sophisticated (auto-)calibration, to improve gaze direction accuracy while preserving acceptance.

3.2.3. Comfort monitoring

An empirical study was conducted to investigate the potential of automated facial expression analysis for discovering action unit changes related to uncomfortable automated driving manoeuvres. The data comes from two driving simulator studies including 81 participants, all experiencing the same automated close-approach manoeuvre to a truck driving ahead three times. More details about the study are available in Mediator Deliverable D1.2 (Borowsky et al., 2020). Results from the study are summarized in sections 3.2.3.1–3.2.3.2. Lessons learnt regarding comfort monitoring include:

- Real-time comfort assessment by automatic video-based facial expression analysis revealed situation-related patterns of visual attention, tension, and surprise. However, these patterns were only found on an aggregated level over all participants, and the technology is not mature enough to detect (dis-)comfort for a certain individual at a specific point in time.
  - Analyses of personal characteristics revealed strong differences in effects, therefore facial expression analyses will not perform equally well for every person, even when personalizing changes at the individual level (e.g., individual high/low facial expressivity).
  - The discovered effects on an aggregated level were found to be shifted in time per person (earlier/later onset of reactions), which already impacts aggregated data analysis and is even more challenging at an individual level.
  - It could be the case that individuals do not only show different strength of the (same) effects, but completely different action units patterns. Identifying these

qualitatively different patterns would require a higher amount of data, especially when aiming to make predictions at individual level.

- Even though there seems to be general potential in facial expressions analyses for contributing useful information about users' satisfaction with the current operations of the automated system, detection and prediction at individual level still needs further research.
- Since real-time discomfort detection was found to be unreliable at an individual level, offline situation-based prediction of discomfort was used in the Mediator experiments (Mediator Deliverable 3.3; Borowsky et al., 2023). The offline approach enabled testing of active Mediator proposals with naïve users without false alarms of a still imperfect real-time discomfort detection system.

3.2.3.1. Face tracking quality

In the first driving simulator study two different camera brands were compared using the same face tracking software (Visage facial feature detection and face analysis SDK version 8.4). One camera (GoPro Hero 5) was mounted in the centre below the instrument cluster behind the steering wheel and the other (Intel RealSense SR300) centrally over the steering wheel. Due to the optimal horizontal angle, face tracking quality was high with only 6% of video frames without tracking for the lower camera and 12% for the upper camera. If the difference is due to the camera brand or the position is not known.

In the second driving simulator study, four video cameras of three different brands were used, capturing the driver's face from different directions. The main aim was to compare the impact of different camera angles on the face tracking results. Two cameras (GoPro Hero 5) were placed below the instrument cluster behind the steering wheel on the left and right side, one camera (Intel RealSense SR300) was placed centrally over the steering wheel, and the fourth camera (AVT Mako G-234B) was placed at the right side from the driver's perspective next to the steering wheel. The percentages of not tracked video frames were lower compared to the first study: 26% (GoPro Hero 5, right), 35% (GoPro Hero 5, left), 43% (AVT Mako), and 43% (Intel RealSense).

Overall, face tracking quality was primarily influenced by the camera angle and by obstructions of the face by the steering wheel. Some, mainly smaller, participants' mouth region was obstructed by the steering wheel, resulting in lost tracking or worse tracking quality. Reflecting eyeglasses or beards were other factors that affected tracking quality.

3.2.3.2. Discomfort-related effects on facial action units

To maximize tracking availability and quality, action unit tracking results from all cameras were combined. The resulting tracking rates during the discomfort sequences were 97% in study 1 and 80% in study 2, respectively. The results show that during the truck approach, participants' showed situation-related pressing and stretching of the lips, a push-back movement of the head, raising of inner brows and upper lids as well as reduced eye closure. These patterns could be interpreted as visual attention, tension and surprise. The results indicate that automatic facial expression analysis can be used in research settings to give information about users' comfort with automated vehicle operations. However, while aggregated results were stable on an aggregated group level, it was not possible to obtain stable and reliable results on an individual level (Beggiato et al., 2021).

3.2.3.3. Situation-based prediction of discomfort

Real-time comfort assessment is reactive in its very nature and aims to increase the drivers' comfort level in case it decreases. However, to ensure a comfortable and positive driving experience, it would be better to anticipate and avoid a potential decrease in comfort before is occurs. Predicting discomfort long enough in advance requires a different approach compared to real-time comfort assessment. Instead of using direct measurements based on facial expressions and physiology, the idea is to predict discomfort by monitoring the environment to identify upcoming uncomfortable driving situations. Mediator Deliverable D1.2 (Borowsky et al., 2020) presents an overview of potentially uncomfortable driving situations along with a time span for detecting the situation in advance. If such a situation is about to occur, an automated system can suggest a take-over from manual to automated driving (or vice versa) well in advance to avoid the uncomfortable situation. Examples of scenarios where discomfort may be avoided by a take-over from manual to (highly) automated driving include car following scenarios (55% a-priori probability of being experienced as uncomfortable), situations with poor visibility at night (54%), and drowsiness (70%). In other cases, discomfort may be avoided by suggesting a hand-over from automated to manual well in advance of situations where the road conditions increase the risk of motion sickness (75%), or in situations that cannot be managed by the automated system (100%).

3.3. Results from discussions with developers

A series of bilateral discussions/interviews were conducted with representatives from driver monitoring system companies, tier 1 suppliers, original equipment manufacturer, and test organisations. The representatives were Raimondas Zemblys from Smart Eye, Clémentine Francois from Tobii, Fabian Faller from Continental, Claus Marberger from Bosch, Caroline Chung from Veoneer, a driver monitoring expert from Stellantis, Mikael Ljung Aust from Volvo Cars, and Rikard Fredriksson from the Swedish Transport Administration and Euro NCAP. Results from these discussions are summarised in the form of questions and answers.

3.3.1. Which driver impairments are most important to detect and mitigate?

All respondents mentioned the same four impairments, (i) Distraction and Inattention, (ii) Fatigue and Sleep, (iii) Sudden sickness, and (iv) Intoxication from alcohol and drugs. The reported reason why the respondents brought forth these four impairments was that they contribute to a substantial proportion of fatal crashes on our roads. With assisted and automated driving features, the list should also be complemented with (v) driver engagement, to ensure that drivers fulfil the requirements for supervision and take-over performance.

Several developers raise the question of occupant monitoring. By keeping an eye on all occupants in the vehicle, it becomes possible to ensure that children are not left behind in hot cars, that all occupants wear their seatbelt, and that safety systems such as airbags are used in a proper way if occupants are out of position.

3.3.2. Which ground truths or gold standards are used to assess driver impairment?

Microsleep and Sleep are often pragmatically defined as closed eyes, even though a more rigorous approach is to assess sleep via polysomnography. Fatigue and Drowsiness are less straight forward to assess. They are typically operationalised via subjective sleepiness ratings on the Karolinska sleepiness scale, where alert is defined as 1–4, 1–5 or 1–6 and fatigued is defined as 6–9, 7–9 or 8–9. The intermediate ratings 6 and 7 are sometimes added as a third class, and sometimes they are left out during system development/training (in case of binary classification of alert vs fatigued). In cases where the Karolinska sleepiness scale ratings are clearly incorrect, some developers adjust (or omit) these erroneous ratings. This trend of using subjective sleepiness ratings has been reinforced with the General and Pedestrian Safety Regulation package (Regulation (EU) 2019/2144), where the EU requires the fitment of “a system that assesses the driver’s alertness through vehicle systems analysis and warns the driver if needed” on new vehicle types from 2022, and that advocates using subjective sleepiness as the evaluation criteria. The same subjective sleepiness ratings are also often used as target values when developing fatigue detection systems. One respondent argued that the only measure that can be used to assess driver fatigue is the driver's performance. Since measures lose their relevance in autonomous driving scenarios, it would be favourable to find new fatigue indicators/metrics that are strictly correlated to driving performance in manual driving. The same metrics could then be used in automated driving settings as well.

When mentioning *distraction,* most developers refer to looking away from the road too often or for too long. Road is here synonymous with the forward windscreen or a corresponding area, and the meaning of too long is usually 2 – 3 seconds. Looking away is measured either via manual annotations or via eye tracking. Some argue that measuring eyes off road is enough to catch most instances of distraction while others state that distraction detection is more complicated than just measuring eyes off road. In the latter case, there is a desire to measure if a driver is attentive enough in a particular situation, but this requires information about the situation and the intentions of the driver. Some also mention activity recognition as a direct indictor of distraction, equating behaviours such as mobile phone use or talking to a passenger with an impaired state. In addition to visual distraction, there is also *cognitive distraction*, or attention that has shifted away from goal relevant information. Cognitive distraction is often operationalised via workload and measured using NASA-TLX (Task load index; Hart & Staveland, 1988), even though this is not necessarily true given that cognitive distraction can happen in situations with both high and low load.

Although thresholds may differ between countries, *intoxication by alcohol and drugs* have clear medical definitions and there are gold standards for testing and quantifying substance levels. The situation is similar for *sudden sickness*, which should be seen as an umbrella term covering a variety of conditions (diabetic shock, cardiac events, seizures, etc.), where the common result is driver incapacitation.

3.3.3. Which driver impairments can be measured today?

For *fatigue*, several respondents said that eye movements and blink behaviour are particularly good indicators. Especially, eye movements slow down and the eyelids are closed for longer periods of time. A difficulty is that the same indictors are also signs of cognitive load and alcohol intoxication. It is feasible to detect even early signs of fatigue. Some developers state that personalised algorithms will improve detection performance, either using stored baseline data from a certain driver, or by collecting baseline information in the beginning of the drive. Other developers state that their fatigue detection system does not need personalised baseline information. Either way, it is necessary to use a time window of several seconds up to several minutes to see how the eye metrics evolve over time, before the fatigue estimate can be derived. By and large, driver monitoring system developers seem confident that fatigue, microsleep and sleep can be measured with camera-based systems. There are, however, difficulties to verify that the systems work as expected since ecologically valid test procedures are complex and time consuming.

Camera-based face and eye tracking has come a long way and is quite robust today. This technological development has paved the way for driver monitoring systems capable of measuring when drivers' eyes are not directed to the forward roadway. In that respect, *distraction* can be measured today. Similarly, the camera can be used for activity recognition to detect non-driving related tasks like eating, drinking, and reading a book. There is also interesting research on assessing attention and intentions, mind wandering and cognitive load, but such systems are not market-ready yet.

Alcohol intoxication can be assessed with breathalysers, also with contact-free, unobtrusive measurement of a driver's breath alcohol level. Such systems are usually integrated with an alcolock, preventing the engine from starting if the driver is under the influence of alcohol. There is research investigating if alcohol can be assessed with a remote sensor such as a camera, but such systems are not market-ready yet.

Sudden sickness is typically not measured directly. Instead, the state is inferred from non-responsiveness as driver incapacitation. Euro NCAP defines this as a driver who either does not return their gaze to the forward road view within 3 seconds of an inattention warning or a driver whose gaze has been away from the forward road view or has been eyes closed for more than 6 seconds.

3.3.4. What are the limitations of today's driver monitoring systems?

Camera-based driver monitoring systems are now so extensively trained and validated that they require no calibration, track faces and eyes almost instantaneously, operate across a near 180-degree range, and work through most sunglasses. The algorithms can manage gender and ethnicity, and works with most hoods, hats, caps, scarves, face masks, hijabs and niqabs. To an extent of course. Certain glasses still pose a problem, either if they cause a lot of reflections, or if they are blocking infrared light, and too much of the face cannot be covered by hoods, face masks etc. Vehicle manufacturers are also worried about the black box machine learning algorithms that are used for face recognition and tracking. It is yet largely *untested* how these algorithms perform when scaling up the number of users. Large-scale deployment always reveals problems that have not been caught in the prelaunch testing phase, and since driver monitoring is intended as a safety system, it must work as expected on all drivers.

Many vehicle manufacturers opt for single camera solutions, low-cost sensors, and power efficient lower end automotive computers. This inevitably leads to problems if the view of the (one) camera is obstructed, for example by a book or a phone. Further, the calibration-free approach comes with the cost of reduced accuracy. If the intended application requires an absolute gaze direction *accuracy* of 1 degree, there will be quality problems. An accuracy of about 5 degrees will however work fine on most participants without calibration. Also, when using one or just a few cameras, it is difficult to get a clear camera view of the drivers' eyes in all gaze directions. Typically, vehicle manufacturers want to position the cameras close to rear-view mirror to be able to see the whole cabin, making it hard to monitor the face of the driver and all the desired signals, because there will only be a few pixels accounting for the driver's face. It will also be challenging to track the face if there is occlusion like baseball caps, or to track the eyes if the driver looks down, because the camera will only be able to see the eyelids. Another common sensor location is on the steering column, which means the line of sight will often be obstructed, typically by the hands and the arms, and especially so when turning the steering wheel. The best camera location for eye and eyelid tracking is right in front of driver, but that may not be so practical. Lower sensor positioning is preferable to cover glances inside the vehicle, while a higher position is better when tracking glances through the windscreen. Similarly, very short or very tall drivers may be difficult to track since only parts of their face fits into the camera image. Using a multi-camera system would solve many of these issues, but few customers are willing to pay for this, so mono-camera systems will be most common on the mass market.

Infrared camera sensors work well in the dark, and after moving from 850nm to 940nm, they also work well in direct sunlight. Depending on where the sun is there might still be issues though. Some developers have started to investigate other wavelengths such as the short-wave infrared band. The advantage of moving to shorter wavelengths is that the sunlight has lower energy in these frequency bands. However, these new cameras are not market-ready for automotive face tracking applications.

