TECHNICAL REPORT

ISO/TR 21959-2

First edition
2020-02

Road vehicles — Human performance and state in the context of automated driving —

Part 2: Considerations in designing experiments to investigate transition processes

Véhicules routiers — Etat et performance humaine dans le contexte de la conduite automatisée —

Partie 2: Principes expérimentaux pour etudier les processus de transition

Reference number
ISO/TR 21959-2:2020(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 22, *Road vehicles*, Subcommittee SC 39, *Ergonomics*.

A list of all parts in the ISO 21959 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Although automation technology is advancing at a rapid pace, the majority of automated driving levels (as defined by SAE J3016, 2016[1]) still require a human to fulfil specific remaining (driving related) tasks. The safety-critical human's task is the takeover task in transition from a higher level to a lower level of automated driving. Researchers and developers continue to seek system design and human machine interface improvements for better takeover performance. Researchers face a challenge in understanding the limitations of a human's ability to perform the takeover task, which involves different human factors. Developers work to evaluate systems to see whether the takeover process is effective at minimum risk in specific scenarios. There are a wide variety of experiments to evaluate takeover performance in transition for many different purposes. This document contains information to consider in the takeover scenario, some of which is still under investigation, in order to help readers design experiments to evaluate takeover performance and design appropriate experiments.

Road vehicles — Human performance and state in the context of automated driving —

Part 2: Considerations in designing experiments to investigate transition processes

1 Scope

This document focuses on system-initiated and human-initiated transitions (Clause 6) from a higher level to a lower level of automated driving. Human factors and system factors that can influence takeover performance are included (Clauses 7 and 8). Although some are still under investigation, there is a need to appropriately set these factors as variables to better understand their effects or to better control/eliminate their influence. This approach will aid research design by ensuring that important factors are considered and support consistency across studies enabling meaningful comparisons of findings. This document also includes information on considerations in test scenario design (Clause 9), common measures for human takeover performance (Clause 10) and considerations in choosing a testing environment (Clause 11) to help readers design experiments comparable to other studies.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

No terms and definitions are listed in this document.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

4 List of Acronyms

DDT	Dynamic Driving Task
DMS	Driver Monitor System
ECG	Electrocardiogram
EEG	Electroencephalogram
HMI	Human-Machine Interface
KSS	Karolinska Sleepiness Scale
MRM	Minimal Risk Manoeuvre
NDRT	Non-driving Related Task

NDRA Non-driving Related Activities

OEDR Object and Event Detection and Response

ODD Operational Design Domain

RtI Request to Intervene

SAGAT Situation Awareness Global Assessment Technique

SDLP Standard Deviation of Lateral Position

SIMS Situational Motivation Scale

SuRT Surrogate Reference Task

TLC Time to Lane Crossing

TTC Time-to-Contact/Collision

5 Purpose

The purpose of this document is to provide considerations in designing experiments to measure human takeover performance in transition situations in order to better understand human limitations, evaluate systems, and improve systems, including human machine interfaces. This document is expected to help users design appropriate experiments for their purposes. This document does not provide any design principles to restrict or direct the system design.

6 Transition process models

6.1 General

Transition processes included in this document are generally based on the models defined in ISO/TR 21959-1[2]. A human's safety-critical task is the takeover task in transition from a higher level to a lower level of automated driving both for system-initiated and human-initiated transitions. ISO/TR 21959-1[2] defines typical transition process models from automated to manual driving (i.e. level 0). However, the models can be adapted for transitions between different levels (e.g. 4→2 or 3→1). This clause reminds readers of the relevant transition process models.

6.2 Transition process model for system-initiated transitions

In a system-initiated transition, the system may issue a request to intervene (RtI) when it finds a dynamic driving task (DDT) performance-relevant system failure or an object/event which cannot be handled by the system for levels 1–4. The system also may issue an RtI when exiting the operational design domain (ODD) for which it was designed, (e.g. exiting a motorway, exiting assumed environmental conditions such as weather and traffic). The driver is expected to take over the DDT in response to an RtI to continue driving. The system may terminate immediately after issuing an RtI for level 2 while it shifts to the takeover mode following an RtI before termination for levels 3 and 4 (Figure 1). There can also be other types of transitions after an RtI, such as transitions from level 2 to level 1 and from level 3 to level 2. The driver's task model can be adjusted depending on the level after the RtI (i.e. object and event detection and response [OEDR] task for transition to level 2, OEDR task +lateral control or OEDR task +longitudinal control for transition to level 1). When the driver does not initiate intervention within the takeover mode, the system may shift to the minimal risk manoeuvre (MRM) to stop the vehicle safely for level 3 and level 4 (see 8.2.2 for details).

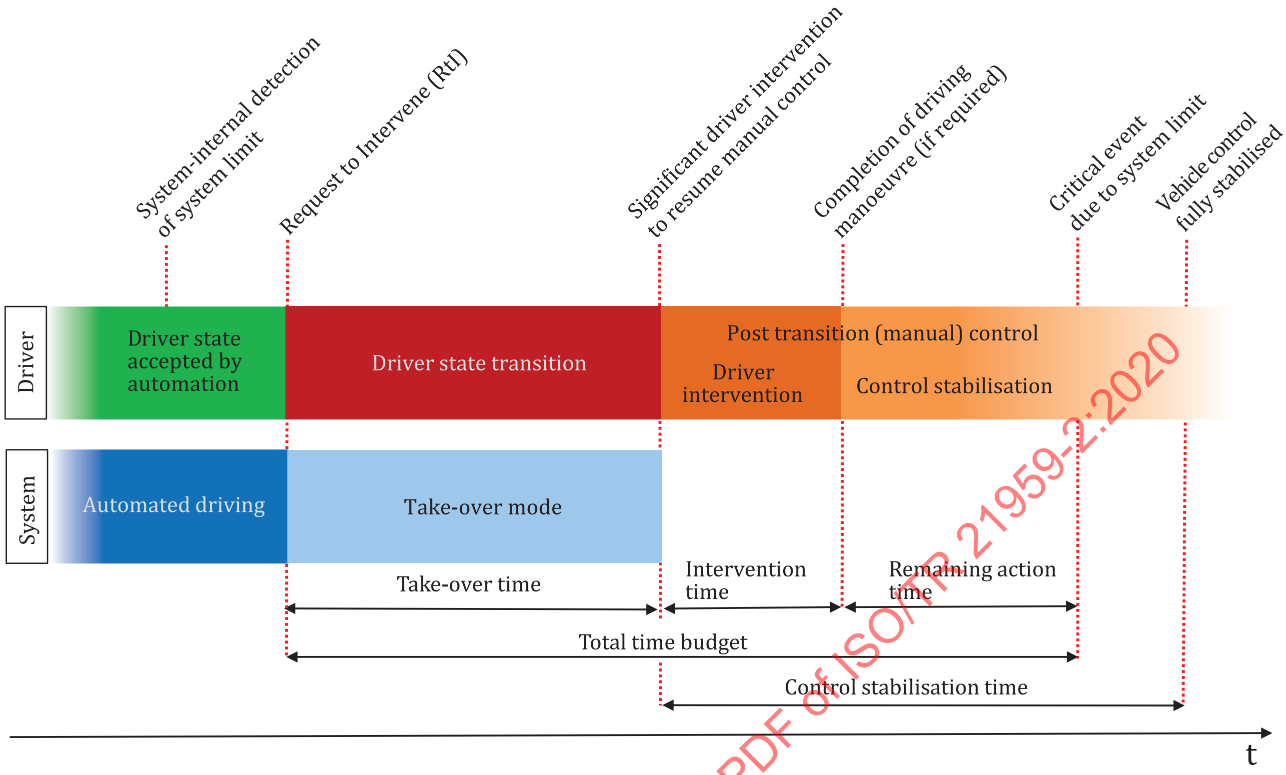


Figure 1 — Transition process model for system-initiated transitions from automated to manual driving

6.3 Transition process model for human-initiated transitions

The driver is authorized to take over the DDT at any point during operation of the automated driving functions, except for some level 4 and level 5 features some or all of the time. The human-initiated transition may be either optional or mandatory. The optional case is the transition where the user wishes to drive manually without being in a safety critical situation. The mandatory case is the transition where the level 2 system fails to avoid an undetected object/event due to the system's functional limitations or where the system suddenly terminates without issuing an RtI due to a DDT performance-relevant system failure. In such mandatory transitions, the driver is expected to detect the object/event or the failure and initiate transition (Figure 2). This type of transition is mainly from level 2 to manual driving but can be from level 2 to level 1. The driver's task model can be adjusted depending on the level after the initiation.

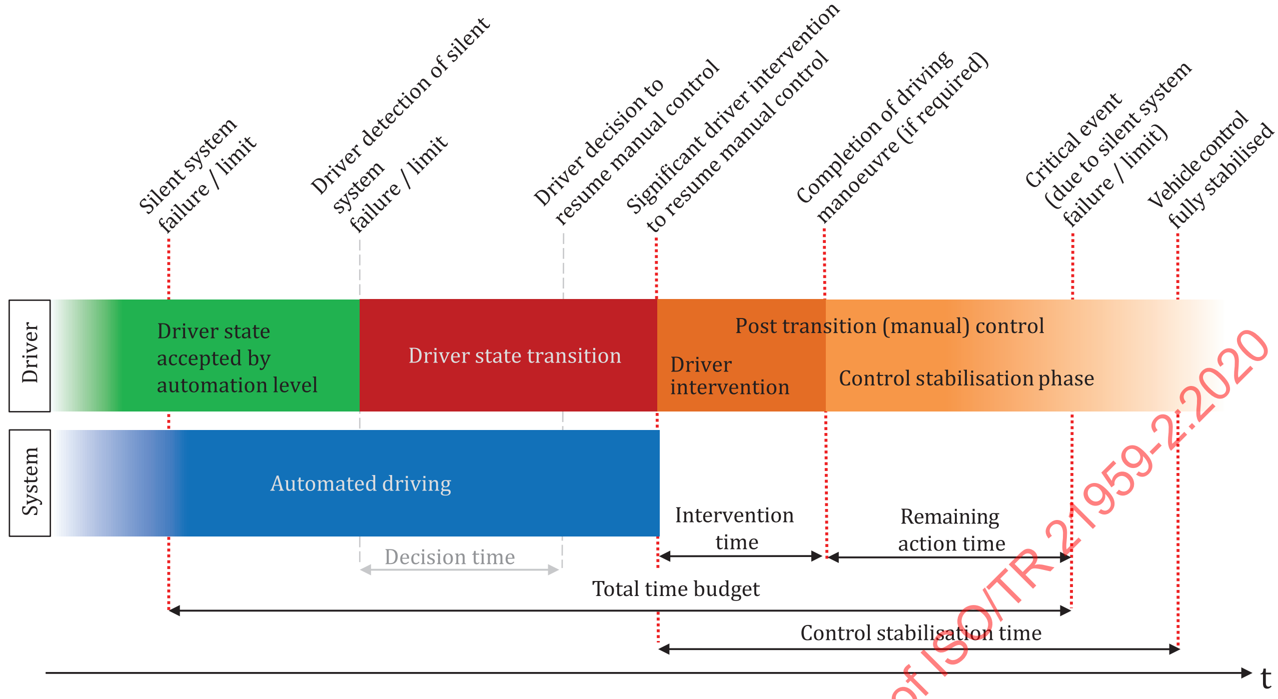


Figure 2 — Transition process model for mandatory human-initiated transitions from automated (level 2) to manual driving due to driver's detection of a safety-critical object/event or a DDT performance-relevant system failure

7 Human factors that influence takeover performance

7.1 General

It is known that a driver's takeover performance varies with the influences of multiple factors. In experiments, there is a need to appropriately set factors as variables for investigating their effects or better controlling/eliminating their influence. This will allow for the design of experiments that are easier to compare with other studies. This clause presents information about "internal" human factors that may influence a driver's takeover performance. Driver's takeover performance includes time in the driver state transition phase (i.e. response time of significant driver intervention to RtI) and quality in the post transition control phase (i.e. how well the driver controls the vehicle right after the significant driver intervention; see also Clause 10).