An indirect problem with sunlight is that drivers squint when they get the sun in their eyes, obstructing the view for the camera. Some developers state that they have solved this issue while other see it as problematic.

Measuring gaze direction and eyelid opening is one thing. Interpreting the meaning of a glance and estimating the information gained via a glance is something else. When it comes to making sense of the eye movement signal, how to interpret it and to decide whether the situation is critical or whether it is reasonable to start interacting (warn) with the driver there is still a long way to go. Camera-based or physiology-based systems are referred to as direct driver monitoring systems, but it must be understood that these systems still use *indicators* to *estimate* the drivers' state. Gaze direction is a proxy for visual attention, but foveal vision is not enough to determine if drivers have perceived and understood their environment. A clear example is the fact that peripheral vision cannot be measured, so there is no way of knowing if a driver has seen a pedestrian just because there was no fixation in that direction. Similarly, activity recognition can be used to measure that a driver holds a mobile phone, but this does not necessarily mean that the driver is inattentive. Reduced heart rate and increased heart rate variability could mean that the driver is drowsy, but it could also mean that the driver is in a relaxed low load state. Such corner cases are hard to get around, and unfortunately the corner cases are quite common.

Tier 1 suppliers have started to merge information from driver monitoring systems with the vehicle's situational awareness sensing systems, but such fusion-based systems are still in the development phase. Fusing information about the driver's state with the prevailing situation and the current automation level will become more important in the near future, to make sure that the driver and the vehicle are level compliant. Taking complexity such as traffic density, weather, and road curvature into account to determine how probable it is that the automation system will work properly, in combination with information about the driver's state, to determine the time needed for the driver to safely take over. This is the main reason for fusing driver monitoring with external monitoring. The application of improving the actual driver state detection by making use of environment sensing appears to be down prioritized by system developers, perhaps since it quickly leads to more complex system setups. For example, if you want to measure the intersection between the measured gaze vector and outside objects, it is not sufficient to use a calibration-free mono-camera eye tracker. With a gaze direction error of up to 5° this will quickly lead to large alignment errors if the object of interest is 20–30 meters away from the vehicle. Building a safety critical system under such premises is not feasible.

When considering devices that determine the state of a driver by physiological signals, dissimilarities in physiologic mechanisms and origins cannot be disregarded. One of the key issues that needs to be resolved is linking the characterization of the impairment as manifested in the physiological parameters with driving task performance, and, conceivably, to the risks associated with a particular driver's condition. Another concern is the validation of monitoring devices, to prove the acceptance, effectiveness, and reliability of physiology-based monitoring devices before they are brought to the market. Indeed, there are currently no agreed-upon physiological parameters to serve as a reference (ground truth, gold standard) for the validation and test protocols that apply specifically to such systems.

3.3.5. How to make best use of driver monitoring information?

The first thought is that a detected driver impairment should be communicated to the driver. This could be in the form of a warning, a suggestion, some advice, or similar. Vehicle manufacturers are not very keen on this solution. A *fatigued* driver is already aware of being sleepy but often continues to drive anyway. For a warning to be efficient and convincing, it could be timed to cooccur with an event such as a lane departure and an automatic steering back into the lane. Unless the warning is connected to an event, it is all too easy to dismiss with a mental remark such as “I am almost home and will surely manage to drive the last 30 minutes as well”. In the same vein, the warning should be coupled with a suggestion on how to best counter the impairment, like guiding the driver to the nearest rest stop.

Distraction warnings provided when looking away for too long can also be problematic for several reasons. First, the driver will most often glance back at the road anyway after a long glance away from the road. Second, if the driver gets a warning, looks back at the road, and finds that everything is ok, then the feeling of being in control is reinforced, resulting in lower acceptance. Third, if an event requires the driver’s immediate attention, it is likely that the warning (that is triggered after looking away for several seconds) will arrive too late. Inopportune glances away from the road can be detrimental even if they are short, and it is not likely that a distraction monitoring system will catch these while providing a warning that gives the driver sufficient time to act. A better solution would then be to prevent or mitigate the unfolding situation by an automatic emergency breaking/avoidance manoeuvre. Such systems are already tuned with the inattentive driver in mind. For example, Euro NCAP suggests that a forward collision warning should not be issued later than when the time to collision is 1.7 seconds. This time allows the driver to look up, get a grasp of the situation, and initiate an evasive manoeuvre. For an attentive driver, 1.7 seconds is quite long and there is a great chance that the warning is perceived as overcautious and often even incorrect. It is often suggested that driver assistance systems, such as forward collision warnings, should be issued earlier if the driver is distracted. However, since the “looking back and react” time is already included in the 1.7 seconds (or similar threshold), this will not add much. On the contrary, and perhaps counterintuitive, vehicle manufacturers instead argue that the role of the driver monitoring system should be to delay warnings or automated evasive manoeuvres. This would prevent incorrect interventions and thus increase acceptance and trust, and consequently avoid disengagement of safety functions, which will increase safety in the long run. In this scenario, the driver monitoring system is not used as a direct safety measure, but rather as a means to avoid false interventions. Vehicle manufacturers are not very interested in monitoring the driver. What they want is to get a better understanding of when it is legitimate for the vehicle to interfere, or not, with the driving task.

A use case where driver monitoring systems can be used more directly is *sudden sickness*, or rather if the driver is clearly incapacitated or simply moves out of position and disappears from the camera image. In such cases it is obvious that the driver is incapable of driving, and consequently the safety systems in the vehicle should make sure to remain in the lane, and preferably also come to a safe stop.

Alcohol intoxication can be measured with a breathalyser, and the mitigation strategy is then to not start the engine to prevent drunk drivers from driving. Given the undisputed negative effects of alcohol intoxication on traffic safety, the natural thing would be to install alcolocks in all vehicles. However, customers do not want this, why vehicle manufacturers will not take this step. And apparently governments do not want this either because otherwise alcolocks would have been mandated by law. Nuisance alarms, where detections arise from alcohol content in food or medication, can be an issue though.

In relation to sickness and alcohol detection, it was clear that automakers are not very keen on installing medical equipment in the vehicles, especially with the accompanying calibration and approval processes. Instead, they are attempting to measure when drivers drive *as if* they are drunk, and if this happens, they use the ordinary safety systems to manage the situation to the extent possible.

3.4. Mapping driver impairments to driver monitoring systems

Inspired by the Safe-by-Design heuristic developed by Jannusch et al. (2021), representative information about driver impairments is here used to identify possibilities and limitations with different sensors/information commonly used in driver monitoring systems. The idea is to identify the need and effectiveness of driver monitoring based on representative usage patterns. The rows in Table 3.1 lists representative driver impairments, whereas the columns list the type of driver monitoring sensor/information that is suitable or needed to evaluate each impairment. Table 3.1 is deliberately broader than the scope of MEDIATOR and aims to give a holistic overview of the diverse and multifaceted information need that is associated with driver monitoring. Examples of what the columns represent is provided below:

- Lateral and longitudinal control: Driving performance metrics such as line crossings, lateral variability, and short headways.
- Automation mode/state: An account of which automated functions are currently activated. The information is needed since different automation levels put different requirements on the driver.
- Surrounding road environment: Information about surrounding road users and their predicted travel path. Could be used to improve workload and distraction metrics, by taking driving context into account.
- Digital maps: Data on the surrounding infrastructure. Could be used to improve workload and distraction metrics, by taking driving context into account.
- Out of position: If the driver is positioned in a way that makes it difficult or impossible to see the road.
- Posture: Related to out of position. Could be used when head/eye tracking fails, and also as a sign of fatigue or sleep (slouching body posture) or incapacitation (collapsed body posture).
- Hands on steering wheel: An indication of driver readiness, perhaps mostly useful in transfers of control, and in shared control situations.
- Head tracking: An estimate of visual information acquisition.
- Eye tracking: A better estimate of visual information acquisition.
- Eyelid opening: Related to fatigue and sleepiness.
- Pupil diameter: Psychophysiological indicator of various impairments such as cognitive load, fatigue, drug abuse and alcohol intoxication.
- Activity recognition: Detection of certain activities, such as holding a mobile phone or eating a sandwich.
- Facial expression recognition: Estimation of sentiments such as anger, disgust, fear, joy, sadness, and surprise. Sometimes also compound emotions such as happily surprised and sadly fearful.
- Respiration: Breathing patterns has a bi-directional relationship with emotional states such as anxiety, depression, anger, stress, and also with fatigue.
- Heart rate: Heart rate and heart rate variability patterns has a bi-directional relationship with emotional states such as anxiety, depression, anger, stress, and also with fatigue.

- Skin conductance: The sympathetic branch of the autonomic nervous system reflects emotions and arousal, which affects sweat gland activity and hence skin conductance.
- Alcolock: Provides an estimate of blood alcohol content before the drive.

Table 3.1. Representative examples of driver impairments versus driver monitoring related information that can be obtained from the interior and exterior of the vehicle. The number of + indicates greater potential/need for the sensor data to meet the requirements evoked by the example impairment. The number of + is our estimate based on the information in this chapter.

	Lateral and longitudinal control	Automation mode/state	Surrounding road environment	Digital maps	Driving time	Out of position	Posture	Hands on steering wheel	Head tracking	Eye tracking	Eyelid opening	Pupil diameter	Activity recognition	Facial expression recognition	Respiration	Heart rate	Skin conductance	Alcolock
Long glances away from the road	+			+					+	++								
Visual time sharing	+								+	++								
Inopportune glances away from the road			++	++					+	++								
Neglect to check for cyclists over the shoulder before turning right			+	++			++		+	++								
Neglect to assure free space before making a lane change			++	++			++		+	++								
Unfit to take over control in relation to takeover request		++				++	++	++	+	++	++							
Anger, sadness, frustration	+						++					+		++	+	++	+	
(Dis-)comfort							+							++	+	+	+	
Eating/drinking	+												++					
Tending to children	+									+			++					
Interact with navigation system	+								+	+			++					
Interact with mobile phone	+								+	+			++					
Talk on handheld mobile phone									+	+			++					
Talk on handsfree mobile phone									+	+			++					
Boredom		+		+			++				+	+		++	+	+	+	
Early signs of fatigue				+	+		+			+	++	++		+	+	+		
Severe fatigue	+	++		+	+		+		+	+	++	++		+	+	++		
Sleep	+	++				+	++	+	+	+	++	++			+	++		
Above legal alcohol limit	+	++								+	+	+	+					++
Intoxication via drugs		++								+	+	+						
Incapacitated driver	+	++				++	++	+	++	+	++					++		

3.5. Safety benefits of driver monitoring systems

The share of fatalities in distraction-affected crashes, i.e., a crash involving at least one driver who was distracted, is 8.1% in the US (Stewart, 2022) and 5 - 25% in Europe (European Commission, 2022). For fatigue, the share of fatalities involving drowsy drivers is 1.6 % in the US (Stewart, 2022) and about 17 % in Australia (Ansari et al., 2023). In Europe, a survey across nineteen countries have shown that the prevalence of falling asleep while driving in the previous 2 years is 17 %, and amongst those who fall asleep, the prevalence of sleep-related crashes is 7 % (Gonçalves et al., 2015). For alcohol, the share of fatal crashes involving alcohol are 17 % in Australia (Ansari et al., 2023), 30 % in the US (Stewart, 2022), and 25 % in Europe (European Commission et al., 2022). Sudden sickness has been found to be the direct cause for about 10% of fatal motor vehicle accidents, where cardiovascular related conditions are the dominating cause (Tervo et al., 2008). All these fatalities can obviously not be avoided by driver monitoring systems, since detection of an impairment is different from preventing it from happening. This makes it difficult to estimate the true safety benefits of driver monitoring systems.

Many new vehicles are equipped with advanced driver assistance systems designed to actively prevent crashes. Examples include forward collision warning, autonomous emergency braking systems, lane departure warnings, and lane keeping assistance. Such systems reduce front-to-rear, single-vehicle, sideswipe, and head-on injury crash rates, i.e., crashes that are often associated with driver fatigue and/or inattention. It has been shown that:

- Lane departure warnings lower involvement rates in crashes of all severities with 18 %, in those with injuries with 24 %, and in those with fatalities with 86 % (Cicchino, 2018).
- Forward collision warning reduces front-to-rear crash rates with 27 % and front-to-rear injury crash rates with 20 % (Cicchino, 2017).
- Low-speed autonomous emergency braking reduce front-to-rear crash rates with 43 % and front-to-rear injury crash rates with 45 % (Cicchino, 2017), and pedestrian crash risk with 25–27 % and pedestrian injury crash risk with 29–30 % (Cicchino, 2022).

It is not known how many of the remaining injury-crashes that could have been avoided by an earlier intervention triggered by a driver monitoring system.

The extent to which more advanced automated functions have safety benefits above and beyond what was listed above has yet to be determined (Mueller et al., 2021). There are however concerns about potential unintended negative consequences. For example, driver fatigue can be higher when using SAE level 2 as compared to manual driving (Dunn et al., 2021; Kundinger et al., 2020), especially during night-time when the sleep pressure is high (Ahlström, Zemblys, et al., 2021). It is also more common to engage in non-driving related tasks when using automated functions (Dunn et al., 2021; Kim et al., 2022). A recent study conducted in Finland aimed to identify to what extent vehicles capable of SAE level 3 automation may improve traffic safety (Malin et al., 2022). In the study, level 3 was implemented as a system able to keep the vehicle in lane and maintain a safe distance to vehicles in front, including functionalities for emergency braking and electronic stability control. Under the assumptions that *all* target crashes are prevented by the system, and that 100 % of the fleet is equipped with such systems, it was estimated that a level 3 system designed for motorways has the potential to affect 3.3 % of injury crashes, 3.1 % of fatalities, and 3.2 % of all serious injuries in Finland. The corresponding fractions for a level 3 system operating in urban environments were 2.2 %, 1.1 % and 2.5 %, respectively. Bjorvatn et al. (2021) found that level 3 automation on motorways improves safety if the operational design domain requirements are fulfilled. However, since the motorway network is limited and since its safety level is already good, the total safety effect (of all injury crashes) is limited to 0.1–1.2 %, with a penetration rate of 5–30 % (Yue et al., 2018).