7.2 Driver attributes

7.2.1 Knowledge

Drivers' knowledge about system functions, limitations and the required driver's role, has been found to influence takeover performance in some studies[3][4][5]. Other studies have found that instructions have limited effects[6][7]. In general, the sources of a driver's knowledge are diverse and may include mass-media, instruction manuals, instructions given at a car-dealership and other various sources. In experiments, such knowledge can be controlled, to some extent, by screening subjects using questionnaires investigating their level of a-priori knowledge and by providing them with controlled information about the functions, limitations and driver's role for the specific system of study. It is to be noted that difficulty in forming a detailed picture of subjects' exact knowledge obtained from various sources for various systems may lead to some variation in the results of takeover performance. In some instances, participants may have incorrect knowledge about system function, limitations and the required driver's role leading to misbehaviour or misuse.

7.2.2 Experience and trust

A driver's experience with using the system has been found to influence takeover performance. How the driver has previously interacted with the system may influence performance in different ways as a result of different levels of understanding and trust. Short term system interaction experiences may lead to a better understanding of the system's functions and limitations and better trust calibration, resulting in better driver takeover performance[8][9][10][11][4][5]. Longer-term experience with no to a few system disengagements may lead to a driver's over-trust of the system and complacency[12], which can degrade takeover performance[6]. In contrast, longer-term experience with too many system disengagements may lead to driver under-trust, which might improve takeover performance[13] but which also may lead to disuse of the system[14]. In experiments, such experiences can be controlled, to some extent, by screening subjects using questionnaires investigating their experiences with specific systems featured in the study and the frequencies of the interactions they have experienced. After screening, new subject experience can be introduced by providing subjects specific driving conditions after providing controlled interactions. It is to be noted that similar systems with the same level of automated driving can differ in functions and limitations (i.e. detection targets, ODD, reliability and others) by brand or even by different models within one brand. Also, different users of the same brand system may use the system in different ways in different traffic environments. Therefore, experience still may lead to some dispersion in the results.

7.2.3 Demographic attributes

7.2.3.1 Age

A driver's age-related perceptual, cognitive and physical limitations may influence takeover performance. Visual impairments of older drivers are diverse and are often accompanied by eye diseases[15]. Such impairments may degrade perception of traffic environment in the OEDR task or in the process of transition. Visual impairments may also cause difficulty in reading system status information displayed in the cockpit[16]. Cognitive impairments may degrade understanding of "complicated" system functions/limitations and the driver's role. These impairments may also lead to problems with divided attention[17] and slow down task switching in transitions. Physical impairment may degrade speed and accuracy of the response behaviour in transitions. Although, as mentioned above, there are several hypotheses for older drivers' degraded takeover performance, the effects of age are still under discussion. Some researchers have found significant negative effects of age[18][19][5][20][21], whereas other researchers found only limited effects[22][23]. In experiments, subjects can be screened not only based on age but also based on the results of perceptual, cognitive and physical response tests. However, it is to be noted that the effects of age have large inter/intra-individual variability and still may lead to some dispersion in the results of takeover performance of subjects who were screened via tests.

7.2.3.2 Other demographic attributes

There are other driver demographic attributes that may influence takeover performance, such as experience and skill of manual driving, style of manual driving with individual and cultural differences, technology-sensitivity and general trust of technology. However, these factors have not yet been well studied.

7.3 Driver readiness/availability

Conceptually, readiness/availability is a driver's dynamic state during automated driving, which influences their takeover performance. Readiness/availability can be continuous; lower readiness/availability than a required level may lead to degraded and unsafe takeover performance[24][25][26][27][28] (see also A.2.2). Considering the definitions of the driver's role for each level of automation, the required level of readiness/availability generally increases with decreasing levels of automation. The readiness/availability is considered to include several components related to motoric/physical and cognitive states (Table 1). Each component can have different metrics and different effects on takeover performance. The required level for each component of readiness/availability can be experimentally determined as the level that leads to a successful takeover by comparing the metrics and the takeover performance in time and quality in certain traffic conditions for a specific system transition design (see also A.2.2).

Table 1 — Components of readiness/availability

Components of readiness/availability	Motoric/physical state	Cognitive state
Sitting position	✔	—
Posture	✔ Hands/arms, feet/legs, trunk	—
Engagement in NDRAs	✔ Hands/arms, manual operation	✔ Visual Cognitive
Drowsiness	—	✔
Mind wandering	—	✔
Situation awareness	—	✔
Operating state/mode awareness	—	✔

7.3.1 Sitting position and posture

When the driver is away from the driver's seat and sitting at another location[170], a significant amount of time will be required to return to that seat. One study reported that a driver with a larger torso angled in a relaxed posture showed poor takeover performance[29]. It may require a certain amount of time to return to the appropriate driving posture from a relaxed posture with, for example, the backrest inclined backward. When there is a large space between the driver and the system controls, with the seat moved backward, the steering wheel moved upward and the legs and/or arms crossed, it may require even more time for the driver to return to the appropriate driving posture[23]. In experiments, the driver's position and posture can be controlled by setting the seat and the steering wheel in the desired position and also by instructing subjects (drivers) to assume a desired posture. When evaluating a driver's takeover performance in the naturalistic setting, the driver's position and posture can be monitored by video recording, steering-touch sensors, a seat pressure monitor[30], seat position sensors, steering position sensors and the seatbelt buckle switch.

7.3.2 Engagement in non-driving related activities

When the driver is engaged in a non-driving related activity (NDRA), such as using a laptop, reading a book, or operating a hand-held device using hand(s) off the steering wheel, it can take more time to take over than when the driver is in the appropriate driving posture with both hands placed on the wheel[25][29][23]. The larger takeover time may include the time required to place the item in a secured place before grasping the steering wheel[31]. When the driver performs takeover with one hand while holding an item with the other hand, the quality of takeover performance may be degraded due to inaccurate one-handed steering operation. Time can be also consumed when the driver is wearing glasses for an NDRA and takes them off for takeover or vice versa[23].

Engagement in NDRAs may induce driver's manual examination or other visual and cognitive loads that may influence takeover performance. Some NDRAs require interactive driver manual operation (e.g. selecting a function on a touch-panel). Texting also requires continuous manual operation. NDRA manual operations induce one-handed steering and also intensive visual load for accurate operation. A visually loaded driver may fail to sample environmental elements that are necessary to develop the appropriate situation awareness[32]. When looking down, the driver may fail/delay to detect an object/event that requires immediate driver-initiated takeover for a level 2 system. This inappropriate situation awareness may degrade the quality of takeover performance after an RtI. When the driver is cognitively loaded by an NDRA, insufficient attention allocated to the road environment may also degrade the quality of takeover performance. Degradation of quality of takeover performance can be an inappropriate manoeuvre (e.g. abrupt manoeuvre, inappropriate choice of manoeuvre) or a delayed response manoeuvre to an event[33][34].

A number of studies have investigated the effects of NDRAs on takeover performance. Effects of specific types of NDRAs (e.g. using a tablet, typing an email, reading news, watching a video, performing an auditory-vocal task) on takeover performance have been investigated by some researchers[35][36][37]. Some NDRAs were found to degrade takeover performance in either time or quality or both. Louw et al. (2019)[38] found that a driver engaged in a visual NDRA (the arrows task in a computer display) with a level 2 system showed degraded performance of the driver-initiated takeover for a “silent failure” due to inappropriate attention allocation. Zeeb et al. (2016)[36] found that NDRAs such as reading a news text and watching a video did not influence the response time of grasping the steering wheel after an RtI (i.e. motor processes were carried out almost reflexively), whereas the lateral control of the vehicle in the control stabilisation phase was significantly degraded. The findings of Zeeb et al. (2016)[36] were consistent with those of Kitazaki et al. (2019)[39] and Choi et al. (under review)[34], who found a visual-manual task using a surrogate reference task (SuRT) (ISO/TS 14198)[40] caused an abrupt steering manoeuvre to change the lane to avoid a stationary object after an RtI, resulting in unstable lateral vehicle control after changing lanes. Kitazaki et al.[39] and Choi et al.[34] also found that a cognitive load using the N-Back task[41] slowed down the steering manoeuvre after an RtI, resulting in a shorter minimum distance to the stationary object. In contrast, Radlmayr et al. (2014)[42] found similar effects of SuRT and N-Back tasks in increased collision rates in the high density traffic situation.

Positive effects of NDRAs have also been reported by some researchers. Automated driving may cognitively underload the driver[43], resulting in development of drowsiness and degradation of performance[36]. NDRAs may counteract the cognitive underload and maintain driver alertness. Neubauer et al. (2012)[44] found that drivers using a cell phone showed a faster braking response following transition to manual driving than drivers without NDRAs. Schömig et al. (2015)[45] reported that the drowsiness of drivers given a quiz task stayed low compared to those without the task.

NDRAs used in experiments can be selected from those that drivers are most likely to be engaged in[46]. Simple representations of the tasks, such as SuRT for the visual manual task and N-Back for the cognitive task, can be also used for the purpose of analysing the influences more systematically and in a standardized way. Video recordings of a driver’s behaviour are useful to extract time slots where the driver is engaged in NDRAs and to identify NDRA types, especially in naturalistic studies. When using an electronic device for an NDRA, performance on the device (e.g. number of touches on the screen) can be used to measure extent of a driver’s engagement in the NDRA. Subjective measures, such as the rating scale for mental effort[47] and NASA's task load index[48] have been widely used to assess the driver’s workload to conduct the NDRAs, even though these measures do not provide continuous information and cannot be used unobtrusively. There are a number of studies investigating biometrics of readiness/availability of a driver engaged in NDRAs. Because a driver’s input on the primary vehicle controls cannot be used as a metric of driver states when driving with the automated system, a majority of studies have focused on metrics obtained from video recordings of the driver’s face. These studies included gaze behaviour, such as gaze distribution and eyes-off-road time, as well as eye movements, such as frequency of saccades, blink duration, blink frequency, Perclos and pupil diameter[24][49][50][26][51]. There are a number of other studies that have estimated NDRA effects for distracted driving[52][53][54][55][50].

Emotional attachment to an NDRA caused by high motivation to the NDRA[56] may also influence the takeover performance. Even after the RtI has been issued, the driver can be strongly motivated to continue certain types of NDRAs (e.g. typing an email, reading a new article) to finish a chunk of activity before takeover. Delayed glance movement from the NDRA to the front after an RtI may explain this effect. The motivation in a given situation can be subjectively assessed by the situational motivation scale[57].

7.3.3 Drowsiness

Automated driving may cognitively underload the driver, resulting in drowsiness and degradation of vigilance[58][36]. Drivers participating in driving simulator experiments tend to develop drowsiness more rapidly than those in on-road experiments[59]. Long driving durations and monotonous driving environments tend to lead to faster development of drowsiness[60]. Higher drowsiness levels correspond to a smaller amount of accessible cognitive resources. Therefore, it is a reasonable hypothesis that drowsiness can degrade takeover performance in terms of time and quality[61][36]. A drowsy driver may fail to detect an object/event that requires driver-initiated transition for the level 2 system. For system-initiated transitions, negative effects of drowsiness on takeover performance were reported by some researchers[62][39], whereas no significant effects were found by other researchers[63][64].

In experiments, subjects' drowsiness levels can be controlled to some extent by conducting experiments considering circadian rhythms, instructing subjects to control their length of sleep before the experiment, or adjusting the driving scenario's duration and level of monotony[65]. However, it is difficult to precisely control subjects' drowsiness levels. Subjective measures, such as the Karolinska sleepiness scale (KSS), have been widely used to assess drowsiness levels[66]. Some biometrics have also been used to measure drowsiness levels, including Perclos[67], blinking duration[68][21], visual scanning, and also some physical actions[69][70]. Electroencephalograms (EEGs) have been used to measure depth and type of sleep as well as micro sleeps during driving[71].

7.3.4 Mind wandering

Mind wandering means thinking about issues unrelated to the ongoing driving task. Mind wandering shifts a driver's attention to internal information[72] and may result in deteriorated takeover performance similar to that caused by drivers with insufficient attention to driving due to an NDRA or low arousal[73]. In general, mind wandering is more likely to occur when a subject's vigilance is low[74]. Therefore, long durations of automated driving and a monotonous driving environment are more likely to induce mind wandering[58]. Although it is difficult to actively manipulate a driver's mind wandering in experiments, some studies have used "thought-sampling methods," which are based on self-report or intermittent questions such as, "Just now, were you mind wandering?" as part of the experimental protocol[75][76]. Some biometrics have also been found for mind wandering, such as gaze direction, pupil size[77], electrocardiogram (ECG)[76] and EEG[78].