Based on the study by Malin et al. (2022), we can assume that about 3 % of all injury crashes can be prevented if manual driving is replaced with perfect automation on motorways. Out of these 3 %, about one third are caused by sensing/perceiving factors or impairment (Mueller et al., 2020). Since driver monitoring systems perform best in similar environments (similar operational design domains) as automated systems, it makes sense to consider injury prevention by perfect automation as an upper limit to what can be achieved with a perfect driver impairment mitigation system, which would then be one third of 3 %.

4. Guidelines

Based on literature, existing recommendations, interviews with experts, and experimental research performed within the Mediator project, several guidelines have been developed for driver monitoring systems. The guidelines proposed in the following are defined based on functionality, technological possibilities, safety relevance and feasibility. In this chapter, first the requirements of a driving monitoring system for vehicles with multiple levels of automation are briefly discussed after which the guidelines are summarized per sensor type and per impairment. The chapter ends with guidelines related to the evaluation of driver monitoring systems.

Requirements for a driver monitoring system are dictated by the automation mode. The automation mode determines responsibilities and affordances, but also the a-priori probability that certain states occur (see sections 2.4.1 and 2.4.3). Requirements also differ depending on whether the driver monitoring system is used for classification of current state variables, or prediction of future states.

In *Continuous mediation*, the driver must be alert and must uphold situational awareness by monitoring the traffic environment as well as essential information systems such as the speedometer, the mirrors, and the state of the automated functions. The main challenges for a driver monitoring system are fatigue and distraction detection, and possibly specification of their causes. For instance, optimal mitigation strategies may differ depending on whether fatigue is sleep related or caused by cognitive underload or overload.

In *Driver standby*, the driver is allowed to be distracted, but must be ready to take over in seconds. The main challenge is fatigue detection, ensuring that a driver will be able to respond in time when a take-over request is made. In addition, it may be desirable to make sure that a driver has sufficient situational awareness to be able to respond adequately. Therefore, occasional visual sampling of essential vehicle information systems may be required. Another potential concern for future vehicles where the driver may be out of position, for example to engage with other vehicle occupants or to facilitate work, is that the driver must face forward before any transfer of control to the driver takes place.

In *Time-to-Sleep*, the automation can handle the driving task for a significant amount of time (e.g., enough to take at least a short nap). If situations arise that the automation cannot handle, it must be capable of performing a safe stop manoeuvre in cases where the driver is not able to take over control. For this automation level any oncoming takeover request will be feasible from a wakeful state and therefore the primary challenge lies in waking up a *sleeping* driver and estimate the time required to sufficiently recover from sleep inertia. Sleep inertia is characterized by a transitory period of hypovigilance, confusion, disorientation of behaviour and impaired cognitive and sensory-motor performance. The duration of this period depends on a large number of factors including sleep history, circadian timing, duration of the sleep episode, which sleep stage the driver is awakened from, and if one eats before sleeping (Hilditch & McHill, 2019; Tassi & Muzet, 2000). Specifically, people are most difficult to wake from deep sleep cycles, a situation which is also associated with the most severe sleep inertia with detrimental and lasting cognitive effects (Ferrara & De Gennaro, 2000). Consequently, accurate estimation of the time required for a driver to retake control of the vehicle requires knowledge of the current sleep cycle, as well as knowledge of idiosyncratic properties that affect this time (Hirsch et al., 2020; Wörle et al., 2021).

4.1. General sensor-related aspects

To adhere to the requirements previously described, a driver monitoring system will need to be equipped with sensors. Here, several types of sensors that can be used to estimate the level of distraction, fatigue and comfort are described and general guidelines for these sensors are provided. Guidelines that are specific to either distraction, fatigue or comfort monitoring are described in subsequent sections.

General guidelines that apply to all driver monitoring systems include:

- No safety critical systems in the vehicle should rely on the driver monitoring system, for example during transitions of control from automation to human.
- The system should self-diagnose failures and test its functioning before the start of its operation.
- The system should operate regardless of adverse environmental conditions (low lighting, wet and dirty conditions, humid and warm, etc.).
- The robustness of driver monitoring methods can be improved by combining multiple, complimentary or redundant, sensory modalities (i.e., *sensor fusion*). If this is done, vehicle manufacturers must guarantee proper integration of the various subsystems.
- The use of a separate Controller Area Network bus for driver monitoring sensors and subsystems is highly recommended for issues of safety and protection against interferences.

4.1.1. Computer vision systems

Computer vision systems are used for classification of distraction, fatigue, and comfort/emotions, by monitoring driver posture, head orientation, gaze direction, eye-closure, facial features, and engagement in NDRA. The requirements for camera systems for monitoring of different driver state variables largely overlap, and the following general recommendations can be made:

- Eye trackers come in a variety of forms, including cameras facing a subject, head-mounted devices, and glasses. To maximize convenience for users and to prevent dis- and misuse of monitoring systems, unobtrusive remote eye trackers should be implemented.
- Eye monitoring cameras and IR omitting sources should not interfere with the driver's view.
- Computer vision systems tend to be susceptible to artifacts caused by rapid changes in illumination common in car driving. Infrared cameras, possibly combined with infrared light emitters, appear most robust to changes in lighting conditions (Hermens, 2020).
- The system should be robust to occlusions of the face and/or body. This can be achieved by using multiple, and at least two cameras, placed in different locations.
- The system must be able to deal with individual differences in appearance such as: face shape, skin tone, eye shape, resting aperture and colour, the presence of facial hair, and various kinds of lenses and spectacles. Each of these factors has been implicated as affecting eye tracking accuracy and precision (Holmqvist et al., 2023).
- Most manufacturers of remote eye trackers recommend a distance between the camera and the eyes within a narrow range of approximately 60–70 centimetres, up to at most approximately 50–100 centimetres. These distances are dictated by the camera systems' optical properties. When a person moves outside of the tracking range, inaccuracies, noise and data loss are likely introduced (Holmqvist et al., 2023). Camera systems should be installed in positions that ensure the optical *sweet spot* matches the range in which a driver is free to move their head.
- The system should evaluate the reliability of its classifications. A warning should be issued when the system is not able to provide classifications of eyes on road and/or NDRA with sufficient certainty, for instance due to the driver wearing a mask.

- Eye trackers often require a calibration procedure to be performed at the outset of system use. Ideally, an eye-tracking system implemented in a vehicle should work without the need for such calibration, because the need for a driver to perform a calibration procedure each time the vehicle is started, even if short, can be considered obtrusive and is error prone, and is therefore detrimental to user acceptance.
- The system should become operational when the vehicle is switched on and remain on as long as the vehicle is operational.

4.1.2. Vehicle-based data

Driver control inputs to the vehicle can be used to determine driver state variables. Fatigue, for example, may be estimated using the time history of lateral position or steering wheel angles. Here, frequent small corrections are indicative of an alert state, whereas infrequent large inputs are indicative of drowsiness. In a comparison of this method to estimation of fatigue based on eye-closure, it was found to be more accurate (79% vs. 55%; McDonald et al., 2014). On the other hand, optimised vehicle-based indictors of fatigue perform worse than a biomathematical model of fatigue based on sleep history alone (Sandberg et al., 2011). An advantage with vehicle-based data, such as lateral and longitudinal positioning, is that these measures directly reflect safety critical performance decrements like swerving and lane departures. At the same time, a disadvantage is that it can be difficult to assess the cause of the event, which in turn makes it difficult to deploy an appropriate countermeasure.

Other types of vehicle-based data, such as when drivers override the automated driving system, can also be used as an indirect indicator of discomfort.

The use of driver control inputs to measure state variables is unobtrusive and can augment estimates from other systems to improve accuracy and precision.

- To be able to extract useful information from these data, the sampling rate must be sufficiently high to capture the relevant range of control input frequencies. A sampling frequency of about 10Hz is adequate to capture the highest frequency components of inputs to the steering wheel, accelerator, and brake pedal alike (Delice & Ertugrul, 2007).
- The accuracy and availability of the data must also be good enough, and above all, it must provide an indication that data is missing in cases when data are unavailable. A typical example here is incorrect lane position data on roads with missing or poor road markings.

4.1.3. Physiological data

Empirical studies typically measure physiological correlates of driver state variables using methods such as adhesive electrodes or other sensor elements placed on various locations of the body. Given the obtrusive nature of such methods, they are not desirable for actual implementation in vehicles. Non-obtrusive alternatives are, for example, capacitive sensors in the steering wheel. Through contact with the fingers, these sensors can be used to measure heart rate (Leonhardt et al., 2018), as an indicator of fatigue or stress/discomfort; and skin conductance, informative of motion sickness (Warwick-Evans et al., 1987). However, such measurements require two hands on the steering wheel. Several alternative methods to monitor physiological state variables may be considered:

- Contact-based sensors can be incorporated into constant-contact areas between driver and vehicle, such as the back of the seat or the seat belt. Such solutions offer the advantage that they allow a driver to take their hands off the wheel but have the disadvantage that sensors must be sufficiently sensitive to obtain readings through other media, or materials such as clothing. This may prove detrimental to estimation of some state variables but may not be of much concern for other cases; for example, monitoring changes in posture using a pressure-mat in order to infer discomfort.

- Wearables and other nomadic sensors can be used as an "opt-in/added value", but sensors not integrated in the vehicle cannot be used for safety-critical decisions.
- Steering wheel sensors should operate without wiring and without obstructing the operation of the airbag.
- As non-contact methods, thermal and RGB camera imaging may be considered. These methods monitor volumetric changes in the facial blood vessels during the cardiac cycle and indicate the timing of cardiovascular events. These methods can be used to estimate heart rate, but they are not sufficiently precise to estimate heart-rate variability (Kranjec et al., 2014).
- Other non-contact sensor systems have been proposed, for instance based on monitoring reflections of laser light, ultrasound, and micro-waves (i.e., optical vibrocardiography, ultrasonic sensors and radar). These methods are potentially sufficiently precise to determine heart rate variability. However, as of yet, these methods should be considered experimental (Kranjec et al., 2014; Leonhardt et al., 2018).
- Meaningful analysis of heart rate variability is mainly dependent on the integrity of the basic cardiac input signal and the temporal accuracy of heartbeat detection. For healthy adult individuals with normal amplitude variability, a minimal sampling frequency of 125Hz and sampling window size of 1 minute is required, although rates of 200 Hz may be necessary for low-amplitude rhythms in some populations (Laborde et al., 2017).
- Even with an adequate digitization rate, temporal accuracy of heartbeat detections can be degraded by noise in the input signal and this timing may require signal filtering or the application of a peak-finding algorithm to improve the localization of the heartbeat. Therefore, it should be preferable to refer to the overall resolution in inter-beat interval detection of the entire recording / analytical system. An overall accuracy in the timing of two consecutive heartbeats of 2 milliseconds is adequate for most applications, but a resolution of 1 millisecond is preferable (Berntson et al., 1997).
- The window size used for heart rate variability analyses should be chosen in accordance with the purpose of data collection. To illustrate, the minimal sampling window size to estimate an indicator of stress by heart rate variability analysis (Vrijkotte et al., 2000) is 1 minute, but using heart rate variability as an indicator of physical fatigue (Ni et al., 2022), requires recommended window size of 5 minutes (Laborde et al., 2017; Malik, 1996). It may be noted that when multiple indices are to be derived from a single data source, sampling frequency and window size should be chosen such to accommodate the most stringent requirements.

It should be noted that there are issues with the specificity of physiological data. In the interviews conducted with driver monitoring system developers, it was noted that reduced heart rate and increased heart rate variability *could* mean that the driver is drowsy, but also that the driver is relaxed. The question of causality is a severe problem for psychophysiological data in general (Persson et al., 2020). Moreover, there is considerable interindividual variability in the manifestation of psychophysiological parameters, this implies that individual baseline data should always be considered when inferring physiological variables.

4.1.4. Estimation performance

Apart from evaluations of *classification* performance (see section 4.6.2.4), which involves the additional step of interpreting classifier input variables, the quality of the input variables themselves can be characterized in terms of *Accuracy*, referring to bias i.e., a systematic error between true and estimated value of a variable; the *Precision*, which refers to the reproducibility of estimates; *data-loss,* which refers to the amount of data lost in a sequentially sampled signal, and *latency*, which is the time it takes to produce an estimate (Holmqvist et al., 2023). The following general principles may be applied:

- *Accuracy* and *precision*. The accuracy and precision of a system affect subsequent classifier performance. In order to simultaneously minimize false positives and negatives and maximize true positives and negatives for a classifier relying on input data, the constant and variable error should be minimal. Exact tolerances depend on the application. For instance, errors in gaze estimates in the order of a few degrees may be acceptable when the data is used to assess whether the driver is looking at the road or not.
- *Data loss*. A system may sample data at a given frequency, but only yield useful estimates for a subset of those samples. Any data loss should not exceed the frequency at which derived measures (i.e., classifications) are evaluated to infer driver fitness. Systems should monitor the presence of reliable data and provide warnings when insufficient data is available.
- *Latency*. The time between an event and the corresponding response to that event produced by a computer system should be below the perceptual threshold for an intuitive understanding of the causal link between the event and the response to that event.

4.2. Distraction

In the most stringent automation mode (i.e., Continuous mediation), a driver is required to be able to take over driving tasks at any given instant, even in cases where automation fails without notifying the driver. To be able to do so, drivers must be physically available and have sufficient situational awareness to respond adequately at any time. Physical availability implies that sleep or high levels of drowsiness violate driver requirements (see section on fatigue); and the driver must be in a physical position to actually take over; that is, in future vehicles that accommodate the possibility of novel seating arrangements, the driver seat must face forward. In addition, drivers are not allowed to engage in many non-driving related activities, such as the use of mobile phones, tablets or laptops. At higher levels of automation, these requirements can be relaxed.

Situational awareness is an ambiguous and multifactorial construct, and a thorough assessment requires both a comprehensive characterization of the situation, and to peer into the mind of the driver to evaluate the extent to which their knowledge of the situation aligns with the ground truth. This is, at least currently, not technically possible. Instead, proxies must be used. Specifically,

1. awareness of the situation at the very least requires a driver to see the road and traffic ahead. That means that the driver must have their eyes on road most of the time (depending on information decay rate; Senders et al., 1967).
2. a driver/vehicle unit must have sufficient situation awareness, and depending on the automation mode, engagement in non-driving related tasks may be detrimental. Detection of many non-driving related activities provides an indication that the driver's situational awareness can be insufficient.

At present, the most practical method to classify whether the eyes are on road and whether non-driving related activities are performed is by means of computer vision. It should be noted that ultimately, the performance of these systems should be evaluated in terms of their ability to correctly discriminate between an attentive and distracted state. This depends not only on the ability of a system to detect gaze and non-driving related activities, but also on the scientific validity of the notion that eyes-off-road and engaging in non-driving related activities are indeed indicative of distraction, what a driver's affordances are in terms of sampling from essential sources of information such as the speedometer, and how these notions are affected by driving context. Nevertheless, in addition to the general recommendations for computer vision systems provided in section 4.1.1, the following recommendations are specifically applicable to distraction monitoring:

- The least obtrusive implementation for simultaneous gaze tracking and non-driving related activity detection is by cameras mounted in the vehicle cabin, monitoring the driver face and body.
- Human glance behaviours can use either an 'owl' strategy – turning the head in its entirety to focus on a target, or as following a 'lizard strategy' – turning the eyes, or a combination of the two. Drivers typically follow the lizard strategy for glances to objects close to the forward field of view, whereas the owl strategy tends to be followed for more distant glances. Methods that infer gaze direction from head pose are thus only suitable to track owl-like glances and impose the risk of missing cues present in eye-movements. Therefore, methods that infer gaze from tracking the eye directly should be preferred over head-tracking.
- The quality of gaze estimation improves with increasing camera resolution and frame rate. However, reasonable accuracy can be obtained even with low resolution, low frame rate cameras, as evidenced by studies that extract gaze from webcam data at a resolution of 640*480 pixels, and with a sampling frequency of 16 frames per second (Lin et al., 2013). Given the pace of technological advances, it is plausible that most currently commercially available camera systems exceed these specifications, and thus provide sufficient resolution and frame rate for eye-gaze tracking purposes. It may be noted that whereas low-quality systems may provide sufficiently detailed data for estimating gaze, these systems are not suitable for applications where high-speed eye-movements (i.e., saccades) need to be recorded, which precludes applications for, for instance, motion sickness detection.
- The system should be able to recognize conditions that indirectly imply that the driver is unable to see the road, for instance due to the obstructions of their field of view.

In addition to the performance measures outlined in section 4.1.1, the following additional considerations apply specifically to data acquisition for the purpose of inferring distraction from eye gaze.

- Accuracy and precision. As an example, method to classify distraction, the AttenD algorithm (Ahlstrom et al., 2013; Ahlström, Georgoulas, et al., 2021) classifies a driver as distracted when the gaze is outside the field relevant for driving. Bias (i.e., inaccuracy) in gaze estimation could inadvertently cause false alarms or misses, and noise (i.e., imprecision) contributes to data loss and reduces the reliability of the system overall. Notwithstanding, classification of distraction based on a general agreement of gaze with the part of the visual field related to driving, places much less stringent constraints on system accuracy and precision than other typical use cases for eye trackers such as when determining gaze targets. This implies that even systems that perform relatively poorly in terms of accuracy and precision (constant and variable errors in the order of degrees; Holmqvist et al., 2023), may still be sufficiently precise for classification of distraction based on gaze.
- Data loss. A system may sample gaze at a given frequency, but only yield useful estimates for a subset of samples. Potential causes are for instance a gaze outside the camera's field of view or blinking. The percentage of samples which do not yield useful estimates of gaze direction quantify the data loss. Blinking may account for approximately 2% of missing observations (Holmqvist et al., 2011). Any data loss should not exceed the temporal buffers for these to be updated sufficiently frequently. The system should be able to identify the cause of any data loss, in particular cases where the face is outside of the camera field-of-view, since this information can be used to infer that the eyes are not on-road.
- Latency. The time between an actual eye-movement and the output of the calculation to obtain the corresponding estimation is typically below 80 milliseconds (Holmqvist et al., 2022). This precedes any additional time it may take a system to issue a warning, and is

significant in urgent cases when considering that typical driver reaction times vary between 0.7–1.5 seconds (Green, 2000). It is known that a delay between a gaze shift and warning in the order of seconds causes drivers to misinterpret the warning as a false alarm (Fiorentino et al., 2023), that is, the causal link between the distraction event and issued warning will not be perceived. Therefore, very short latencies are required for distraction detection, and warnings should never be triggered when the driver is looking straight at the road.

Wherever available, the system should use both direct and indirect methods for highest accuracy and reliability of detecting driver disengagement. In other words, direct observations on gaze direction and non-driving related activities may be combined with inferences derived from vehicle data such as steering inputs.

4.3. Fatigue

As for distraction monitoring systems, requirements for fatigue monitoring systems are dictated by the driver requirements and affordances set by the automation modes. In Continuous mediation as well as in Driver standby modes, drivers should be alert. Conversely, in the Time-to-Sleep mode, people are allowed to sleep. Here, the challenge is to get people to recover from *sleep inertia* when driver input is again required (see section 3.1.2). Most effects of sleep inertia dissipate within 30 minutes, depending on prior sleep deprivations and the time of day in which a person awakens (Hilditch & McHill, 2019). A system being able to take these factors into account when deciding when to wake up the driver before a takeover request might be more feasible than monitoring sleep inertia in real-time.

Predicting future fatigue is essential since knowing the progression of driver fatigue is needed for long term planning on who (i.e., the driver or automation) should drive and when and who should drive. Biomathematical fatigue models based on information such as amount of sleep, activities, personal characteristics, time of the day and caffeine intake may predict periods of increased fatigue (Mollicone et al., 2019). These models can be used in combination with activity trackers to register prior sleep (McCormick et al., 2012). In predicting sleep, it is also necessary to take into account differences in the development of fatigue for different automation levels, since drivers using automation are likely to become fatigued faster than manual drivers (Schömig et al., 2015; Vogelpohl et al., 2019).

Fatigue can be task-related and sleep-related. The former is due to extended periods of high or low workload, while the latter is caused by a lack of sleep. Whereas these forms of fatigue have distinct causes and different remedies (see 3.1.2), their overt effects overlap. As of yet, there do not appear to be methods to determine the causality of fatigue symptoms through real-time driver monitoring systems. Therefore, the guidelines presented here address fatigue assessment in general, without discriminating its causes. Taken together, general requirements for fatigue monitoring systems in automated vehicles are that they should:

- Assess fatigue in real-time and classify a driver as either alert, drowsy or asleep.
- Forecast the progression of fatigue in awake drivers, to facilitate timing of upcoming transfers of control between human and automation.
- Account for effects of sleep inertia, in cases where the automation facilitates sleep.

Concerning methods to assess fatigue in real-time, the following considerations and recommendations should be considered:

- Eyelid parameters currently appear to be the most practical objective fatigue indicator. As an example, the PERCLOS (PERcentage of eyelid CLOSure) metric is often used to detect fatigue (Dinges & Grace, 1998; Trutschel et al., 2011). Other measures such as eye blinks, eye gaze patterns, and pupil diameter have also been suggested.
- Although environmental factors, intra-individual, and inter-individual differences obscure the relationship between fatigue and many physiological measures (heart rate, respiration, and heart rate variability etc.), these metrics have an advantage over camera-based systems since data are available also in cases where the face is not within the cameras field of view (blocked by a nomadic device, out of position during automation, etc.). Physiological metrics, measured via wearable devices, are therefore recommended as a complement to camera-based metrics.
- Since combinations of metrics appear to outperform measurements of PERCLOS alone (Bakker et al., 2021; Kerick et al., 2013), multi-modal and multi-feature approaches are recommended.
- There are individual differences regarding what should be considered a long blink duration, a low percentage of eye closure, a large heart rate variability, etc. Because of the large individual variability in how symptoms of fatigue express, fatigue detection systems should make use of personalised algorithms, especially when estimating moderate levels of fatigue. This may not be needed when estimating severe sleepiness or sleep where the fatigue indictors are more evident (difficulty keeping eyes open, long eye closures, swerving out of the lane).
- The latency (the time between the presence of fatigue and the detection of fatigue) in real-time fatigue detection based on eye movements depends on the sensitivity of the measurement method as well as the required time windows to detect sleepiness reliably. This delay can range between 10 seconds and 5 minutes (Borowsky et al., 2020). At lower levels of fatigue, timing is not critical and a delay of several minutes can be tolerated. For microsleep events in a manual driving setting, the required latency is in the order of a few seconds.
- In terms of validity, the method of real-time fatigue assessment that has the strongest evidence in scientific literature is electroencephalography (EEG). However, such systems require an EEG sensor positioned on the head, which is obtrusive even with wearable dry electrode solutions. Moreover, these systems are mostly used in research settings which allow for much better control of surrounding electrical fields that affect EEG measurements, and where people are not subject to vibrations and accelerations as they are in a vehicle. This reduces the usefulness of EEG measurements in daily practice.

As for the validation of fatigue monitoring systems:

- Subjective sleepiness ratings can be used to validate fatigue measurement methods. Self-assessment scales are often used as the ground-truth when designing sleepiness detection systems. The most used subjective fatigue and sleepiness indicator is the Karolinska sleepiness scale (Åkerstedt et al., 2014). It is recommended that fatigue detection systems are tuned to avoid false negatives corresponding to Karolinska sleepiness scale levels 8 and 9.
- Whereas driving behaviour can obviously not be used as a sleepiness indicator during automated driving, it can be used to validate monitoring systems, next to observer ratings or self-assessment scales. Indicators of driving behaviour that indicate sleepiness are lane- and speed keeping and steering wheel movements (Liu et al., 2009). It is recommended that fatigue detection systems are tuned to avoid false negatives in connection to sleep-induced line crossings.

4.4. Comfort

Subjective measures are still the gold standard when assessing the levels of (dis)comfort in experimental research (Song & Vink, 2021). Objective assessment of comfort requires more research and development to enable reliable, accurate, unobtrusive, real-time detection of discomfort based on sensor data. It is therefore difficult to provide clear guidelines on real-time comfort monitoring.

Estimations of discomfort based on lookup tables with potentially uncomfortable driving situations is more manageable. Instead of directly measuring the comfort level, discomfort is inferred from contextual variables derived from digital maps, traffic management systems, weather services, other driver state monitoring systems, etc. Upcoming situations with high a-priori risk of inducing discomfort may then be mitigated by suggesting a take-over from manual to automated driving (or vice versa) where such transfers are possible.

General requirements for comfort assessment systems in automated vehicles include:

- Safety should precede preference and comfort. For example, lane keeping assistant systems and adaptive cruise control systems may monitor driver behaviour during manual driving to infer preferences in lateral lane position and following distance to leading vehicles. Adopting such preferences will likely improve subjective experiences and comfort. However, such personalisation can have a negative impact on safety, for example if the systems adapt to an aggressive driving style, or if the automated functions start to drive less predictably (from an outside observer's point of view).
- Indirect comfort prediction, based on external data sources and lookup tables with a-priori defined uncomfortable situations can increase comfort by proposing timely take-overs, either by the vehicle or by the driver. An example is to propose automated driving when approaching an upcoming traffic jam, where the driving task is typically tedious. Information about how drivers activate or deactivate automation systems can be used to personalise the take-over suggestions.
- Measurements of acceptance or rejection of take-over suggestions can also be used to personalise the threshold for minimum automation availability durations. For example, if the vehicle reaches a road stretch that affords Driver standby, but only for 1 minute, then this 1-minute stretch could be supressed by setting it to "not available". The personalisation step would then be to adapt the maximum x-minute stretch that is supressed.
- Context and driver state information can be used to increase comfort by adapting the timing of communication between the vehicle and the driver. This is similar to a workload manager, with the difference that it tries to avoid discomfort rather than overload.

4.5. Techniques for prediction of (near-)future driver states

A central concept in MEDIATOR is that of time-to-driver-(un)fitness. Accurate forecasts of such time budgets are critical when deciding if and when either the driver or automation should control the vehicle. Obtaining predictions on these time budgets can be achieved in various ways, and to stimulate innovation it may not be pertinent to prescribe particular forms of the implementation. Notwithstanding, in the Mediator system, the decision logic worked by means of decision trees (Bakker et al., 2023). The decision trees use current estimates of distraction (see section 4.2) and fatigue (section 4.3) obtained through driver monitoring systems and combine these with time budget affordances based on expert opinion to predict the time-to-driver-(un)fitness. These budgets allow for personalization and can also consider recent driver histories of the amount, timing, and quality of sleep, the time since the last sleep period, time of day, workload and time on task, which is for instance also done to determine pilot fitness-for-duty in aviation (FAA, 2013).