7.3.5 Situation awareness

Situation awareness (SA) is the perception of environmental elements and events with respect to time or space, the comprehension of their meaning and the projection of their future status[79] and models of SA have been developed for driving (e.g. Reference [80]). Understanding the traffic situation (e.g. positions and speeds of neighbouring vehicles) and road environment (e.g. number of lanes, road shape) can influence takeover performance in time and quality when selecting an appropriate driving tactic to cope with a critical event both in driver-initiated and system-initiated transitions[81]. This influence becomes larger when traffic is denser and the road environment is more complex[42][33]. An example use-case is a transition before an obstacle in the same lane with another vehicle in the neighbouring lane. The driver is expected to take over control of the vehicle and change lanes before or after the neighbouring vehicle, depending on the gap and the relative speed to the neighbouring vehicle. Insufficient situation awareness may delay the decision making for an appropriate lane change or may lead to a collision with the neighbouring vehicle due to an inappropriate lane change initiation. Strategic level of situation awareness may also be important. An example use-case is an instance where a driver engaged in a NDRA receives an RtI and takes over control of the vehicle without being aware of the current location on the route to the destination. The driver may not be properly prepared for taking a turn at a forthcoming intersection[82], which may result in a risky situation if the driver takes the turn too late or abruptly.

Situation awareness can be influenced by NDRAs, drowsiness and mind wandering as mentioned above. It can also be influenced by trust and experienced comfort[83]. Although there are some studies which used a foggy environment to degrade a driver's situation awareness in a driving simulator experiment[84], it is difficult to control situation awareness in general. The situation awareness global assessment technique (SAGAT[85]) has been used to measure the level of situation awareness. However, it is difficult to use SAGAT to measure dynamic changes in situation awareness. A driver's gaze behaviour can be monitored as a part of situation awareness[86]. Situation awareness is a useful concept but difficult to control or measure precisely, continuously and unobtrusively.

7.3.6 Operating state/mode awareness

The level of a driver's understanding of the state of the system operation/mode may influence takeover performance. Particularly when multiple levels of automated driving exist within the system (e.g. levels 0, 1, 2, 3), an insufficient awareness of the operating mode — called mode confusion — can be an issue[87][83]. Mode confusion can result in a missed or delayed initiation of takeover (both driver-initiated and system initiated) if the driver wrongly believes the system is working at a higher level than is actually engaged[88]. An example use-case is a situation in which the driver believes the system is operating as level 3 and fails to perform the OEDR task when the system is actually operating as level 2. Insufficient operating mode awareness may also lead to automation surprise when the system is actually working without the driver being aware[83].

Operating state/mode awareness may be dependent on an understanding of the driver's role for each level of the system and an understanding the dynamic state/mode of the current system operation. In experiments, instruction of the driver's role for each level and a human machine interface (HMI) displaying the dynamic state/mode of the system operation may control the driver's operating state/mode awareness to some extent[89]. However, it is difficult to control or measure the driver's dynamic operating state/mode awareness. Observation of behaviour in transition may explain some part of the driver's state/mode awareness although it is difficult to separate the effects of state/mode awareness from those effects caused by other factors. Post-drive subjective assessment may be used to assess the operating state/mode awareness of the driver. However, it remains difficult to assess operating state/mode awareness in dynamic shifts between multiple levels.

7.3.7 Attentiveness

The cause of a driver's inattentiveness to the environment can include visual/cognitive loads induced by NDRAs, drowsiness, or mind wandering. Inattentiveness to the environment may result in low situation awareness.

7.3.8 Receptivity

The driver/operator is expected to be receptive to RtIs and evident vehicle system failures as a fallback-ready user with level 3, according to SAE J3016[1]. Receptivity is considered to be a level of readiness/availability rather than a component of it. When extremely drowsy or sleeping, the driver may not notice the RtI signal that determines the threshold of receptivity. When the driver/operator is concentrating on an NDRA, the same situation might occur. The threshold may be determined by the response time instead of the border between the discrete states of receptive and unreceptive. Design of the HMI issuing the RtI and the HMI for other evident vehicle system failures (e.g. intensity of the signal, sensory modality of the stimulus) may influence the threshold of receptivity.

8 System factors that influence takeover performance

8.1 General

It is known that a driver's takeover performance varies with the influences of multiple factors. In experiments, there is a need to appropriately set factors as variables for investigating their effects or better controlling/eliminating their influence. This will allow for designing experiments that are easier to compare with other studies. There are a number of studies seeking improvements of system design and HMIs for better takeover performance. This clause collects information about system factors that may influence driver takeover performance in critical situations. Exploratory studies for HMIs are included in Annex A. This clause does not provide any design principles to restrict or direct the system design. Driver's takeover performance includes time and quality in the driver state transition phase and the post transition control phase (see Clause 10).

8.2 System behaviour

Some aspects of system behaviour may influence the driver's takeover performance. Therefore, factors related to such system behaviour are expected to be controlled or eliminated in experiments depending on the purpose of the study. When evaluating a driver's takeover performance with commercialized systems, system behaviour and related factors are expected to be known (if obtainable) to correctly interpret the acquired data.

8.2.1 Type of transition

For assessing human takeover performance at downward transitions, it is crucial to distinguish between critical events that are communicated by means of an RtI and critical events that need to be recognized by an attentive driver without an RtI. The latter may occur with level 2 systems operating under specific functional limitations or that terminate the function without notifying the driver due to a DDT performance-relevant system failure (for transitions in non-critical situations, see 9.3.2).

8.2.2 System behaviour within takeover mode

In system-initiated transitions, the driver is expected to take over control of the vehicle and cope with a critical event within the total time budget. The total time budget is defined as the time from onset of an RtI to a critical event (see Figure 1). Therefore, vehicle speed influences the total time budget (i.e. lower speeds have larger total time budgets). It was suggested by Gold et al. (2018)[33] that automatic braking application after an RtI prolonged the total time budget and improved the takeover performance. In level 2 system-initiated transitions, the system may terminate before the driver significantly intervenes to resume control. How long the system continues working before termination may influence the driver's takeover performance. Also, the way in which the system terminates (either suddenly shutting down or fading out) may influence the driver's takeover performance[90].

8.2.3 System-initiated risk mitigation strategy

When the driver does not resume control within the takeover mode (Figure 1) after an RtI, the system may activate an MRM, such as bringing the vehicle to a controlled stop on the road shoulder or in another safe spot and turning on the hazard lamps (SAE J3016[91]; Figure 3). The MRM is optional for level 3 and mandatory for level 4 and level 5 systems. Conditions for activation of an MRM may depend on the environment and the system state and may vary among manufacturers. Human-initiated transitions are still possible during the MRM, although the required method to resume control may differ between the takeover mode and the MRM.

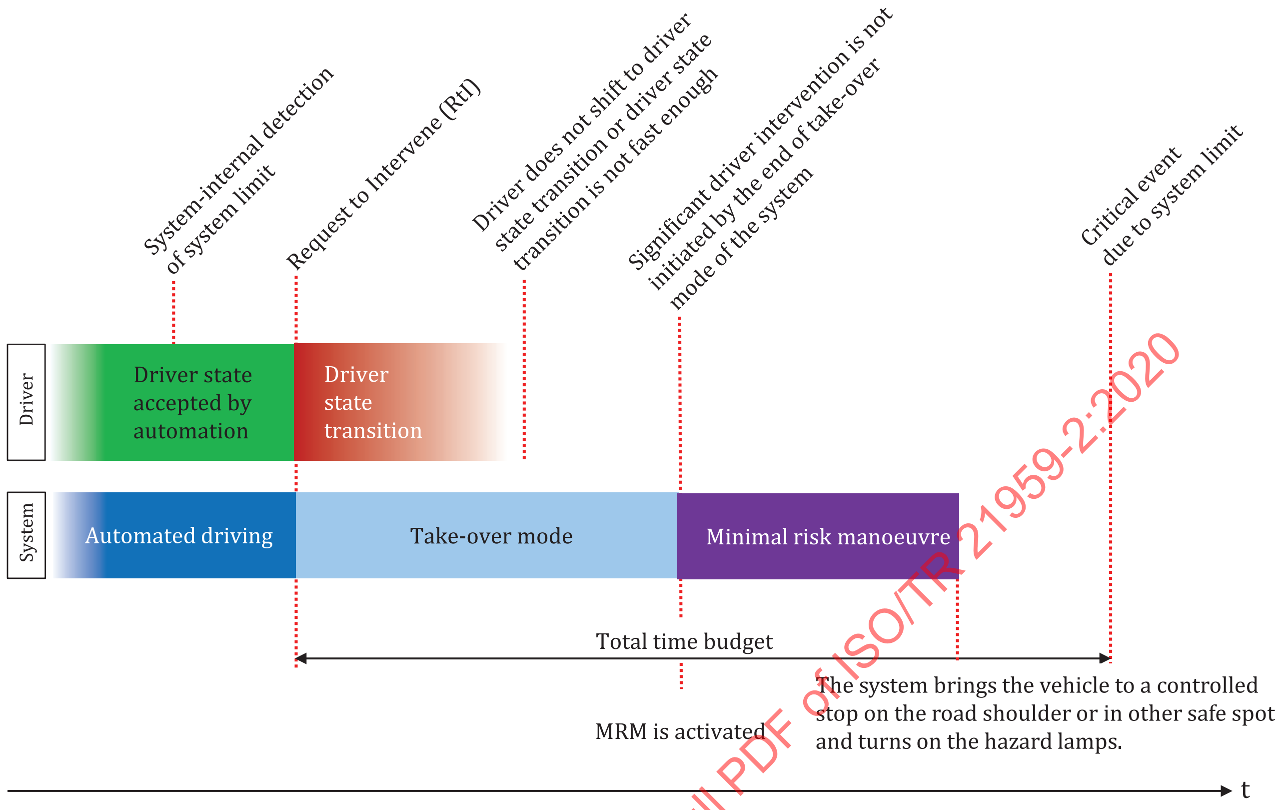


Figure 3 — Transition process model for an MRM by level 3 systems and above

8.2.4 System limitations and failures

A level 2 system may fail to avoid an undetected object/event under its functional limitations. The system may also suddenly terminate without notifying the driver of a failure (i.e. a “silent failure”). In such situations, the driver is expected to detect the object/event or the failure and initiate transition. In experiments, assumed system functional limitations (e.g. what objects can and cannot be detected by the system) may influence the driver-initiated takeover performance in relation to instructions of the system functions/limitations given to the driver (i.e. subject). Experimental conditions for occurrence of “silent failures” (what failure and when it happens) may also influence the performance of the driver-initiated takeover in relation to an instruction of the system given to the driver (i.e. subject) and have an effect on the predictability of an unexpected situation in the scenario.

8.2.5 Stability and reliability of the system functions

Stability and reliability of the system functions may influence a driver’s trust in the system and, as a result, may influence take-over performance[92]. For example, when the longitudinal or lateral control of the system is unstable, with fluctuation in the front gap or in the lateral position, the driver’s trust may remain low and may result in higher vigilance to the system and environment, which may lead to better takeover performance. When the system is unreliable and disengaged frequently, similar positive effects may be seen unless the driver distrusts the system and stops using it. The link between system reliability and trust has been studied by some researchers[92][93][83]. Figure 4 shows a schematic expression of the link between reliability, trust and other factors proposed by Carsten and Martens (2018)[83].

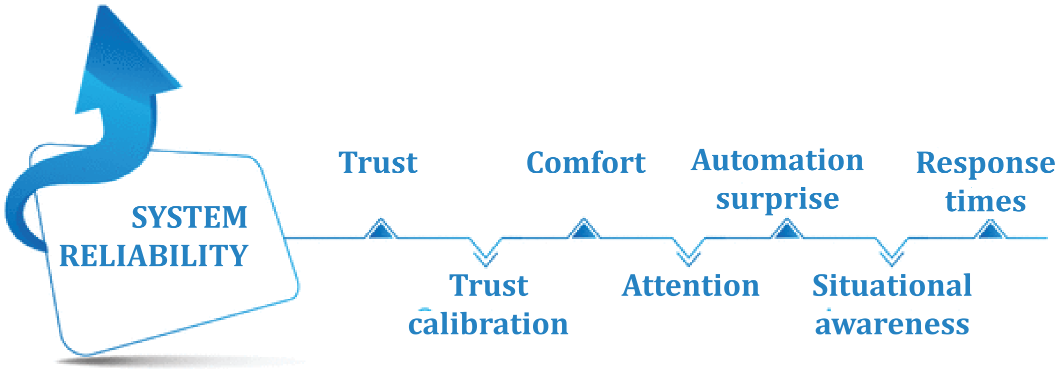


Figure 4 — Relation between system reliability and the concepts trust, trust calibration, comfort, attention, automation surprise, situational awareness and response time (Carsten and Martens, 2018)[83]

8.2.6 Level of automated driving to which the system shifts in transition

The human's safety-critical task is the takeover task in transition from a higher level to a lower level of automated driving. The level of automated driving to which the system shifts in transition may influence the driver's takeover performance. The majority of studies have investigated transitions from higher levels (i.e. levels 2 and 3) to level 0 (fully manual driving). However, a driver's takeover performance may be different when the system shifts from higher levels down to the levels above level 0 (e.g. from level 2 to level 1, from level 3 to level 2). An example use-case is a scenario in which the level 3 system detects a failure in the longitudinal control part of the system. The system issues an RtI and shifts to level 1 while maintaining the lateral control part of the system. The driver's takeover performance may be better than the shift from level 3 to level 0 with simpler re-engagement in driving or may be worse due to confusion about the driver's role (i.e. mode confusion[87][83]).

8.3 Human machine interfaces for RtI

An RtI is issued through an HMI for the system-initiated transitions to alert and shift the driver to the driver state transition within the time budget. There have been a number of studies on the effectiveness of HMI design parameters. When measuring a driver's takeover performance, such design parameters are expected to be set as independent variables to investigate their effects or controlled to obtain results comparable to results from other studies. When evaluating a driver's takeover performance with commercialized systems, specifications of the HMI are expected to be known (if obtainable) to interpret the acquired data correctly.

8.3.1 Design parameters for HMI

HMI design parameters that may influence a driver's takeover performance are listed below.

1) Sensory modality: visual, auditory, haptic and a combination of those.

   — Multimodal signals were generally more effective than unimodal signals for better takeover performance[94][95][96][97][23].

2) Expression of the signal: design of a visual icon, visual and auditory messages with linguistic expressions, tone of an auditory beep and type of a haptic stimulus (frequency of vibration, pressure and others).

   — Linguistic cues appeared to be more advantageous than abstract cues in critical situations, whereas multimodal abstract cues were advantageous in non-critical situations[96].

   — Beeps with shorter inter-pulse intervals were perceived as more urgent, with Stevens' power law yielding an accurate fit to the data[98].

— Intensity: size, colour, brightness and blinking frequency of a visual icon, volume of an auditory alert and magnitude of a haptic stimulus.

— Red blinking visual alert was more effective than yellow stationary alert[95].

3) Location: location of visual, auditory and haptic signals.

— Location of visual signals may influence their noticeability to the driver in relation to the driver's visual field. It is to be noted that the driver is expected to spend much time looking forward to perform the OEDR task with level 2 systems, while the driver may be looking away from forward for an NDRA with a level 3 system.

— Location of visual, auditory and haptic signals could provide the driver with additional directional information about the cause for issuing an RtI (e.g. sound from a right speaker as an RtI for a neighbouring vehicle approaching to the right side of the driver's vehicle). Petermeijer et al.[97][99] found limited effectiveness of auditory and vibrotactile directional information of an RtI in the takeover performance, yet recommended investigating further.

8.3.2 Total time budget

In system-initiated transitions, the driver is expected to take over control of the vehicle and cope with a critical event within the total time budget designed for drivers with targeted attributes and states (see Clause 7). When the total time budget is too short, the takeover performance may result in a crash[35][100][25][22][101][102]. In contrast, it has been also reported that a lengthy total time budget does not improve performance because such an RtI does not raise the driver's alertness[95][103]. When the HMI is designed to express urgency as a function of total time budget (e.g. notifications with longer total time budget and an alert for an RtI with shorter total time budget), the takeover performance was reported to improve[95][104].

8.3.3 Other human machine interfaces to improve drivers' takeover performance

There have been a number of exploratory studies searching for HMI interactions to support the driver for better takeover performance. Information on such studies are summarized in Annex A.

9 Test scenarios

9.1 General

This clause aims to provide considerations on how to derive and document test scenarios for the evaluation of human performance in safety-critical transition situations. This is important in order to replicate empirical results, compare them across studies and, ultimately, to gain sustainable knowledge in this field of research. Characteristics of the testing environment influence human takeover performance in addition to human factors (Clause 7) and system factors (Clause 8), as Radlmayr et al. (2014)[42] showed during an investigation of the effects of traffic density on takeover time and quality.

The major interacting elements of a transition process that can be considered in this context are the human driver (or user of a driving automation system), the system itself and the environment in which the system and the user operate. In order to assess human performance in transition situations, all three elements are expected to be carefully specified as experimental conditions. A test scenario — as defined in this document — primarily covers environmental aspects of the experimental conditions and can be described on different abstraction levels. Hungar (2017)[105] distinguishes between functional, logical, and concrete abstraction levels. Whereas functional scenarios can be described by natural language, logical scenarios specify the parameter space in the state space. Concrete scenarios depict a concrete representative of a logical scenario. The total number of scenarios decreases with the level of abstraction. The status of the system development determines which level is most appropriate. A functional description of scenarios facilitates the understanding of scenarios and the communication between different researchers and developers involved in early phases of the system development.

Logical scenarios are used during system development and concrete scenarios are required for test and validation purposes.

In general, the design of a particular test scenario is strongly linked to the type of transition under investigation (see 9.3) and the chosen human performance measures (see Clause 10).

9.2 Parameters for specifying test scenarios

In order to describe a test scenario in a standardized way, classification schemes can be utilized to make sure all relevant elements of the scenario, including its physical and psychological dimensions, are specified and documented.

The public project PEGASUS[1)], which aimed to derive test specifications for highly automated driving functions, can serve as a basis to describe the physical dimension of a test scenario. It explicitly distinguishes between scenes and scenarios [105].

— A **scene** describes a particular state of the environment in terms of traffic infrastructure (e.g. number of lanes, traffic regulations and road curvature), environmental conditions (e.g. road friction, visibility limitations) and traffic constellation (e.g. own vehicle and others with respect to type, position and speed difference).

— Based on the definition of scene, a **scenario** describes a particular evolution of scenes. It consists of a timed sequence of scenes with a fully defined start scene and transitions between subsequent scenes (including the actions of the ego vehicle as well as events in the surrounding environment). Table 2 includes parameters to describe the physical properties of a test scenario on a functional level.

For a comprehensive description of a test scenario, additional parameters are recommended to describe the environmental conditions preceding the actual test scenario as well as environmental conditions succeeding the critical event. The pre-transition phase can be specifically designed to elicit certain user states, e.g. monotonous driving conditions to reduce human vigilance or highly available system performance to induce automation trust. The test scenario for the post-transition phase may also be specifically designed according to the requirements of the targeted human performance measures, e.g. giving no options for lane change in order to measure standard deviation of lateral position during manual driving.

The public Ko-HAF project[2)] aimed at investigating human takeover performance in the context of level 3 automated driving. Within this project, test scenarios were additionally classified with respect to a psychological dimension. Based on a review of the literature, the factors urgency, predictability, criticality and complexity of driver response have been selected to describe the demands of the test scenario on the user[106].

— The **urgency** of a testing scenario indicates how fast a driver takeover reaction is required. This is associated with the driver's time budget to take over control and react to the system limit. The time budget can be rather limited (e.g. an obstacle is blocking the lane in front of the vehicle) or very long (e.g. approaching a highway exit miles away). Taking over vehicle control can be challenging, as the task switching from possible non-driving related activities (NDRAs) to the manual driving task requires attentional resources and time-consuming reconfiguration of the driver's physical and cognitive state to meet the demands of the respective scenario[107]. Restricting the available time for these processes may increase the speed of the takeover process, as the driver will invest cognitive and attentional resources, but may likewise decrease the quality of the subsequent driver input[35]. The urgency of a test scenario is often manipulated in studies in order to analyse the minimum time budgets a specific system design needs to provide for human takeover.

— The **predictability** of a testing scenario refers to available knowledge about the existence and location of a system limit. A transition situation may be highly predictable for a user due to early notifications of upcoming system limits (e.g. by car-to-X communication or centralized safety servers). But general user knowledge about the ODD of a particular automation feature (e.g. operation limited

1) See https://www.pegasusprojekt.de/en/home .

2) See www.kohaf.de .

to highways) also allows the driver to build expectations about upcoming transition situations. Even dynamic system limits like critical weather conditions can be highly predictable for drivers either due to previous experience or early system notifications. On the other hand, instantaneous system failures, such as malfunctioning of a sensor or unknown functional performance limitations (potentially occurring with level 2 features), are highly unpredictable for users.

— The **criticality** of a testing scenario refers to the cost of failing to take over vehicle control in time. The impact of failing in a takeover scenario can be rather low, like missing a highway exit, or severe, for example when colliding with other road users or obstacles. Therefore, criticality is determined by the situation and the characteristic of the automated systems. A minimal risk manoeuvre, for example, can reduce the risk of a collision and thereby decrease the criticality of a testing scenario. In addition, the test environment (see Clause 11) has significant influence on the perceived criticality; real-road studies with physical collision objects suggest a higher level of criticality than simulator studies or vehicle tests on closed test-tracks.

— The **complexity of the required drivers' response** refers to the human action needed to resolve a transition demand. Depending on the test scenario, different driver reactions may be necessary or appropriate. Depending on the number and type of the required cognitive and motoric operations (see Naujoks et al., 2014)[94] the response may either be complex (e.g. deciding between evasive steering and/or emergency braking) or rather simple (e.g. continue driving in straight lane). Also, environmental factors, such as road shape or behaviour of surrounding traffic, modulate the level of task demand. The complexity of the required driver's response is considered a crucial element in designing a test scenario (see 9.3).

Although the selected factors are related in many cases, testing scenarios can be designed for almost any combination of urgency, predictability, criticality and complexity of driver response. Gold et al. (2017)[106] propose a three-level classification (e.g. low/medium/high) for each factor with indications of when to use which category. Table 2 shows how these four factors can be considered in characterising an exemplary test scenario.

Table 2 — Template to specify a test scenario on a functional level

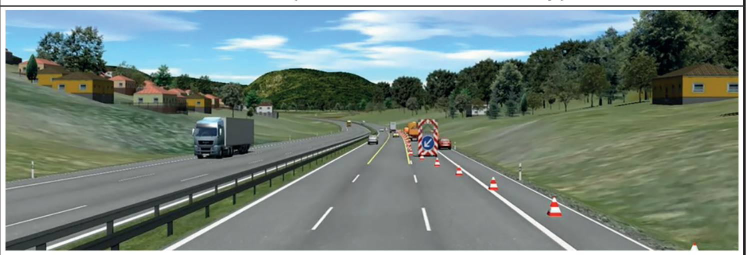
Test scenario parameter		Parameter values for an exemplary test scenario
Physical dimension	Characteristics of automated driving environment preceding the test scenario	20 minutes of uninterrupted automated driving (level 3) on three-lane highway with automatic lane changes in dense traffic
	Characteristics of start scene	**Traffic infrastructure:** Three lane highway, straight section, clear lane markings, traffic pylons indicating end of right lane because of upcoming construction site **Environmental conditions:** Good visibility, no sun glare, dry road **Traffic constellation:** Ego vehicle driving in middle lane at a speed of 80 km/h; no surrounding traffic
	Evolution of test scenario	System-initiated RtI for full manual control of the vehicle 150 meters before entering the construction site (system limit): reduction of lane width to 2,5 meters within construction site
	End of test scenario	End of construction site (500 meters after entry)
	Visual sketch of relevant scene(s)	

Table 2 *(continued)*

Test scenario parameter		Parameter values for an exemplary test scenario
Psychological dimension	Urgency of driver intervention	Medium: RtI activation 7 sec before reaching critical event (entry of construction site)
	Predictability of driver intervention request	Low: Test subjects do not know in advance about this particular system limit
	Perceived criticality in case of missing intervention	Medium: Automation feature is deactivated at the end of the takeover mode; uncontrolled vehicle may cause collisions with infrastructure
	Complexity of required driver response	Low: Driver can easily resume control by intervening on primary driving controls; no change of speed or lane required

9.3 Considerations for selecting/designing adequate test scenarios

In order to adequately select and design a test scenario, the aim of the assessment and the type of transition under investigation can be clarified first. Depending on the scope of the study, considerations about how to optimally set high level scenario parameters (e.g. criticality, urgency, predictability and complexity of the required driver response) can be given.