Prediction of distraction could also be based on monitoring of the information systems. For example, when drivers receive a message, it is likely that they will become tempted to read the message. Trials conducted in MEDIATOR demonstrated that people appreciated a proposal to switch to a higher automation mode when they were classified as distracted based on gaze estimates (Fiorentino et al., 2023). A further improvement in terms of road safety and possibly comfort may be achieved if automation is proposed even before a driver gets distracted, as sort of a proactive workload manager triggered when, for instance, a message is received. Further methods that may prove useful for more accurate predictions of time budgets are cybernetic models founded in control theory that predict the progression of driver states. An example applicable to sickness is the Oman model of motion sickness progression (Oman, 1982), which can generate predictions of future motion sickness taking into account personal sensitivity data and knowledge of the vehicle route (Irmak et al., 2021).

From chapter 3 it is clear that many challenges remain before driver monitoring reach high accuracy and reliability for all individuals. Predicting future occurrences of impairments, or the development of an ongoing impairment, is even more challenging. For a prediction to be useful, it must be accurate both in terms of the actual driver state and in terms of temporal accuracy. For distraction, it is not possible to estimate the onset of the next distraction event. However, given information about ongoing non-driving related activities, it may be possible to give an estimate of the time until the driver becomes attentive again, based on statistics of typical task completion times. For fatigue, the development of sleep-related fatigue can be predicted with biomathematical models, but the accuracy of such models is not very good on an individual level (Van Dongen, 2004). There are biomathematical models describing task-related fatigue as well, taking task complexity and driving time into account, but again, the accuracy is low on an individual level. The challenges are similar for comfort prediction, where future discomfort can be estimated based on a-priori information about upcoming uncomfortable situations (see further section 3.2.3.3).

Taken together, predictions of future driver states are uncertain and should not yet be used to make safety critical decisions. Areas of application where predictions can be used is for example a situation where the vehicle proactively suggest that the driver should switch to automation when approaching a traffic jam. This provides added value for the driver without compromising safety, under the assumption that the traffic jam assist function is as safe as a human driver.

4.6. Evaluation procedures

Existing regulations typically specify that vehicle manufacturers must provide type approval authorities with documentation that describes the technical implementation of systems in sufficient detail for the authority to assess the robustness and functioning of the system. However, in lieu of technical regulations or guidelines, ultimate proof that a driver monitoring system meets requirements is in testing classification performance for a representative sample of human participants.

We propose that driver monitoring system testing should be conducted stepwise for (1) data quality and availability, (2) driver impairment detection performance and (3) intervention strategy, based on a combination of theory and realistic expectations. Data quality and availability are influenced by the ability of the sensors to track the driver's features reliably under as many different circumstances as possible, including variations in personal characteristics, clothing and accessories, weather related factors, etc. Using these data as input, algorithms estimate the driver's level of impairment. Ensuring good feature quality is the basis for the algorithms to be able to function properly. The impairment detection algorithms can vary in complexity, in the type and amount of information they consider, and in the specificity of their output. Finally, the way in which the driver monitoring outcome is used for intervention – with direct warnings or as modifier of other systems – can differ. Evaluation procedures are needed for all three levels to effectively detect and prevent potential weaknesses in the systems. Step (3), evaluation of the intervention strategy, is covered in Mediator Deliverable D4.2 (van Grondelle, 2023) and will not be described here.

4.6.1. Evaluation of feature data quality

Since the number of combinations of idiosyncratic properties quickly becomes unfeasible large, it is recommended that data quality evaluations at this level are tested in a controlled setup, in a stationary vehicle or even in a lab setup.

The extent to which idiosyncratic properties must be represented in the sample may depend on the type of classifier. For instance, to mitigate risks of bias in AI, machine learning techniques must be trained on representative data sets. In contrast, this is not the case for expert systems.

4.6.1.1. Evaluation using pretended/acted impairments/behaviours

As a starting point, the systems should be tested to verify that they can measure the base features that they are intended to measure. As a first step, this can be done by asking participants to blink slowly (verify eyelid opening metrics), to yawn and talk (mouth detection), to look in various direction (gaze direction) including over the shoulder (head tracking), and to use mobile phones etc. (driver activity recognition). Similarly, physiology-based systems should be tested to verify for example heartbeat and respiration detection. If the driver monitoring system is based on an end-to-end system this step can obviously not be done.

4.6.1.2. Study population

To ensure high quality and data availability for all potential users, the systems must be tested on a broad range of the population. Therefore, people of different ages, genders, weight, height, body types, skin tone, ethnic groups, eye shapes, facial hair, etc., should ideally be included. Exempt cases are for instance children, as the typical minimum age at which one is allowed to drive is between 16–18 years. Individuals showing nonconformity with stereotypes (corner cases) should be considered carefully. In case of occupant monitoring, participants of all ages, including infants, should be included.

4.6.1.3. Facial occlusions and body movement

The effect of facial occlusions such as face masks, hoods, hand activities, glasses, and phones or laptops, should be evaluated. The latter is especially relevant in vehicles with automated driving functionalities. Also, detection and tracking of hands for recognition of driver distraction actions is a challenging problem. “Difficult” lenses and IR-blocking sunglasses should be included in the evaluations.

Similarly, physiological sensors should be evaluated for motion artifacts in different body positions and under various degrees of motion. Non-contact sensors, for example in the seat or the seatbelt, should also be tested with different types of clothes.

4.6.1.4. Light conditions

There are frequent and sudden variations of lighting in real-life driving. These changes happen quickly and depend on daytime (day/night), weather, driving environment (streets lined with trees, driving under a bridge) and artificial light (headlights, street lighting). Such conditions should be evaluated if the system use a camera sensor. These tests should also monitor for unintended side effects such as squinting in strong sunlight that causes the lower eyelid and eyelashes to occlude the pupil partially or fully, making it difficult to track the eyes.

4.6.2. Evaluation of impairment detection performance

Evaluation of detection performance should be based on actual, not acted, impairments. It should also be conducted in more ecologically valid environments rather than in a lab/office setting.

4.6.2.1. Context and road environment

Studies should be performed under various ambient conditions that may be encountered during driving. This may include, but is not necessarily limited to, day versus night-time driving, varying ambient lighting conditions, and driving on urban, rural and highway roads, which vary in speed, presence and type of other road users, and also the likelihood of distractions and fatigue occurring.

4.6.2.2. Controlled experiment versus real life

Findings obtained in experimental studies performed under laboratory conditions are not necessarily representative of behaviours in naturalistic driving. An illustration is the Hawthorne effect (Landsberger, 1958). Participants in driving studies may feel obliged to behave in some way differently than they would in real life. In addition, it is plausible that an inverse relation exists between the level of experimental control and ecological validity of a study. Further, confounding factors are limited by design in controlled experiments (Lu, 2009). For example, a homogenous study population in combination with strict experimental design where only sleepiness is allowed to vary will for sure find an effect of sleepiness on heart rate variability, but this does not mean that the same change in heart rate variability is caused by sleepiness in an uncontrolled situation, where the given change can also be due to some of the other confounding factors that also affect heart rate variability. The link between cause and effect is often weakened when moving from a controlled to a naturalistic setting. Therefore, it is desirable to also test systems in naturalistic driving studies or field operational tests which approximate real driving conditions as closely as possible. Alternatively, if simulation or simulator studies are used, the validity of the simulator as an alternative to naturalistic driving studies must be established, and/or rules should be derived which allow estimation of the real-world outcome given findings obtained in the simulator (Fors et al., 2018; Talsma et al., 2023).

4.6.2.3. Ground truth, annotations, and labelling

Experimental conditions may be designed such that they induce certain driver states (Wörle et al., 2023). For example, to manipulate fatigue, researchers can include conditions where participants are sleep deprived versus fully alert. Experimental manipulation of distraction is possible by manipulating the presence and/or workload associated with non-driving related activities. Similarly, discomfort can be induced by uncomfortable driving situations. In addition, it is desirable to quantify the extent to which the experimental manipulations indeed had the desired effect. For fatigue, study participants may report fatigue symptoms by means of the Karolinska sleepiness scale. For distraction, sufficient situation awareness can be verified via hazard perception paradigms, and non-driving related tasks can be quantified as device usage and task performance. For comfort, drivers can indicate their level of comfort in real-time by using for example pressing a button when they experience discomfort. It may be desirable to account for covariates such as workload and motivation, whether it is to remain alert, engagement in non-driving related tasks or the driving task itself. This can be achieved by motivating instructions and scenario descriptions, via incentives, or similar experimental manipulations such as the choice of non-driving related tasks. Like the assessment of the constructs that are of primary interest (i.e., drowsiness, inattention), questionnaires can be used to evaluate the effectiveness of these motivating manipulations, to subsequently account for possible mediating effects in the statistical analyses.

4.6.2.4. Detection performance

To assess the quality of detection algorithms, metrics are required that reflect their performance. These metrics may be constructed using knowledge of the ground truth, as discussed in section 4.6.2.3. and the estimates obtained from a classification algorithm. Given a ground truth and an estimate, where in the dichotomous case, both are either ‘positive’ or ‘negative’, a confusion matrix can be established. This confusion matrix specifies the number of observed combinations of possible ground truths versus the predicted states. For a single observation on a dichotomous variable, there are four possible outcomes: the ground truth and estimate are both negative (true negative), the ground truth is negative but the estimate is positive (false positive), the ground truth is positive but the estimate is negative (false negative), or both the ground truth and estimate are positive (true positive). A perfect classifier will have 100% true negative rate (specificity) and 100% true positive rate (sensitivity). Although there are algorithms that provide dichotomous estimates directly, many implementations instead yield an estimate of the probability of a positive outcome. The final estimate is then made by setting a criterion threshold for the estimated probability below which a negative classification is made, and above which a positive classification is made. By varying the criterion, the classifier’s specificity and sensitivity will change. This interplay is reflected in the so-called Receiver Operating Characteristic curve. The resulting curve will lie somewhere between the identity line for a classifier that does not perform better than chance, and a line where the true positive rate equals 1 for all false positive rates. Hence, the area under the receiver operating characteristic curve reflects the classifier’s performance, varying between 0.5 for a classifier that does not perform better than chance, and 1 for a perfect classifier. As guideline on the interpretation of the obtained area under the curve, values between 0.7–0.8 can be considered acceptable performance, 0.8–0.9 can be considered excellent performance, and anything above 0.9 is outstanding performance (Hosmer Jr et al., 2013). In practice, a more nuanced evaluation could be implemented by also considering the relative importance (utility) of different outcomes in the form of a cost function which combines the probabilities of the alternative cells of the confusion matrix with their utility, and an optimal criterion may be chosen by optimizing this cost function.

Ultimately, the goal of a Mediator system is to provide safety. Classifications of driver state variables provided by driver monitoring system will be used to inform a decision logic that mediates between the driver and vehicle on who is most fit to drive. Achieving this goal depends not only on performance of the driver monitoring system, but also on the intervention strategy and how the intervention is communicated to the driver. Guidelines on intervention strategies established in the Mediator project are described in Mediator Deliverable D4.2 (van Grondelle, 2023). A comprehensive systematic fatigue risk management framework that addresses prevention, monitoring, and mitigation of fatigue-induced risks during on-road long-term testing of automated driving functions can be found in Favaro et al. (2022). It is it is noteworthy that real-time driver monitoring is but 1 out of 20 countermeasures used to counter fatigue-induced risks.

4.6.3. Ethics

Ethical standards for any research involving human participants are generally aligned with the Declaration of Helsinki, which outlines ethical principles for medical research involving human subjects[1]. A guiding principle is that the gained knowledge should outweigh any risk or discomfort while not taking precedence over individual rights and interests. For a driver monitoring solution, this means the system should work equally well for different types of people, and that the system is affordable and cost-effective. For evaluation purposes, an independent ethical review committee should judge the study, prior to the study being conducted. We recommend that all driver monitoring system evaluations involving study participants undergo an independent ethical review, to ensure that the study proposal, the information provided for participants, the informed consent form, and the data management plan aligns with the declaration of Helsinki. The information provided to the ethical review committee must be sufficiently detailed for the committee to decide whether it is pertinent that the research question be answered, whether the study design and number of participants are appropriate and theoretically sufficient to answer the research question, and whether the risk to participants is acceptable.

4.6.4. Data Protection

As of May 25, 2018, any case where personal data is collected must, by law, adhere to the General Data Protection Regulation[2]. This thus applies to implementation of driver monitoring system in vehicles, as well as to studies designed to evaluate these systems. In general, data should only be collected that is essential to a particular purpose and should be stored securely in anonymized form, meaning that any information that directly identifies, or which may be used to indirectly identify an individual person must be discarded. For driver monitoring solutions, this implies that personal data such as video should not be stored beyond its need for detection and prediction purposes. For evaluation studies, researchers must formulate a data management plan. This plan shall list all data types collected within the scope of the study, the purpose of their collection and their format. It should also detail who will be able to access the data and any used anonymization methods. A data protection officer should be appointed, who is the contact point for any concerns regarding data usage.