9.3.1 Investigating driver state transitions during automated driving

There are transitions that do not require the driver to completely resume the DDT but rather request a different driver state. A downward transition from a level 3 to a level 2 automation feature, for example, requires the driver to understand the meaning of the associated HMI message and to perform the OEDR task. Another example relates to critical driver states during level 2 automated driving: if a driver monitoring system (see A.2.2) detects critical levels of driver inattention, HMI messages may be issued to raise the driver's vigilance.

In order to assess required driver state transitions during automated driving, a test scenario could include relevant objects and events that are expected to be monitored by the driver. As the focus of this research primarily relates to the user's (mental) understanding of HMI messages (without the need to cope with a specific road hazard), test scenarios can be designed to be rather generic. A potential test scenario might be one in which automated driving takes places on a generic two-lane highway with mixed traffic and regular system-initiated lane changes.

9.3.2 Investigating takeover performance in non-critical transitions

Another type of transition relates to user-initiated switches between automated and manual driving in non-critical situations. Possible examples are as follows:

— During conditionally automated driving, the driver likes to drive manually and requests manual control without being requested.

— The driver gets an early notification about the upcoming end of the ODD ("End of automation in about 3 km. Please prepare for taking over."). The driver chooses to take over sooner rather than waiting until the last moment.

Research questions for these types of transitions usually refer to acceptance, driving comfort or usability. Since the transitions are self-initiated (and thereby self-regulated) test scenarios can be derived from common or standard situations within the ODD. Accordingly, the situational urgency can be kept at a low level along with low criticality and low complexity of required driver reaction. A potential test scenario might be releasing/resuming control on a straight road with surrounding traffic.

9.3.3 Assessing takeover performance at system limits

Most of the reported empirical studies on human performance in the context of automated driving focus on how human drivers are capable of handling system limits requiring the driver to takeover manual control. System limits may either be communicated by an RtI or not (see 6.3). System manufacturers may obtain knowledge about system limits by simulation of the system-environment-interaction during the development phase or by analysis of available field data for a particular system. System limits are either based on environmental characteristics (e.g. missing lane markings or critical event on the road) or independent of environmental aspects (e.g. internal system failure).

The focus of this research is typically on assessing the driver's performance limits, which requires a non-predictable, urgent and critical test scenario. Delayed or missing driver responses could lead to a critical situation. Otherwise, if the situation is not urgent and critical, there may be no need for the driver to exhibit maximum performance levels. Test scenarios for assessing human performance at system limits can be designed with varying levels of required driver response. The following sections describe potential driver tasks at system limits and considerations for how to treat them in the design of a suitable test scenario.

9.3.3.1 Low complexity of driver response

In a simple driver response scenario, the driver's task is to take over manual control and smoothly continue driving without having to deal with further safety threats. Driver intervention at this level typically refers to the DDT on a skill-based level[108]. In order to make human performance measures more sensitive, the scenario may include non-critical curves and other traffic participants acting in a non-critical way. These types of test scenarios may be most representative for real-life situations, in which these "non-critical" situations might be most prevalent, making it possible to determine comfort and acceptance aspects of human performance in routine situations.

9.3.3.2 Medium complexity of driver response

At the next level, test scenarios can be designed to require instant longitudinal or lateral control interventions, typically performed at a skill- or rule-based level[108]. Events can be included in the test scenario either referring to specific system limits (e.g. upcoming critical curvature of road) or to test human performance under more demanding situations. In order to assess the quality of longitudinal intervention, a test scenario could require a brake (only) reaction, for example, caused by a strongly decelerating or stopped vehicle in the ego lane. On this level, no option to change lanes could be given. In order to assess the quality of lateral control interventions, a test scenario can include an event specifically disturbing lateral control, e.g. a wind gust that needs to be compensated by a steering reaction. Medium levels of response complexity are mainly considered in empirical studies[118] since they allow drawing conclusions on human performance limits under challenging (yet representative) takeover conditions.

9.3.3.3 High complexity of driver response

On this level, test scenarios may require higher cognitive processes, such as problem-solving and decision making, typically performed at a rule- or knowledge-based level[108]. In order to integrate this aspect, test scenarios can be designed with an obstacle in the ego lane that requires the driver to decide between two collision mitigation strategies: full braking in the ego lane vs. changing to a non-blocked lane (see Gold, 2013). Also, decisions on a strategic level (e.g. selecting the correct route) can be integrated in test scenarios (see Damböck, 2013[109]). According to the specific system design or the experimental condition, the available total time budget may vary. Test scenarios requiring complex driver responses may be most sensitive to detect performance decrements; however, they may not be representative of real-life situations.

10 Takeover performance

10.1 Introduction

Human performance addresses the assessment of human intervention in takeover situations with respect to its effects on traffic safety. A broad range of human performance measures can be used in empirical studies with specific properties.

The assessment is primarily based on comparing human performance in transition situations with human performance in a baseline condition. Baseline conditions typically involve driving a comparable test scenario at the same automation level the transition leads into (e.g. coping with the test scenario in manual driving or assisted driving). Alternatively, the assessment can also be based on predefined normative acceptance thresholds of driving performance (such as driving errors) or traffic safety (such as near misses). The proposed measures can also be used to compare different experimental conditions in HMI studies, such as different RtI modalities.

In this clause, a taxonomy is introduced to organize different measures according to their unique profile (see Figure 5). The following sections will introduce the organizing principles and give examples of related human performance measures.

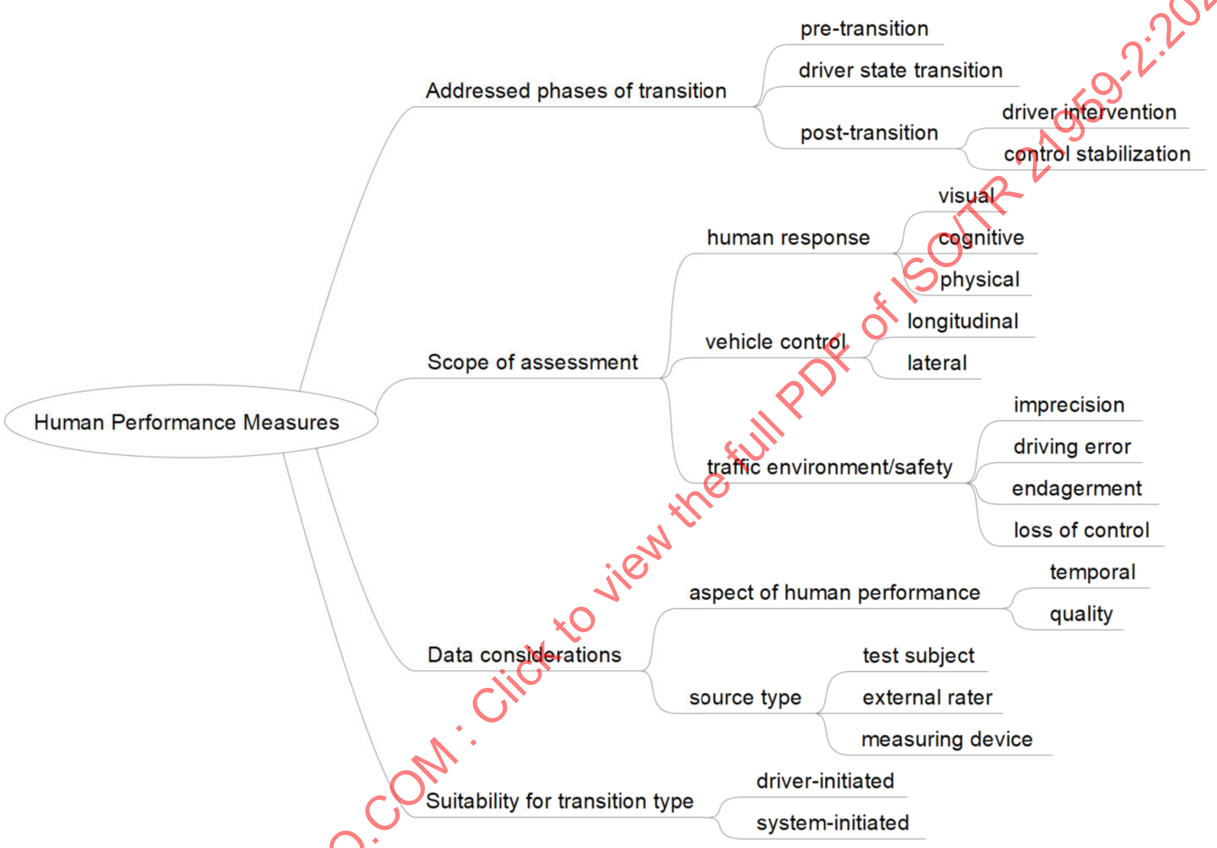


Figure 5 — Taxonomy of human takeover performance measures

10.2 Taxonomy of human performance measures

10.2.1 Addressed phases of transition

The transition from a higher level to a lower level of automated driving can be characterised in four stages: pre-transition, driver state transition, post transition-driver intervention and post transition-control stabilisation. Each phase has its own characteristics and set of potentially relevant performance measures that can be used to capture human performance and human interaction with the control of the vehicle. A brief description of each phase and example of relevant measures are provided. For a complete overview of relevant performance measures for each phase see Table 3.

NOTE Physiological measures such as EEG and ECG were not included in Table 3 because they more specifically assess driver state rather than takeover performance.

a) **Pretransition:** In the pretransition phase, the vehicle is performing part or all of the DDT. However, the activities and performance of the driver during this phase may be of interest since these may influence performance in the later phases of transition (e.g. driver readiness/availability) for both driver- and system-initiated takeovers. Relevant types of human performance at this stage include whether there is engagement in an NDRA/NDRT and the nature of that engagement (perceptual, cognitive, manual; duration of engagement and demand/workload associated with the NDRA/NDRT). Measures of human visual behaviour may be used to provide an indication/confirmation of level of engagement (or lack thereof) with respect to the driving environment and contribute to assessments of workload and situation awareness. Subjective ratings by the driver and expert assessors as well as open-ended questions may also be used.

b) **Driver state transition:** This phase covers the period and processes during which the driver transforms from a “not-in-motion-control” state to a driver state that is suitable to resume more aspects of the driving task. There are a number of important changes that can take place during this time. The challenge of measuring performance during this phase is that there are many possible actions/behaviours that may be relevant. The particular measures may vary across driving situations, vehicles, environments, drivers and for a given driver in different situations. Many of these measures involve reaction times. Multiple performance measures will likely be required to properly characterise driver behaviour. Clearly defined operational definitions are essential. Video analysis of driver behaviour using naturalistic driving approaches may provide valuable insight for performance during this phase. There are many relevant human performance activities and related measures for this phase. Examples include time/latency with respect to first driver reaction, noticing the RtI, start time for a movement to first reaction (brake, steer, accelerate), measures of hesitation and disengagement with a NDRA/NDRT. Measures of visual behaviour/performance are also useful here and include reorientation of visual attention to forward view and a wide variety of other visual performance measures. Subjective ratings of experience and performance and expert ratings are also relevant.

c) **Post transition-driver intervention:** In this phase, the driver initiates specific actions to regain more aspects of the DDT (or to change to a different level of automation). This may be accomplished by taking control of steering and/or braking of the vehicle or by using an activation/deactivation control, such as a button or switch. In this phase, relevant performance measures fall into two broad categories: human response (often physical) and vehicle control. Examples of relevant human performance activities and related measures include decision time, performance (time, quality) of first reaction (brake, steer, accelerate) and following actions. Measures of visual behaviour are also relevant. Vehicle control measures include the many methods of characterising driving in terms of lateral and longitudinal control, including acceleration, braking, time-to-collision (TTC), time-to-lane-crossing (TLC), standard deviation of lane position (SDLP), etc. The more qualitative aspects of this phase are captured in driver subjective ratings (e.g. workload) as well as expert ratings, which provide information on operating errors, safety relevant measures and other quality measures.

d) **Post transition-control stabilisation:** In this last phase, the performance measures of interest characterise what is required to reach a suitable level of manual driving performance after the transition is completed. Comparisons may be made to what would be expected by an average driver or in comparison to a driver’s own driving when automation is not engaged or some other standard for comparison. Examples of relevant human performance include measures of visual behaviour, measures of engagement/interaction with vehicle controls and displays (e.g. checking mirrors) and interactions with the driving environment and other vehicles. There are many relevant vehicle control measures, such as time to stabilise driving performance, steering, acceleration and braking profiles, TTC, TLC, SDLP, etc. There are many measures of expert evaluation and driver self-assessments that can be applied to assess the quality and safety of performance at this stage.

10.2.2 Suitability for transition type

Measures for takeover performance differ based on their suitability to defining driver-initiated versus system-initiated transfers of control, as these two types of transition have different stages of transition. The significant difference in measurement between the two is the initiating factor of either the driver deciding to intervene or the system issuing an RtI. From the point of significant driver intervention, these two transition types have the same set of relevant measures. For a complete overview of transition measures according to this organizing principle, see Table 3.

a) **Driver-initiated transitions** include 1) driver's detection of an object/event or system failure, 2) driver decision to resume manual driving, 3) start of significant driver intervention, 4) driver intervention time and 5) control stabilisation time.

b) **System-initiated transitions** include 1) RtI, 2) driver takeover time, 3) start of significant driver intervention, 4) driver intervention time and 5) control stabilisation time.

10.2.3 Scope of assessment

Traffic safety can be understood as a complex and dynamic interaction of human response, vehicle control, and traffic environment. Accordingly, human takeover performance could ideally be measured on a system level considering all three elements. A number of human takeover performance measures solely address the human driver as the initiator of a control transition and quantify details about the human response (e.g. time to react to an RtI or omission of mirror glances). On a next level, the effect of the human driver (or other vehicle systems) on vehicle control can be quantified by measures such as maximum longitudinal acceleration or yaw rate. Finally, there are measures that refer to the effects of the vehicle's motion on the traffic environment. TTC or TLC are prominent examples of this type of measure, incorporating all system elements — the driver, the vehicle and the environment. For a complete overview of transition performance measures according to this organizing principle, see Table 3.

a) **Human response**

The process for resuming control involves a set of resources described by visual, cognitive, and physical components. Adapted from descriptions within Zeeb et al. (2015)[25], and McDonald et al. (2019)[110], a takeover starts at the presentation of a salient, precipitating event, and initiates perceptual, cognitive, and physical readiness processes. These processes are individually measurable, and can generally be grouped as follows:

— Perceptual processes (especially related to visual processes) include redirecting gaze to the forward scene (if away from the forward road) and scanning the roadway to gather information to support action selection and evaluation.

— Cognitive processes include cognitive readiness, action selection and evaluation.

— Physical processes include motor readiness and action execution. Specifically, these processes involve repositioning the hands to the steering wheel and the feet to the pedals in measuring readiness, and in the amount and type of steering/braking input to execute an "as necessary" action in measuring action execution.

Measures for takeover performance can be described in terms of the human resource type(s) involved (see Table 3 for an overview of transition measures according to this organizing principle).

b) **Vehicle control**

Vehicle control is characterised by measures of longitudinal and lateral control.

— Longitudinal control measures reflect a driver's inputs to increase and decrease vehicle speed. Performance measures include acceleration, braking and speed variability but also apply to maintenance of a safe following distance, TTC and speed in relation to posted speed limits.

— Consistent lateral position is often taken as an indicator of vehicle control. When lane changes are intentional, measures of variability during the lane change (smoothness) and timing both indicate degree of control. Measures of lateral control can be obtained from variations in a driver's steering inputs, vehicle measures such as SDLP and TLC, as well as subjective ratings by either drivers or expert observers.

c) **Traffic environment/traffic safety**

On the highest level, human-performance measures indicate the effects of a human's intervention in vehicle control on traffic safety. In particular, criteria-based performance measures can be attributed to specific levels of how traffic safety is affected. This subclause adopts a four-level classification of safety threats previously proposed by Naujoks et al. (2018)[31] for an expert-based rating of the take-over controllability (TOC-Rating). For a more complete list of corresponding human performance measures, see Table 3.

1) Driving imprecisions (decreased quality of human vehicle control or situation awareness not yet having a direct impact on safety).

Example measure: Increased standard deviation for lateral position or steering wheel angle.

2) Driving errors (sub-optimal manual vehicle control or situation awareness not yet causing harm to other traffic).

Example measure: effective, but inappropriate longitudinal intervention (too early, too strong).

3) Endangerment of oneself or others (takeover situation or human intervention harms driver, passengers, or other traffic participants not yet causing damages).

Example measure: Occurrence of near misses.

4) Loss of control (takeover situation or human intervention leads to damage to people and/or objects).

Example measure: Percentage of test subjects causing the vehicle to run off road.

10.2.4 Data considerations

There are many ways in which performance data relevant to the transition process can be further categorized. Two additional categorisations follow:

a) **Aspect of human performance**

Measurements for human takeover performance can be defined by temporal and quality aspects of transition. Because a transfer of control is time-based, and its success depends on the quality of the transition, each performance measure can be grouped by this organizing principle.

— Temporal measures of takeovers reference driver response times from the precipitating event, the first demonstrable steering or pedal input from the driver and the resumption of stable control. Measures include time between the RtI and the redirection of the driver's gaze, return of the hands or feet to the controls, automation deactivation and the initiation of the last evasive action.

— Quality measures of takeovers reference driver control performance. Measures include lateral and longitudinal acceleration values, and collision imminence values such as TTC and TLC. A complete set of metrics used to measure takeover time and quality are provided in Table 3; a recent review article extensively defines and references each of the time- and quality-based measures of take-over performance[110].

b) **Source type**

Measurements for human takeover performance can have different sources. In this overview, three categories of data sources are distinguished: the test subject (driver), external rater(s) and measuring devices. In order to analyse the driver's subjective assessment of the takeover performance, rating scales (e.g. assessing workload) can be used as a measure. Other types of performance measures require an external observer who rates a particular driver behaviour or the effects on the traffic environment. Often, the detection of driver errors fall into this category. Since human raters cannot be fully objective by nature, minimum requirements on interrater reliability are required. Finally, technical devices to measure parameters of the takeover performance, such as onboard sensors (e.g. CAN bus data) or additional technical equipment (e.g. eye-trackers), can be used. The different types of data sources have unique diagnostic quality profiles in terms of objectivity, reliability and validity. For a complete overview of transition performance measures according to this organizing principle, see Table 3.

10.3 Overview of measures and characteristics

Table 3 illustrates the described organizing principles by listing exemplary human performance measures with their associated profile. The collection of measures is not intended to be exhaustive, nor does it prescribe or recommend specific measures. For a detailed description of many measures on human performance the SAE J2944[111] can be referred to as a main reference.

Table 3 — Measures of takeover performance

	Addressed phases of transition				Suitability for Transition type		Scope of assessment									Data considerations					
							Human response			Vehicle dynamics		Traffic environment / Traffic safety				Aspect of performance		Source type			
Human Performance Measures (examples)	Pre-transition	Driver State transition	Driver intervention	Control stabilization	Driver-initiated	System-initiated	Visual	Cognitive	Physical	Longitudinal	Lateral	Imprecision	Driving error	Endangerment	Loss of control	Temporal	Quality	Test subject	External rater	Measuring device	Example reference
Assessment of collision severity / injury level (e.g. AIS)			x		x	x									x		x		x	x	Civil et al., 1988 [112]
Control stabilization time				x	x	x						x				x				x	Naujoks, 2019 [113]
Crash rate			x		x	x									x		x		x	x	Damboeck, 2013 109]
Decision time (related to "silent failures")		x			x			x	x							x			x	x	ISO/TR 21959 Part 1 [2]
Distance headway to the lead vehicle			x	x	x	x				x							x			x	Louw et al., 2015 [117]
Effective, but inappropriate longitudinal intervention (too late, too early, too strong)			x		x	x							x				x		x		Naujoks, 2018 [31]
Expert based ratings of human take-over performance (e.g. TOC ratin g)		x	x	x	x	x						x	x	x	x		x		x		Jarosch, 2018 [114]
High frequency steering control input				x	x	x			x		x						x			x	Merat et al., 2014 [100]
Intervention time			x		x	x				x	x					x			x	x	ISO/TR 21959 Part 1 [2]
Inverse time to collision			x		x	x				x							x			x	Wiedemann et al., 2018 [116]
Lane change error rate			x	x	x	x							x				x		x		Kerschbaum et al., 2015 [115]
Lateral acceleration			x	x	x	x					x						x			x	Zeeb et al., 2016 [36]
Lateral control errors (uncritical lane exceedance, low TLC values)			x	x	x	x					x		x				x		x		Naujoks, 2018 [31]
Longitudinal acceleration			x	x	x	x				x							x			x	Radlmayr et al., 2014 [42]
Maximum derivative of the control input that drivers used to avoid the collision			x		x	x				x	x						x			x	Louw, Markkula, et al., 2017[84]
Maximum resultant acceleration			x		x	x				x	x						x			x	Wandtner et al., 2018b [37]
Maximum steering wheel velocity			x		x	x					x						x			x	Wiedemann et al., 2018 [116]
Method to override / deactivate automation feature		x			x	x			x	x	x						x			x	Naujoks, 2019 [113]
Number of test subjects causing a collision with other traffic participants			x	x	x	x				x					x		x		x	x	Gold, 2013 [35]
Number of test subjects causing the vehicle to run off road			x	x	x	x			x		x				x		x		x	x	Zeeb et al., 2016 [36]
Occurrence of near misses (defined by min TTC)			x	x	x	x								x			x		x	x	Naujoks, 2018 [31]
Omission of required visual checks/mirror use			x	x	x	x	x						x				x		x	x	Naujoks, 2018 [31]
Operating errors			x	x	x	x		x	x				x				x		x	x	Naujoks, 2018 [31]
Percentage eyes-on-road	x	x	x	x	x	x	x	x									x			x	Morando et al., 2016 [119]
Percentage glance durations > 2.0s	x			x	x	x	x	x									x		x	x	Victor et al., 2005 [120]
Reduced awareness of traffic rules			x	x	x	x		x				x					x		x	x	Naujoks, 2018 [31]
Remaining action time		x	x		x	x						x	x	x	x	x				x	Gold, 2013 [35]
Side mirror gaze time		x	x			x	x									x				x	Vogelpohl et al., 2018 [121]
Speedometer gaze time			x	x		x	x									x				x	Vogelpohl et al., 2018 [121]
Standard deviation of lateral position				x	x	x					x						x			x	Naujoks et al., 2017 [46]
Standard deviation of steering wheel angle				x	x	x					x						x			x	Clark & Feng, 2017 [18]
Standard deviation of velocit y				x	x	x				x							x			x	Brandenburg & Skottke, 2014 [104]
Subjective criticality rating of takeover situation		x	x	x	x	x						x	x	x	x		x	x			Berghoefer, 2018 [122]
Takeover time (with respect to RtI)		x				x			x							x				x	ISO/TR 21959 Part 1 [2]
Time headway to the lead vehicle			x	x	x	x				x							x			x	Zeeb et al., 2017 [118]
Time to complete a lane change			x	x	x	x			x		x					x				x	Louw et al., 2015 [117]
Time to first driver reaction (e.g. interruption of non-driving related task)		x				x			x							x				x	Berghoefer, 2018 [122]
Time to lane crossing			x	x	x	x					x						x			x	Zeeb et al., 2017 [118]
Time to move (at least one) hand to wheel / feet to pedals (beginning of movement);		x				x			x							x				x	Berghoefer, 2018 [122]
Time to operate relevant vehicle controls (e.g. blinker) or steering / pedal operation (beginning of operation)		x				x			x							x				x	Berghoefer, 2018 [122]
Time to start of visual re-orientation		x				x	x	x								x				x	Berghoefer, 2018 [122]
Time to touch pedals (feet-on reaction time)		x				x			x							x				x	Petermeijer, 2017 [99]
Time to touch steering wheel (Hands-on reaction time)		x				x			x							x				x	Petermeijer, 2017 [99]
Time to visually fixate road centre		x				x	x									x				x	Berghoefer, 2018 [122]
Type of first intervention in vehicle control (e.g. steering input, braking, accelerating, ...)			x		x	x			x								x			x	Gold, 2013 [35]
Type of first user reaction (e.g. change of gaze direction, foot movement, hand movement, ...)		x			x	x	x	x	x								x			x	Naujoks, 2019 [113]
Violation of (pre-defined) safety distance long./lat.			x	x	x	x							x	x			x		x		Naujoks, 2018 [31]
Violation of speed regulation or other traffic rules				x	x	x				x	x		x				x		x		Naujoks, 2018 [31]
Workload scales (e.g. NASA-TLX, DALI)	x	x	x	x	x	x		x									x	x			Wandtner et al., 2018 [37]
Yaw rate error				x	x	x					x	x					x			x	Eckstein, 2000 [123]