Data protection notwithstanding, there is a growing consensus that data should be shared in order to maximize their utility. In MEDIATOR, the FAIR principle was adopted, which refers to data being Findable, Accessible, Interoperable and Reusable. In practice this means that anonymized and well annotated data should be shared in public repositories. Numerous public repositories exist, and journals regularly offer options to include data in the form as online supplementary material to scientific articles. A global registry of research data repositories is for example made available by the 're3data.org' initiative (Pampel et al., 2013), and a list of various repositories is also made available in the Open Access Directory[3].

[1] Declaration of Helsinki: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/
[2] EU (2016) GDPR - General Data Protection Regulation: https://gdpr.eu/tag/gdpr
[3] http://oad.simmons.edu/oadwiki

5. Discussion

The presented guidelines integrate the state-of-the-art according to the literature with practical results from MEDIATOR. Because scientific knowledge and technical possibilities continually evolve, insights into the relation of latent constructs that contribute to driver fitness and the best way to operationalize, measure, and predict these constructs are subject to change. Therefore, suggested methodologies and recommendations on their implementation are best considered a starting point, or a means to an end, rather than a definitive conclusion on the optimal way to implement driver monitoring systems.

It should be noted that much of the knowledge that is needed to formulate definitive guidelines with a reasonable scientific validity is still lacking, and that more (empirical) research is required to achieve a consensus on the operationalisation of factors that contribute to the construct of driver (un-)fitness, their interactions, and the accuracy and precision of driver monitoring systems intended to estimate these constructs. Also, it should be noted that it may not be desirable to prescribe technical implementations to measure factors that contribute to driver (un-)fitness, because this may hamper innovation. Suggestions have been made based on existing knowledge, but a guiding principle for any evaluation should be classification performance.

Notwithstanding, the guidelines described in this report can be summarized as a set of high-level guidelines, or principles, and a set of more specific, tangible guidelines.

General principles:

- *Minimally obtrusive sensors*. Camera-based systems have several advantages here, since they have the potential to capture rich information about humans, objects, and their interaction. Unobtrusive sensing is needed to facilitate high adoption rates, to avoid deactivation, and to avoid interfering with drivers' operation of the vehicle.
- *Real-time operation and timely detections*. Impairment detection, and subsequent interventions, have different demands on acceptable latencies. Detection of early signs of fatigue is not time critical (order of minutes) while severe fatigue, microsleep, and long off-road glances are time critical (order of seconds or less). In some situations, discomfort can be detected offline several minutes in advance, for example when approaching harsh weather or a traffic jam. Proactive impairment interventions, in contrast to reactive detection/intervention, is favourable. This requires forecasts of drivers' future readiness levels.
- *Robustness to environmental conditions.* System performance should not be significantly influenced by environmental conditions such as traffic, landscape, weather, and darkness.
- *Automation level dependent*. The drivers' responsibilities change with the level of vehicle automation, which in turn affects the requirements for a driver monitoring system. As an example, continuous distraction detection is highly relevant in manual and assisted driving. In higher levels of automation, where non-driving related task engagement is allowed, it is sufficient to ensure that the driver is attentive in relation to transitions of control.
- *Situational awareness*. A driver/vehicle-unit should have sufficient situational awareness to be able to drive safely. With higher levels of automation, the responsibilities for situational awareness are gradually shifted from the human to the vehicle. Similarly, to be able to provide relevant impairment detections, a driver monitoring system should also be situationally aware and take contextual factors into account. For example, fatigue warning systems would benefit from knowledge about sleep history and driving time, and distraction

detection systems would benefit from knowledge about which areas in the surroundings that needs to be sampled to gain sufficient situational awareness.
- *Ecological validity*. Final evaluations/testing of driver monitoring systems should be conducted in ecologically valid settings with naturalistically induced impairments. Lab testing can and should be used in earlier evaluation stages, for example, when testing if an eye tracking system provides high quality tracking throughout a broad range of the population.
- *Minimal intrusion on privacy*. Driver monitoring systems should avoid privacy intrusions. For example, video data should be deleted continuously and should not be stored beyond what is needed for impairment detections.

More tangible guidelines include:
- *Sensor related aspects*:
  - Driver monitoring systems are not perfectly reliable. Therefore, critical, safety-related systems should function independently from these systems.
  - Driver monitoring systems should self-diagnose, and the reliability of classifications should be considered by the system. Reliability may be improved by integrating information from redundant sources of information, and by combining complimentary information. For safety and for protection against interferences, a separate Controller Area Network bus is recommended.
  - Computer vision-based driver monitoring systems primarily rely on information derived from the eyes. To ensure that video data is suitable for subsequent inferences and robust to disturbances, it is recommended to use multiple infrared cameras, mounted such that the driver's face is in the cameras' sweet spot (typically 60-70cm from the face).
  - Vehicle based data, such as steering inputs, swerving, and lane departures, directly reflect safety critical performance, but is limited to manual driving.
  - Physiological data have limited specificity but can augment other sources of information. Contact sensors may be placed in the steering wheel to, for example, measure skin conductance, and sensors in the seat or seatbelt can be used to measure heart rate and heart rate variability.
- *Estimation performance*: classifier performance depends on the quality of input data. For driver monitoring systems, input data are, for instance, gaze direction and eye closure, which are inferred from sensor readings. The quality of input data can be expressed in terms of *accuracy*, *precision*, *data loss* and *latency*. Threshold values for these metrics depend on application and situation.
- *Distraction*: a state of distraction is typically inferred from gaze direction and engagement in non-driving related activities. These can be determined by computer vision systems. Gaze direction should be inferred from directly monitoring eye movements rather than head direction, because the eyes can move within the head. Typical system requirements are a resolution of at least 640*480 pixels and a sampling frequency of 16Hz, with a system latency in the order 0.1s.
- *Fatigue*: typically inferred from eye closure. As for distraction, the most practical method to do so is by computer vision-based systems. This method can be augmented using vehicle-based data, as lane- and speed-keeping and control inputs are indicative of fatigue as well. Forecasting fatigue is critical for higher automation levels, to ensure that future take-over requests can be met. To this end, information such as driver sleep history, time of day, and driving time, have found to be predictive.
- *Comfort*: methods for real-time monitoring of driver comfort are still in their infancy. Current methods rely on inference of emotions from elements of facial expressions. However,

comfort may also be inferred from a-priori knowledge of (un)comfortable situations and previous preferences or may be predicted on the basis of knowledge of the road ahead. For instance, motion sickness may be mitigated by suggesting manual driving on provocative routes.
- *Prediction*: a critical concept in automation is estimation of 'time budgets', which reflect how long a driver can be relieved from the driving task. This requires forecasts of driver states.
- *Evaluation*: driver monitoring system performance is expressed in terms of the ability to correctly classify states. Due to the multitude of factors that affect performance, it should be evaluated empirically. Such evaluations should be ecologically valid, in that they reflect the range of situations and users that can occur in real-life. Performance should be expressed in terms of classifier *sensitivity* and *specificity.*

Whilst it remains to be seen whether interior sensing will lower road deaths, the use of emotional AI and other biometric profiling raises other high-level societal risks (McStay & Urquhart, 2022). On one hand, the EU Vehicle Safety Regulation welcomes safety technologies and driver monitoring solutions, but at the same time, other regulations such as GDPR flag human-state measures and emotion profiling as risky. New guidelines must strike a balance between the two, weighting privacy risks versus safety benefits. For instance, a large online retailer that equips its vehicles with driver monitoring systems do not only use the system to reduce dangerous behaviours, but also to score and penalize personnel. Such use of DMS may be considered excessive, or even 'Orwellian', especially considering that these systems can be inaccurate[4].

Another practical consideration on driver monitoring systems is the added costs for vehicles. For vehicle automation to be broadly adopted by consumers, the technology should be attractive -it must offer readily apparent advantages over manually controlled vehicles and should also be affordable. A synthesis of multiple surveys (Elvik, 2020) indicates that a majority of consumers may not be willing to pay extra for (fully) automated vehicle functionalities. This synthesis also shows that there is considerable variability in what consumers are willing to pay for vehicle automation, with differences between countries, but also with skewed distributions within countries. Whereas some consumers are willing to pay more than the estimated added costs of automated vehicles ($10,000 US), up to 60% of people are not willing to pay any premium for automated vehicle functionality. It is thus likely that a majority of consumers will initially find automated vehicles too expensive. However, the price of automated vehicles can be expected to fall as technology matures and vehicles are manufactured in larger numbers.

The consensus, supported by accident statistics, is that the driver state variables of fatigue, distraction, (sudden) sickness, intoxication, and as a mediating factor, comfort, affect driver fitness. However, there are various outstanding issues in the science behind this.

- First, the definition of the constructs is a topic of active scientific debate. For example, some people argue that distraction is a form of inattention, whereas others say it is the other way around (Regan et al., 2011), and yet others claim that it is futile to define distraction without first sorting out what it means to be attentive (Kircher & Ahlstrom, 2016).
- Given that these are latent psychological constructs, there typically is not a single perfect measure that reflects their level. That is, their operationalization (i.e., how constructs, provided that they are well-defined, are best measured), may be inherently vague.

[4] The Telegraph 2022: https://www.telegraph.co.uk/business/2022/05/22/amazon-installs-ai-cameras-monitor-delivery-drivers/

* Numerous directional and mediating effects between the latent constructs can be postulated. As an illustration, Figure 2.1 graphically represents the relation between variables as they are considered in the present report. Note that this representation is likely to be an oversimplification.
* In addition to other or additional relations between the state variables considered here, others may prove to be of importance to the notion of driver fitness. For example, driver experience and attitude to risk are not considered, but are known to affect behaviour (Hatfield & Fernandes, 2009).
* Although fatigue, distraction, (sudden) sickness and intoxication are indicated in a large percentage of crashes, it is not known how an encompassing construct of fitness is predictive of safety; nor, for that matter, is it apparent how 'safety' should be operationalized precisely. Nor is the impact of a driver monitoring system on safety known. Here one may consider, for example, the likelihood of *potentially* dangerous situations; the likelihood of minor crashes; and/or the likelihood of major crashes.

5.1. Future research

The information presented in these guidelines outline a framework on how certain driver state variables can be measured, and under which conditions these systems should be evaluated to increase the likelihood that they will operate as intended in practice. However, apart from some minimum requirements for sensors to ensure that useful data are obtained, neither the literature nor results from MEDIATOR provide definitive answers on what is best practice, considering multiple alternative methodologies. In particular, the following research questions should be addressed in the coming years:

* How can driver monitoring data be fused with external data such as digital maps and proximity data to determine if the driver is sufficiently aware of the surroundings given the current automation level?
* How can multiple driver impairments be assessed simultaneously in uncontrolled environments with multiple confounding factors?
* How can the cause of a detected driver impairment be determined? For example, is the driver fatigued due to sleepiness, underload, or overload. The distinction is important for correct countermeasure deployment.
* How should comfort and emotion assessments go beyond stereotypical facial expressions analysis and reveal the psychological meaning of those signals?
* How can various driver impairments be forecasted with sufficient accuracy and foresight to be useful when deciding if and when either the driver or automation should control the vehicle?
* How do different driver impairments and their respective indicators, including all interactions, contribute to driver fitness?
* How does driver fitness impact safety and how well should driver fitness be known to ensure safety (split out by different state variables)?
* How should new hardware and data/sensor fusion techniques be used to improve accuracy and availability/uptime in case of motion artifacts, camera view obstructions, out of position scenarios, interference, etc.? What other considerations might contribute to the choice of different assessment sensors/techniques (e.g., cost)?

6. Conclusion

The presented guidelines integrate state of the art knowledge from the literature with knowhow from the industry and practical results from the Mediator project. In summary, a driver monitoring system should use minimally obtrusive sensors, operate in real-time with minimal latencies, be robust to environmental conditions, be adaptive to situational circumstances such as context, road environment and automation capabilities, and be respectful to privacy concerns. Final evaluations of driver monitoring systems should be conducted in ecologically valid settings, to ensure that sufficient performance is achieved under these conditions.

The goal of driver monitoring is to increase road safety. Achieving this goal depends not only on performance of the driver monitoring system, but also on the intervention strategy and how the intervention is communicated to the driver. We believe that this is best achieved by combing warning and intervention strategies such as, for example, adapting the sensitivity of driver assistance systems when a driver is not attentive.

Though knowledge is still lacking to be able to formulate definitive operational guidelines on driver monitoring systems, and though there are still limitations to the capabilities of driver monitoring systems, this should not prevent or delay their introduction in new vehicle types. Instead, available technologies should be used to address and mitigate impairments to the extent possible, starting with severe behaviours such as alcohol intoxication, microsleeps, long glances away from the road and incapacitation.

7. References

Ahlstrom, C., Kircher, K., & Kircher, A. (2013). A gaze-based driver distraction warning system and its effect on visual behavior. *IEEE Transactions on intelligent transportation systems*, *14*(2), 965-973.

Ahlstrom, C., Victor, T., Wege, C., & Steinmetz, E. (2011). Processing of eye/head-tracking data in large-scale naturalistic driving data sets. *IEEE Transactions on intelligent transportation systems*, *13*(2), 553-564.

Ahlström, C., Georgoulas, G., & Kircher, K. (2021). Towards a Context-Dependent Multi-Buffer Driver Distraction Detection Algorithm. *IEEE Transactions on intelligent transportation systems*, *23*(5), 4778-4790.

Ahlström, C., Kircher, K., Nyström, M., & Wolfe, B. (2021). Eye Tracking in Driver Attention Research—How Gaze Data Interpretations Influence What We Learn [Perspective]. *Frontiers in Neuroergonomics*, *2*.