11 Testing environments

11.1 General

This clause describes different test environments typically applied when developing and evaluating HMI solutions in the context of driving automation systems. Numerous factors influence the selection of test environments during the development cycle of driving automation systems and their associated HMIs. These include, but are not limited to, types of research questions under investigation, safety reasons, HMI prototype maturity, availability of time and financial resources, etc. Thus, while simulator studies may be more beneficial for inexpensive and rapid iterations of early-stage HMIs, roadway studies may be more useful for later-stage realistic safety validation. This clause will describe the various test environments, as well as the associated trade-offs commonly observed.

11.2 Types

Testing environments may vary in their level of controllability, fidelity, and validity. When testing driving automation systems and their corresponding HMIs, it is beneficial to have a variety of testing environments to choose from depending on the variables of interest. This clause will break down these different testing environments into two primary categories: simulator studies and roadway studies.

11.2.1 Simulator studies

Due to the ease of implementing new testing procedures and the assurance of high levels of control over driving variables, simulator studies have been a preferential testing environment (particularly in early stages of development) for new HMI solutions. Known benefits of simulator studies include the following[124]:

— high level of control over the driving variables;

— repeatability of the test scenario;

— objective performance scoring;

— relatively low cost of implementation;

— greater levels of safety for the experimental participant;

— test scenario can include safety-critical situations.

Although these benefits are common to different typologies of driving simulators, there is still a great deal of variability regarding technical complexity, fidelity, and level of immersion. One crucial technical distinction between driving simulators is regarding the existence of a motion platform. In that regard, we list below the different categories of fixed-based driving simulators (11.2.1.1) and motion-based driving simulators (11.2.1.2).

11.2.1.1 Fixed-based

Fixed-based driving simulators enable the experimental participant to control the vehicle (longitudinally and laterally) using a mock steering wheel and pedals. These types of simulators may vary in their realism of vehicle controls and the complexity of the visualisation system. Table 4 shows the distinction between two levels of fixed-based driving simulators.

Table 4 — Classification for fixed-based driving simulators (adapted from Parkes, 2012)[124]

Level	Technical requirements	Target studies
A	**Lowest technical complexity for driving simulators:** — Enables control of the movement (vertical and lateral) of the ego vehicle — Control through the use of mock steering wheel, pedals or joystick — Kinaesthetic feedback for driving controls are not provided — Gearshift may not be included — Single channel of projection and small field of view — No mock-up vehicle body or realistic vehicle cab — Interactivity with other simulated road users is low — Low number of simulated road environments	— Design audits — Expert reviews — Comprehensibility and learnability tests — Familiarisation with new HMI
B	**Includes all features from level A plus:** — Full provision of vehicle controls, including pedals, gearshift and steering wheel — Provision of a mock-up vehicle — Realistic feel of vehicle controls (e.g. steering wheel force-feedback) — Wider field of view through multi-channel projection — Provision of side and rear views — 3D acoustic stimuli — Considerable number and realism of road scenarios and road user behaviour	**As for level A plus:** — HMI interaction assessment (including human reaction times to RtI) — Assessment of perceived criticality

11.2.1.2 Motion-based

Motion-based driving simulators often include a mock-up vehicle or a realistic vehicle cab coupled with a motion platform for simulation of rotational accelerations in pitch, roll and yaw. Some driving simulators of this type might also simulate longitudinal accelerations using rail motion systems. Table 5 shows the distinction between two levels of motion-based driving simulators.

Table 5 — Classification for motion-based driving simulators (adapted from Parkes, 2012)[124]

Level	Technical requirements	Target studies
C	**Includes all features from level B plus:** — Provision of realistic vehicle cab — Includes a 6 degree of freedom motion platform — Change of underlying vehicle model possible — Addition of advanced driver assistance systems possible — Large number and realism of road scenarios and road user behaviour — Behaviour of other road users can be influenced — Simulation of different atmospheric conditions	**As for level B plus:** — Driver performance assessment (with relative validity) during human or system-initiated transitions — Training of complex manoeuvring tasks including interaction with other road agents
D	**Includes all features from level C plus:** — Provision of 1 degree of freedom yaw ring (turntable) and extended x and y motion systems (rails)	**As for level C plus:** — Wider range of complex manoeuvring tasks and increasing fidelity

This categorisation is an indication of the most common technical requirements in each category. It is possible that some driving simulators do not fall completely into one category, fulfilling some requirements of high-order categories while framing most of the technical specifications in lower categories.

11.2.2 Roadway studies

Similar to simulator studies, there are a variety of different roadway study approaches that can be leveraged to evaluate driving automation systems and corresponding HMIs. We propose three primary categories, which are described below: test-track, field operational tests and naturalistic driving studies.

11.2.2.1 Test-track

Test tracks can provide a more realistic driving environment but still have the added benefit of experimental control. Test-track research evaluating human performance in the context of automated driving[95] can investigate many of the same variables as simulator research (as mentioned above) with the added benefit of quantitative measures that are more comparable to real world driving.

11.2.2.2 Field operational tests

According to the FISTA Handbook (Version 7, 2017)[125], a field operational test can be defined as, "A study undertaken to evaluate a function, or functions, under normal operating conditions in road traffic environments typically encountered by the participants using study design so as to identify real-world effects and benefits." Typically, this requires having a baseline condition for appropriate comparisons.

While on-road studies present the most realistic environment in which to measure driving performance, they also present the riskiest driving environments in that experimenters cannot control all elements of the driving demands. Using other research vehicles through quasi-choreographed scenarios allows control of certain environmental elements during trials, but there are still many other factors that cannot be controlled, such as other drivers, traffic density, etc.

Some research teams will use a combination of both test track and field tests to ultimately provide a very realistic experience with driving automation systems, but during certain trials have a controlled setting for evaluating the transition processes[126]. While in the field, it is important to understand the risks involved with evaluating control transitions and to provide the right balance between realism and safety.

11.2.2.3 Naturalistic driving studies

Human performance in the context of driving automation systems has been evaluated in simulators, test-track studies and field operational tests. While these methods can be helpful in identifying the benefits and drawbacks of driving automation systems, naturalistic driving studies may also be useful in furthering our understanding of each level of automation. According to ISO/TR 21974-1:2018[127], a naturalistic driving study is defined as "any driving study where research subjects are recruited to drive on public roads (not in a simulator or on a test track), where there is no in-vehicle experimenter or confederate vehicles, and where driving conditions are not experimentally controlled or manipulated."

Common requirements of this type of study are that participants not be provided any instruction(s) on how to drive, equipment used to collect data can be unobtrusive, and participation lasts for an extended period of time (several weeks to possibly even years). Vehicles hosting driving automation systems can be instrumented with a variety of sensors, cameras, data recorders and drivers can be continuously recorded as they do their daily driving over some specified participation period. Naturalistic driving studies allow for a robust evaluation of driver performance, vehicle performance, the context of the surrounding environment (e.g. roadway types, other vehicles, weather) and how drivers use driving automation systems in real-world driving. Naturalistic driving studies may help address gaps in research that simulator, test-track and some field studies cannot. For example, how drivers engage and disengage a driving automation system over time (e.g. weeks, months).

Recently, two naturalistic driving studies were completed that focus on vehicles with driving automation systems. Fridman et al. (2017)[128]recently completed a naturalistic driving study that included 78 participants and over 275 500 miles (443 374 km) of data collected using approximately 25 production vehicles with driving automation systems. Also recently completed, Russell et al. (2018)[129] completed a naturalistic study of level 2 driving automation functions that included 120 participants and over 216,000 miles (347 618 km)of data collected [70 384 miles (113 272 km) driven with both lateral and longitudinal driving automation system features active] using 10 production vehicles with driving automation systems. Results of this study found drivers were engaging in non-driving tasks; however, somewhat surprisingly, these were not related to driving automation system feature use. In addition, RtIs were not associated with any safety-critical events. (No statistical relationship between safety-critical event rates and driving automation system activation level.)

11.3 Advantages and disadvantages

11.3.1 Realism-to-safety trade-off

Different test environments offer different levels of realism. Typically, experimental test environments that can include all of the constraints of a real-world roadway event are regarded as a more effective way of generalising results to the real world. However, there are trade-offs that occur when researchers strive for realism while also considering the safety of all roadway participants involved, which are described in the following sections.

11.3.1.1 Benefits and limitations of roadway studies

Field and naturalistic studies are often regarded as the ultimate validation stage for assessing behavioural models, safety-critical HMI functions, safety measures, and new designs of road infrastructure or vehicle equipment. Experimental participants are in contact with similar versions of the systems that will be later deployed and the environmental conditions are close to what participants will experience in the “real world”. The realism of field and naturalistic studies allows new findings to be more easily transitioned into new policies, directives, or regulations, making this type of test environment the most suitable for generalisation of results. In addition, naturalistic studies often identify unexpected driver behaviours, such as misuse or even disuse of driving automation systems over time. However, ethical, technical and economic factors may constrain the feasibility of field and naturalistic studies in some situations. Santos et al. (2005)[130] stressed some limitations of these types of studies:

— Experimental participants have to be protected from hazardous traffic conflicts.

— The surrounding traffic cannot be controlled.

— To record, synchronise and analyse the relevant data is harder and more time consuming when compared with simulator studies.

— Planning and execution of these types of studies can be time consuming (when compared with other test environments).

11.3.1.2 Benefits and limitations of simulator studies

The use of driving simulators in transportation research has been greatly motivated by the need to have a relatively easy way to rapidly implement a test scenario. A driving simulator is capable of solving some of the disadvantages of field and naturalistic studies, as previously mentioned. Simulators can provide a high level of controllability over experimental variables, such as road configuration, traffic density and the number and type of hazard situations. Driving simulators are also capable of reproducing and repeating any well-specified driving situation without putting the driver at physical risk. Finally, driving simulators allow for high levels of precision and frame rate acquisition and the easy implementation and synchronisation of new research tools, such as eye and motion trackers.

Despite the advantages driving simulators can bring to transportation research, there are still some drawbacks resulting from the use of this type of test environment over that of field and naturalistic approaches. De Winter et al. (2012)[86] and Purucker et al. (2018)[134], provide some known disadvantages of driving simulators:

— Simulator discomfort and motion sickness are possible, affecting especially older drivers and experimental participants in highly demanding test scenarios.

— Limited physical, perceptual and behavioural fidelity. This is particularly relevant in fixed-based simulators of low-fidelity categories and, when detected, these limitations could help to constrict the type of tests that might be selected.