Ahlström, C., Wörle, J., Ljung Aust, M., & Diederichs, F. (2023). Road vehicle automation and its effects on fatigue, sleep, rest and recuperation. In A. J. Filtness & C. M. Rudin-Brown (Eds.), *The Handbook of Fatigue Management in Transportation - Waking Up to the Challenge*. CRC Press.

Ahlström, C., Zemblys, R., Jansson, H., Forsberg, C., Karlsson, J. G., & Anund, A. (2021). Effects of partially automated driving on the development of driver sleepiness. *Accident Analysis & Prevention*, *153*, 106058.

Albadawi, Y., Takruri, M., & Awad, M. (2022). A review of recent developments in driver drowsiness detection systems. *Sensors*, *22*(5), 2069.

Anjani, S., Kühne, M., Naddeo, A., Frohriep, S., Mansfield, N., Song, Y., & Vink, P. (2021). PCQ: Preferred Comfort Questionnaires for product design. *Work*, *68*(s1), S19-S28.

Ansari, S., Du, H., Naghdy, F., & Stirling, D. (2023). Factors influencing driver behavior and advances in monitoring methods. In *AI-enabled Technologies for Autonomous and Connected Vehicles* (pp. 387-414). Springer.

Athmer, C., Spaan, M. T. J., Li, Y., Liu, Y., & Bakker, B. (2023). *Simulation of Decision Making in the Mediator System. Deliverable D3.2 of the H2020 project MEDIATOR*.

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *NeuroImage*, *59*(1), 36-47.

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

Bakker, B., Knauss, A., Rodrigues de Campos, G., Tanov, N., Andersson, S., Mano, D., . . . Enhuber, S. M. (2023). *Functional and technical specifications of the Mediator system. Deliverable D2.11 of the H2020 project MEDIATOR*.

Bakker, B., Zabłocki, B., Baker, A., Riethmeister, V., Marx, B., Iyer, G., . . . Ahlström, C. (2021). A multi-stage, multi-feature machine learning approach to detect driver sleepiness in naturalistic road driving conditions. *IEEE Transactions on intelligent transportation systems*, *23*(5), 4791-4800.

Barnden, C. (2019). *ADAS & Automotive IR Vision Driver Monitoring Systems (2020 Edition). Market Analysis and Forecasts to 2026.*

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, *20*(1), 1-68.

Beggiato, M. (2015). *Changes in motivational and higher level cognitive processes when interacting with in-vehicle automation* University of Technology Chemnitz].

Beggiato, M., Hartwich, F., & Krems, J. (2018). Using smartbands, pupillometry and body motion to detect discomfort in automated driving. *Frontiers in human neuroscience*, *12*, 338.

Beggiato, M., Rauh, N., & Krems, J. (2021). Multi-camera Face Tracking for Estimating User's Discomfort with Automated Vehicle Operations. Human Systems Engineering and Design (IHSED2021): Future Trends and Applications. AHFE (2021) International Conference.,

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

Berntson, G. G., Bigger, J. T., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., . . . VanderMolen, M. W. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, *34*(6), 623-648.

Bjorvatn, A., Page, Y., Fahrenkrog, F., Weber, H., Aittoniemi, E., Heum, P., . . . Torrao, G. (2021). *L3Pilot Deliverable D7.4: Impact evaluation results* [Report].

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

Borowsky, A., Schwarz-Chassidim, H., Hollander, C., Rauh, N., Enhuber, S., Oron-Gilad, T., & Beggiato, M. (2023). *Results of the MEDIATOR driving simulator evaluation studies . Deliverable D3.3 of the H2020 project MEDIATOR*.

Bryant, D. A., & Howard, A. (2019). A comparative analysis of emotion-detecting AI systems with respect to algorithm performance and dataset diversity. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society,

Carsten, O., & Martens, M. H. (2019). How can humans understand their automated cars? HMI principles, problems and solutions. *Cognition, Technology & Work*, *21*(1), 3-20.

Cicchino, J. B. (2017). Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates. *Accident Analysis & Prevention*, *99*, 142-152.

Cicchino, J. B. (2018). Effects of lane departure warning on police-reported crash rates. *Journal of safety research*, *66*, 61-70.

Cicchino, J. B. (2022). Effects of automatic emergency braking systems on pedestrian crash risk. *Accident Analysis & Prevention*, *172*, 106686.

Cori, J. M., Manousakis, J. E., Koppel, S., Ferguson, S. A., Sargent, C., Howard, M. E., & Anderson, C. (2021). An evaluation and comparison of commercial driver sleepiness detection technology: a rapid review. *Physiol Meas*, *42*(7).

de Winkel, K. N., Irmak, T., Happee, R., & Shyrokau, B. (2023). Standards for passenger comfort in automated vehicles: Acceleration and jerk. *Applied Ergonomics*, *106*, 103881.

de Winkel, K. N., Pretto, P., Nooij, S. A., Cohen, I., & Bülthoff, H. H. (2021). Efficacy of augmented visual environments for reducing sickness in autonomous vehicles. *Applied Ergonomics*, *90*, 103282.

de Winter, J., Stanton, N. A., Price, J. S., & Mistry, H. (2016). The effects of driving with different levels of unreliable automation on self-reported workload and secondary task performance. *INTERNATIONAL JOURNAL OF VEHICLE DESIGN*, *70*(4), 297-324.

Delice, I. I., & Ertugrul, S. (2007). Intelligent modeling of human driver: A survey. 2007 IEEE intelligent vehicles symposium,

Dinges, D. F., & Grace, R. (1998). *PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance* (FHWA-MCRT-98-006).

Domeyer, J., Alsaid, A., Liu, S.-Y., & Lee, J. D. (2019). Passenger emotional response type and timing during automated vehicle intersection negotiation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,

Doudou, M., Bouabdallah, A., & Berge-Cherfaoui, V. (2019). Driver drowsiness measurement technologies: Current research, market solutions, and challenges. *International Journal of Intelligent Transportation Systems Research*, 1-23.

Dunn, N. J., Dingus, T. A., Soccolich, S., & Horrey, W. J. (2021). Investigating the impact of driving automation systems on distracted driving behaviors. *Accident Analysis & Prevention*, *156*, 106152.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues* (Vol. 10). Ishk.

El Khatib, A., Ou, C., & Karray, F. (2019). Driver inattention detection in the context of next-generation autonomous vehicles design: A survey. *IEEE Transactions on intelligent transportation systems*, *21*(11), 4483-4496.

Elbanhawi, M., Simic, M., & Jazar, R. (2015). In the passenger seat: investigating ride comfort measures in autonomous cars. *IEEE Intelligent transportation systems magazine*, *7*(3), 4-17.

Elvik, R. (2020). The demand for automated vehicles: A synthesis of willingness-to-pay surveys. *Economics of Transportation*, *23*, 100179.

ERTRAC. (2022). *Connected, Cooperative and Automated Mobility Roadmap*. European Road Transport Research Advisory Council.

Euro NCAP. (2022). *Assessment Protocol - Safety Assist Safe Driving, Version 10.0.1*.

European Commission. (2022). *Road Safety Thematic Report – Driver distraction*.

European Commission, Modijefsky, M., Janse, R., Spit, W., Jankowska-Karpa, D., Buttler, I., & Eikefjord, B. (2022). *Prevention of driving under the influence of alcohol and drugs : final report*.

FAA. (2013). Advisory Circular: Fatigue Risk Management Systems for Aviation Safety. In: US Department of Transportation, Federal Aviation Administration.

Favaro, F., Hutchings, K., Nemec, P., Cavalcante, L., & Victor, T. (2022). *Waymo's Fatigue Risk Management Framework: Prevention, Monitoring, and Mitigation of Fatigue-Induced Risks while Testing Automated Driving Systems*.

Feldhütter, A., Hecht, T., Kalb, L., & Bengler, K. (2019). Effect of prolonged periods of conditionally automated driving on the development of fatigue: with and without non-driving-related activities. *Cognition, Technology & Work*, *21*(1), 33-40.

Ferrara, M., & De Gennaro, L. (2000). The sleep inertia phenomenon during the sleep-wake transition: theoretical and operational issues. *Aviation, space, and environmental medicine*, *71*(8), 843-848.

Fiorentino, A., C., A., Andersson, S., & A., T. (2023). *On-road evaluations of the vehicle-integrated Mediator system. Deliverable D3.4 of the H2020 project MEDIATOR*.

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

Foley, J. P., Young, R. A., Angell, L., & Domeyer, J. E. (2013). Towards operationalizing driver distraction. Proceedings of the 7th International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design,

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

Fredriksson, R., Lenné, M. G., van Montfort, S., & Grover, C. (2021). European NCAP Program Developments to Address Driver Distraction, Drowsiness and Sudden Sickness [Perspective]. *Frontiers in Neuroergonomics*, *2*(33).

Garrisson, H., Scholey, A., Ogden, E., & Benson, S. (2021). The effects of alcohol intoxication on cognitive functions critical for driving: A systematic review. *Accident Analysis & Prevention*, *154*, 106052.

Golz, M., Sommer, D., Trutschel, U., Sirois, B., & Edward, D. (2010). Evaluation of fatigue monitoring technologies [Article]. *Somnologie*, *14*(3), 187-199.

Gonçalves, M., Amici, R., Lucas, R., Åkerstedt, T., Cirignotta, F., Horne, J. A., . . . Téran‐Santos, J. (2015). Sleepiness at the wheel across Europe: a survey of 19 countries. *Journal of sleep research*, *24*(3), 242-253.

Green, M. (2000). " How long does it take to stop?" Methodological analysis of driver perception-brake times. *Transportation human factors*, *2*(3), 195-216.

Greenlee, E. T., DeLucia, P. R., & Newton, D. C. (2019). Driver vigilance in automated vehicles: Effects of demands on hazard detection performance. *Human Factors*, *61*(3), 474-487.

Halin, A., Verly, J. G., & Van Droogenbroeck, M. (2021). Survey and synthesis of state of the art in driver monitoring. *Sensors*, *21*(16), 5558.

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

Hatfield, J., & Fernandes, R. (2009). The role of risk-propensity in the risky driving of younger drivers. *Accident Analysis & Prevention*, *41*(1), 25-35.

Hayley, A. C., Shiferaw, B., Aitken, B., Vinckenbosch, F., Brown, T. L., & Downey, L. A. (2021). Driver monitoring systems (DMS): The future of impaired driving management? *Traffic Injury Prevention*, *22*(4), 313-317.

He, X., Stapel, J., Wang, M., & Happee, R. (2022). Modelling perceived risk and trust in driving automation reacting to merging and braking vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *86*, 178-195.

Hecht, T., Feldhütter, A., Radlmayr, J., Nakano, Y., Miki, Y., Henle, C., & Bengler, K. (2018). A Review of Driver State Monitoring Systems in the Context of Automated Driving. Congress of the International Ergonomics Association,

Hergeth, S., Lorenz, L., & Krems, J. F. (2017). Prior familiarization with takeover requests affects drivers' takeover performance and automation trust. *Human Factors*, *59*(3), 457-470.

Hermens, F. (2020). *Fatigue and distraction detection; A review of commercially available devices to detect fatigue and distraction in drivers*.

Hilditch, C. J., & McHill, A. W. (2019). Sleep inertia: Current insights. *Nature and science of sleep*, *11*, 155.

Hirsch, M., Diederichs, F., Widlroither, H., Graf, R., & Bischoff, S. (2020). Sleep and take-over in automated driving. *International journal of transportation science and technology*, *9*(1), 42-51.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Holmqvist, K., Örbom, S. L., Hooge, I. T., Niehorster, D. C., Alexander, R. G., Andersson, R., . . . Chuang, L. L. (2023). Eye tracking: empirical foundations for a minimal reporting guideline. *Behavior research methods*, *55*(1), 364-416.

Horswill, M. S., & McKenna, F. P. (2004). Drivers' hazard perception ability: Situation awareness on the road. *A cognitive approach to situation awareness: Theory and application*, 155-175.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

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

Irmak, T., de Winkel, K. N., Pool, D. M., Bülthoff, H. H., & Happee, R. (2021). Individual motion perception parameters and motion sickness frequency sensitivity in fore-aft motion. *Experimental Brain Research*, *239*(6), 1727-1745.

Irmak, T., Kotian, V., Happee, R., De Winkel, K. N., & Pool, D. M. (2022). Amplitude and Temporal Dynamics of Motion Sickness. *Frontiers in Systems Neuroscience*, *16*.

Jannusch, T., Shannon, D., Völler, M., Murphy, F., & Mullins, M. (2021). Cars and distraction: How to address the limits of Driver Monitoring Systems and improve safety benefits using evidence from German young drivers. *Technology in Society*, *66*, 101628.

Kashevnik, A., Shchedrin, R., Kaiser, C., & Stocker, A. (2021). Driver distraction detection methods: A literature review and framework. *IEEE Access*, *9*, 60063-60076.

Kerick, S., Metcalfe, J., Feng, T., Ries, A., & McDowell, K. (2013). *Review of fatigue management technologies for enhanced military vehicle safety and performance*.

Kim, H., Song, M., & Doerzaph, Z. (2022). Is Driving Automation Used as Intended? Real-World Use of Partially Automated Driving Systems and their Safety Consequences. *Transportation research record*, *2676*(1), 30-37.

Kircher, K., & Ahlstrom, C. (2016). Minimum Required Attention: A Human-Centered Approach to Driver Inattention. *Hum Factors*.

Klien, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a" team player" in joint human-agent activity. *IEEE Intelligent Systems*, *19*(6), 91-95.

Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors*, *18*(2), 401.

Koay, H. V., Chuah, J. H., Chow, C.-O., & Chang, Y.-L. (2022). Detecting and recognizing driver distraction through various data modality using machine learning: A review, recent advances, simplified framework and open challenges (2014–2021). *Engineering Applications of Artificial Intelligence*, *115*, 105309.

Koch, K., Maritsch, M., van Weenen, E., Feuerriegel, S., Pfäffli, M., Fleisch, E., . . . Wortmann, F. (2023). Leveraging driver vehicle and environment interaction: Machine learning using driver monitoring cameras to detect drunk driving. *arXiv preprint arXiv:2301.08978*.

Koesdwiady, A., Soua, R., Karray, F., & Kamel, M. S. (2017). Recent Trends in Driver Safety Monitoring Systems: State of the Art and Challenges. *IEEE Transactions on Vehicular Technology*, *66*(6), 4550-4563.

Kotseruba, I., & Tsotsos, J. K. (2022). Attention for vision-based assistive and automated driving: A review of algorithms and datasets. *IEEE Transactions on intelligent transportation systems*.

Kranjec, J., Beguš, S., Geršak, G., & Drnovšek, J. (2014). Non-contact heart rate and heart rate variability measurements: A review. *Biomedical Signal Processing and Control*, *13*, 102-112.

Kundinger, T., Riener, A., Sofra, N., & Weigl, K. (2020). *Driver drowsiness in automated and manual driving: insights from a test track study* Proceedings of the 25th International Conference on Intelligent User Interfaces, Cagliari, Italy.

Körber, M., Cingel, A., Zimmermann, M., & Bengler, K. (2015). Vigilance decrement and passive fatigue caused by monotony in automated driving. *Procedia Manufacturing*, *3*, 2403-2409.

Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research–recommendations for experiment planning, data analysis, and data reporting. *Frontiers in psychology*, *8*, 213.

Landsberger, H. A. (1958). Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry.

Lee, J. D. (2014). Dynamics of Driver Distraction: The process of engaging and disengaging. *Ann Adv Automot Med*, *58*, 24-32.

Leonhardt, S., Leicht, L., & Teichmann, D. (2018). Unobtrusive Vital Sign Monitoring in Automotive Environments—A Review. *Sensors*, *18*(9), 3080.

Li, Y., Moreau, J., & Ibanez-Guzman, J. (2022). Unconventional Visual Sensors for Autonomous Vehicles. *arXiv preprint arXiv:2205.09383*.

Lin, Y.-T., Lin, R.-Y., Lin, Y.-C., & Lee, G. C. (2013). Real-time eye-gaze estimation using a low-resolution webcam. *Multimedia Tools and Applications*, *65*, 543-568.

Liu, C. C., Hosking, S. G., & Lenné, M. G. (2009). Predicting driver drowsiness using vehicle measures: recent insights and future challenges. *Journal of safety research*, *40*(4), 239-245.

Llaneras, R. E., Salinger, J., & Green, C. A. (2013). Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention to the forward roadway. Driving Assesment Conference,

Lu, C. Y. (2009). Observational studies: a review of study designs, challenges and strategies to reduce confounding. *International journal of clinical practice*, *63*(5), 691-697.

Lu, K., Karlsson, J., Dahlman, A. S., Sjöqvist, B. A., & Candefjord, S. (2021). Detecting Driver Sleepiness Using Consumer Wearable Devices in Manual and Partial Automated Real-Road Driving. *IEEE Transactions on intelligent transportation systems*.

Lu, K., Sjörs Dahlman, A., Karlsson, J., & Candefjord, S. (2022). Detecting driver fatigue using heart rate variability: A systematic review. *Accident Analysis & Prevention*, *178*, 106830.

Malik, M. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use: Task force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology. *Annals of Noninvasive Electrocardiology*, *1*(2), 151-181.

Malin, F., Silla, A., Mesimäki, J., Innamaa, S., & Peltola, H. (2022). Identifying crashes potentially affected by conditionally automated vehicles in Finland. *Journal of Intelligent Transportation Systems*, 1-12.

Maurage, P., Masson, N., Bollen, Z., & D'Hondt, F. (2020). Eye tracking correlates of acute alcohol consumption: A systematic and critical review. *Neuroscience & Biobehavioral Reviews*, *108*, 400-422.

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: Psychology and Behaviour*, *12*(3), 218-224.

McCormick, F., Kadzielski, J., Landrigan, C. P., Evans, B., Herndon, J. H., & Rubash, H. E. (2012). Surgeon fatigue: a prospective analysis of the incidence, risk, and intervals of predicted fatigue-related impairment in residents. *Archives of Surgery*, *147*(5), 430-435.

McDonald, A., Carney, C., & McGehee, D. V. (2018). *Vehicle owners' experiences with and reactions to advanced driver assistance systems*.

McDonald, A. D., Lee, J. D., Schwarz, C., & Brown, T. L. (2014). Steering in a random forest: Ensemble learning for detecting drowsiness-related lane departures [Article]. *Human Factors*, *56*(5), 986-998.

McStay, A., & Urquhart, L. (2022). In cars (are we really safest of all?): interior sensing and emotional opacity. *International Review of Law, Computers & Technology*, *36*(3), 470-493.

Miller, D., Sun, A., Johns, M., Ive, H., Sirkin, D., Aich, S., & Ju, W. (2015). Distraction becomes engagement in automated driving. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,

Mollicone, D., Kan, K., Mott, C., Bartels, R., Bruneau, S., van Wollen, M., . . . Van Dongen, H. P. (2019). Predicting performance and safety based on driver fatigue. *Accident Analysis & Prevention*, *126*, 142-145.

Moslemi, N., Soryani, M., & Azmi, R. (2021). Computer vision ‐ based recognition of driver distraction: A review. *Concurrency and Computation: Practice and Experience*, *33*(24), e6475.

Mueller, A. S., Cicchino, J. B., & Zuby, D. S. (2020). What humanlike errors do autonomous vehicles need to avoid to maximize safety? *Journal of safety research*, *75*, 310-318.

Mueller, A. S., Reagan, I. J., & Cicchino, J. B. (2021). Addressing Driver Disengagement and Proper System Use: Human Factors Recommendations for Level 2 Driving Automation Design. *Journal of Cognitive Engineering and Decision Making*, *15*(1), 3-27.

Ni, Z., Sun, F., & Li, Y. (2022). Heart Rate Variability-Based Subjective Physical Fatigue Assessment. *Sensors*, *22*(9), 3199.

Noble, A. M., Miles, M., Perez, M. A., Guo, F., & Klauer, S. G. (2021). Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems. *Accident Analysis & Prevention*, *151*, 105959.

Noy, I. Y., Shinar, D., & Horrey, W. J. (2018). Automated driving: Safety blind spots. *Safety Science*, *102*, 68-78.

Oman, C. M. (1982). A heuristic mathematical model for the dynamics of sensory conflict and motion sickness. *Acta Oto-Laryngologica*, *94*(sup392), 4-44.

Ortega, J. D., Cañas, P. N., Nieto, M., Otaegui, O., & Salgado, L. (2022). Challenges of Large-Scale Multi-Camera Datasets for Driver Monitoring Systems. *Sensors*, *22*(7), 2554.

Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., . . . Dierolf, U. (2013). Making research data repositories visible: the re3data. org registry. *PloS one*, *8*(11), e78080.

Paprocki, S., Qassem, M., & Kyriacou, P. A. (2022). Review of ethanol intoxication sensing technologies and techniques. *Sensors*, *22*(18), 6819.

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

Perkins, E., Sitaula, C., Burke, M., & Marzbanrad, F. (2022). Challenges of driver drowsiness prediction: The remaining steps to implementation. *IEEE Transactions on Intelligent Vehicles*.

Persson, A., Jonasson, H., Fredriksson, I., Wiklund, U., & Ahlström, C. (2020). Heart Rate Variability for Classification of Alert Versus Sleep Deprived Drivers in Real Road Driving Conditions. *IEEE Transactions on intelligent transportation systems*, *22*(6), 3316-3325.

Qian, K., Koike, T., Nakamura, T., Schuller, B., & Yamamoto, Y. (2021). Learning multimodal representations for drowsiness detection. *IEEE Transactions on intelligent transportation systems*.

Regan, M. A., Hallett, C., & Gordon, C. P. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis & Prevention*, *43*(5), 1771-1781.

Rolnick, A., & Lubow, R. (1991). Why is the driver rarely motion sick? The role of controllability in motion sickness. *Ergonomics*, *34*(7), 867-879.

SAE. (2021). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (Standard J3016_202104). In: SAE International.

Sandberg, D., Åkerstedt, T., Anund, A., Kecklund, G., & Wahde, M. (2011). Detecting Driver Sleepiness Using Optimized Nonlinear Combinations of Sleepiness Indicators. *IEEE Transactions on intelligent transportation systems*, *12*(1), 97-108.

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

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.

Senders, J. W., Kristofferson, A., Levison, W., Dietrich, C., & Ward, J. (1967). The attentional demand of automobile driving.

Shekari Soleimanloo, S., Wilkinson, V. E., Cori, J. M., Westlake, J., Stevens, B., Downey, L. A., . . . Howard, M. E. (2019). Eye-blink parameters detect on-road track-driving impairment following severe sleep deprivation. *JOURNAL OF CLINICAL SLEEP MEDICINE*, *15*(9), 1271-1284.

Silva, J. B. S., Cristino, E. D., Almeida, N. L. d., Medeiros, P. C. B. d., & Santos, N. A. d. (2017). Effects of acute alcohol ingestion on eye movements and cognition: a double-blind, placebo-controlled study. *PloS one*, *12*(10), e0186061.

Singh, S. (2015). *Critical reasons for crashes investigated in the national motor vehicle crash causation survey* (DOT HS 812 115).

Solís-Marcos, I., Ahlström, C., & Kircher, K. (2018). Performance of an additional task during Level 2 automated driving: an on-road study comparing drivers with and without experience with partial automation. *Human Factors*, 0018720818773636.

Song, Y., & Vink, P. (2021). On the objective assessment of comfort. 3rd International Comfort Congress 2021,

Stewart, T. (2022). *Overview of Motor Vehicle Crashes in 2020*.

Talsma, T. M., Hassanain, O., Happee, R., & de Winkel, K. N. (2023). Validation of a moving base driving simulator for motion sickness research. *Applied Ergonomics*, *106*, 103897.

Tan, C., Delbressine, F., Chen, W., & Rauterberg, M. (2008). Subjective and objective measurements for comfortable truck drivers' seat. Proceedings of the Ninth International Symposium on Advanced Vehicle Control (AVEC 2008),

Tassi, P., & Muzet, A. (2000). Sleep inertia. *Sleep medicine reviews*, *4*(4), 341-353.

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

Tervo, T. M., Neira, W., Kivioja, A., Sulander, P., Parkkari, K., & Holopainen, J. M. (2008). Observational failures/distraction and disease attack/incapacity as cause (s) of fatal road crashes in Finland. *Traffic Injury Prevention*, *9*(3), 211-216.

Trutschel, U., Sirois, B., Sommer, D., Golz, M., & Edwards, D. (2011). PERCLOS: An alertness measure of the past.

Van Dongen, H. P. (2004). Comparison of mathematical model predictions to experimental data of fatigue and performance. *Aviation, space, and environmental medicine*, *75*(3), A15-A36.

van Grondelle, E. D. (2023). *Guidelines safe HMI design. Deliverable D4.2 of the H2020 project MEDIATOR*.

Vicente, J., Laguna, P., Bartra, A., & Bailón, R. (2016). Drowsiness detection using heart rate variability [journal article]. *Medical & Biological Engineering & Computing*, *54*(6), 927-937.

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(2), 167-190.

Vogelpohl, T., Kühn, M., Hummel, T., & Vollrath, M. (2019). Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving. *Accident Analysis & Prevention*, *126*, 70-84.

Vrijkotte, T. G., Van Doornen, L. J., & De Geus, E. J. (2000). Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension*, *35*(4), 880-886.

Wang, N., Gao, X., Tao, D., Yang, H., & Li, X. (2018). Facial feature point detection: A comprehensive survey. *Neurocomputing*, *275*, 50-65.

Warwick-Evans, L., Church, R., Hancock, C., Jochim, D., Morris, P., & Ward, F. (1987). Electrodermal activity as an index of motion sickness. *Aviation, space, and environmental medicine*.

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

Wörle, J., Metz, B., & Baumann, M. (2021). Sleep inertia in automated driving: Post-sleep take-over and driving performance. *Accident Analysis & Prevention*, *150*, 105918.

Wörle, J., Metz, B., & Prill, A. (2023). How to Induce Drowsiness When Testing Driver Drowsiness and Attention Warning (DDAW) Systems. *IEEE Transactions on intelligent transportation systems*.

Yamani, Y., Samuel, S., Knodler, M. A., & Fisher, D. L. (2016). Evaluation of the effectiveness of a multi-skill program for training younger drivers on higher cognitive skills. *Applied Ergonomics*, *52*, 135-141.

Yamazaki, E. M., & Goel, N. (2020). Robust stability of trait-like vulnerability or resilience to common types of sleep deprivation in a large sample of adults. *SLEEP*, *43*(6), zsz292.

Yue, L., Abdel-Aty, M., Wu, Y., & Wang, L. (2018). Assessment of the safety benefits of vehicles' advanced driver assistance, connectivity and low level automation systems. *Accident Analysis & Prevention*, *117*, 55-64.

Zhi, R., Liu, M., & Zhang, D. (2020). A comprehensive survey on automatic facial action unit analysis. *The Visual Computer*, *36*, 1067-1093.

Åkerstedt, T., Anund, A., Axelsson, J., & Kecklund, G. (2014). Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *Journal of sleep research*, *23*(3), 240-252.