— Inconclusive research demonstrating validity of simulation. There is a growing body of knowledge indicating that driving simulator findings are predictive of on-road behaviour[130][131][132]. Nonetheless, comparison between simulator and on-road behaviour often reveals relative validity instead of absolute validity[133].

— Training efforts. To overcome some of the abovementioned shortcomings and to observe realistic behaviour, lengthy simulator training sessions might be necessary.

These disadvantages highlight the two main constructs researchers need to consider when using simulators for performance and behavioural research: fidelity and validity. While fidelity is concerned with multisensory realism and the capacity to accurately simulate the visual, auditory and motion/haptic environment of road scenarios, validity is concerned with the capacity that a driving simulator shows to replicate on-road driver performance behaviour.

One of the few studies exploring the simulator validity problem in driving automation systems was conducted by Erikson et al. (2017)[132]. In this study, researchers compared takeover times following a system initiated RtI, workload during the takeover performance, and perceived usefulness and satisfaction with the HMI in a simulator versus a field operational test. Participants' performance on non-critical system initiated RtIs was assessed on a fixed-base simulator or on an instrumented vehicle during drives on public roads and highways (inter-subject design). To enable comparison between groups, drivers from the simulated test scenario were matched with drivers from the on-road test scenario regarding gender, age and driving experience. Results showed significant differences in response time for both manual-automated and automated-manual transitions, with participants on the simulator averaging a longer response time. Correlation analysis of the sorted response time data showed a positive significant correlation between behaviour during simulated and during "real world" scenarios, suggesting similar participant distributions for both types of transitions. There were no differences between groups on workload and on perceived usefulness and satisfaction assessments.

Findings like those above can call attention for the need to distinguish between absolute and relative validity (see also Wang et al., 2010)[131] when presenting the results coming from simulator research. While absolute validity is obtained when the absolute size of the effect measured in the simulator is the same as the absolute size of the effect measured in the roadway study, relative validity is obtained when the results on the simulator describe well the relative size or direction of an effect measured in the real-world scenario[132]. Transfer functions for results obtained in different test environments may allow some research results to be generalised (see Purucker et al., 2018[134] and Andersen and Sauer, 2007[135]). Researchers can always consider the difference between absolute and relative validity when choosing the test environment for assessing behaviour in safety-critical systems or situations.

11.4 Considerations for test environment selections

Factors such as the nature of the research questions, task conditions and dependent measures can be taken into account when selecting the test environments for evaluating human performance and state in the context of automated driving.

While simulators can be useful tools for the study of driver behaviour since they are sensitive to population variability (e.g. age or driving experience related effects), cognitive factors (e.g. measurement of workload, attention and fatigue) and the effect of environmental factors on performance, results often offer only relative validity which is less than ideal for system validation purposes. If the goal is something other than system validation (e.g. traffic regulation), other performance parameters may be more relevant (such as standard deviation of velocity or SDLP upon a RtI). To answer some types of research questions, data from roadway studies will be a better source of information for evaluating human performance.

Task condition, or test scenario, can be the second factor to take into consideration. As discussed in Clause 9, the **urgency** of the RtI, the **criticality** of the driving situation and the **response complexity** required from the driver are all factors influencing the design of the test scenario. Any combination of a high, medium and low classification for each one of these factors is possible in the scope of a test scenario and this will affect the definition of the test environment. Public roadway studies could not be used when there is likelihood of the occurrence of a safety-critical event. Thus, high levels of urgency and criticality that require complex responses from the drivers can be tested in more controlled environments.

One technique that can be used when attempting to maintain realism and still safely evaluate human performance during critical/urgent scenarios is to leverage what is often referred to as Wizard-of-Oz experimentation[136]. For instance, when evaluating human performance in the context of driving automation systems, it can be useful for experimenters to create the appearance that the driving automation system is indeed controlling the vehicle as intended, when in actuality it is being manually controlled by one or more experimenters. This allows researchers to investigate human interaction with driving automation system features that are under development and not likely to perform ideally (e.g. vary in performance and reliability). This has significant benefits both in terms of investigating future human interaction with driving automation systems and for maintaining a safe experimental environment.

Finally, the definition of the dependent variables of an automated driving study could be the last filter in the test environment selection. Mullen et al. (2011)[137], conducted a literature review that presents a comparison between variables measured in simulator and on-road test environments. While driver performance variables mostly show relative validity from simulators, other variables, such as self-reported measures, physiological measures and traffic regulation compliance behaviour[138], appear to have higher levels of behavioural validity.

Annex A
(informative)

Human machine interfaces/interactions for automated vehicles

A.1 Taxonomy of HMIs related to automated vehicles

There are many types of HMIs in modern vehicles, including those for personal devices that drivers and passengers bring in. The introduction of automated vehicles also raised discussion for new types of HMIs. This section defines taxonomy of HMIs related to automated vehicles in order to avoid miscommunication about HMIs in discussion.

1) Vehicle HMI (vHMI): vHMIs are displays and control devices embedded in the cockpit and support the primary driving task (e.g. speed, remaining fuel).

2) Automation HMI (aHMI): aHMIs are displays and control devices embedded in the cockpit and associated with managing level 1–3 systems. (Eventually, vHMI and aHMI will be considered as one category, but during this transition period it may be prudent to separate them for the purpose of discussion).

3) Infotainment HMI (iHMI): iHMIs are displays and control devices embedded in the cockpit for accessing and operating infotainment contents.

4) Nomadic HMI (nHMI): nHMIs are associated with portable devices brought into the vehicle by the driver and passenger(s). Some of these devices can be connected to the vehicle to enable the iHMI to access and operate the contents through the device.

5) External HMI (eHMI): eHMIs are mounted on the exterior of the automated vehicle and provide information to other road users about intent or state of the vehicle. Blinkers and brake lights are included in this category, but discussion is ongoing as to whether automated vehicles may need more signals because other road users cannot always expect signals from the driver (e.g. eye contact, body gestures). There may be also displays inside the vehicle (aHMI) that provide information to the driver and passenger about what is signalled to other road users as well as information about what the vehicle senses about the outside scene[139].

6) Dynamic HMI (dHMI): Dynamic behaviours of the vehicle (e.g. speed, deceleration profile) are often used as signals from the vehicle showing the driver's intent. Research shows that some 80 % of cues received by pedestrians come from this medium[140][141][142]. Therefore, there is discussion that such dynamic behaviour may need to be explicitly designed for automated vehicles. This is a form of human machine interface or rather human machine interaction and is defined as dHMI. While eventually dHMI and eHMI will be coupled together, they are separated here for the sake of a clearer distinction.

All six different HMI elements defined here are based, to some extent, on the general principles of human machine interaction related to automated vehicles. Therefore, the taxonomy includes eHMI and dHMI, which are out of scope of this document.

A.2 Human machine interfaces/interactions to support the driver to improve takeover performance

A.2.1 General

There have been a number of exploratory studies searching for measures to support the driver for better take-over performance. The measures include HMIs and also human machine interactions. The majority of such measures are still under investigation to obtain evidence for their effectiveness. This annex collects information about such studies.

A.2.2 Driver monitoring system

The concept of a driver monitoring system (DMS) involves automated driving monitors in real time driver's readiness/availability that are necessary for performing the driver's task (defined for each level) and a successful performance in takeover that may happen in the future. A DMS is considered to be primarily effective for level 2 and level 3 systems. When a DMS detects declines in metrics of readiness/availability below the required level, it reports to the main system. The main system then activates interventions (e.g. visual/auditory alerts, prompts, other stimuli) to retrieve the required level of readiness/availability of the driver or to shift the system to level 0. The main system also may initiate an MRM to bring the vehicle to a controlled stop. Selection of the intervention may be dependent on the level of readiness/availability and also dependent on conditions of the system, traffic and road environment. The hands-off detection and alerts/termination of automated mode are a function of the DMS and interventions that are already deployed widely in the commercialized systems. Some of the commercially available systems are implemented with drowsiness detection and interventions (e.g. General Motors, Lexus).

There are other types of DMSs that are not included in the previous paragraph. There are studies and developments of DMSs that detect declines in the driver's health condition (e.g. heart attack, stroke, epileptic seizure, diabetic glycemia) so that the vehicle is brought to a controlled stop by a system to avoid a crash. However, the driver states and metrics to be monitored can be different from readiness/availability for automated driving. Also, a variety of DMSs have been studied and developed for impairments of manual vehicle drivers, such as drowsiness, eyes-off-road and drunkenness. However, metrics of such states largely include behavioural driving changes (e.g. steering entropy) that cannot be used for readiness/availability for automated driving.

There have been a number of studies investigating the effectiveness of DMSs and metrics of readiness/availability to be monitored by DMSs[143][144][145][95][146][49][50][26][147][148][149][150]. The operational definition of readiness/availability is still under discussion and there are no DMS design principles or standards established yet. However, DMSs may evolve to monitor various aspects of impaired driver states by including more components of readiness/availability, as described in 7.3.

A.2.3 HMIs to maintain driver attentiveness to the road when engaged in an NDRA

Some previous research has been done for level 2 system human machine interfaces /interactions to maintain the visually loaded driver engaged in an NDRA to be attentive to the environment (i.e. appropriate time-share of attentional allocation between the environment and the NDRA). Blanco et al. (2015)[95] tested a prompt that alerts drivers if they look away from the front roadway for certain periods of time and found it to be effective. Llaneras et al. (2017)[149] advanced Blanco et al.'s study and proposed a visual attention management strategy. The strategy integrated off-road glance detection and progressive staged alerts. When drivers' off-road glance time became longer, the intensity of the alert became stronger over three stages, with different combinations of visual, auditory and haptic stimuli. The strategy also included increments of system demand in the three stages for the driver to stop the alert (i.e. stage 1: move glance back to the front, stage 2: grab the steering wheel and reengage in manual driving and stage 3: stop the vehicle and recycle ignition). The strategy was found to be effective at improving drivers' compliance with the alerts and maintaining attention to the road. The strategy also increased drivers' detection of unexpected events, such as DDT performance-relevant failures. Driver education about presence and meaning of the alerts was found to improve the effectiveness further.

A.2.4 HMIs to mitigate development of drowsiness

Human machine interfaces/interactions to mitigate development of driver's drowsiness and support driver maintenance of appropriate level of readiness/availability have been studied by researchers. Some of the studies used NDRAs to raise arousal of the driver who was cognitively underloaded in automated driving. NDRAs used in the studies included cell phone conversation[44], talking with a passenger[151] and a quiz task[45]. All of the studies found these NDRAs to be effective at mitigating drowsiness. However, implementation of NDRAs into the system may be an issue. There are also some studies that used sensory stimuli. Hirano et al. (2018)[151] tested auditory stimuli in the context of automated driving but did not find it to be effective. Gaspar et al. (2017)[152] reported auditory-visual alerts were effective to mitigate development of drowsiness in manual driving. They also found multi-stage alerts were more effective than a single-stage alert. Kato et al. (2011)[153] found that the presence of intermittent odours mitigated changes in EEG signals and the degradation of response performance over time in laboratory simulated long-term manual driving. The studies of Gaspar et al.[152] and Kato et al.[153] were based on manual driving and applicability of their findings to automated driving is not known. Some researchers investigated effectiveness of scheduled and repeated driver engagement/disengagement in manual driving (i.e. disengagement in manual driving means automated driving). Blommer et al. (2015)[154] found that the scheduled driver engagement strategy appeared to be effective when the driver was visually loaded. On the other hand, Wu et al. (2019)[21] found no effects of the scheduled driver engagement strategy for younger drivers and negative effects for older drivers, possibly due to their impaired task-switching ability.

A.2.5 HMIs for supporting driver's situation awareness

Drivers' situation awareness is important to ensure that the driver takes an appropriate driving tactic to cope with a critical event after an RtI[81]. To mitigate driver's reduced sampling of environmental elements due to a visual load caused by an NDRA in a level 3 system, the head-up display and steering switches for the NDRA interaction were evaluated and recommended by Wulf et al. (2015)[155]. To enhance the situation awareness, an augmented reality technique was applied to the head-up display by Langlois and Soualmi (2016)[156] to show positions and movements of neighbouring vehicles and the road to follow. The evaluation results showed its effectiveness on drivers' takeover performance. A haptic seat was developed by Telpaz et al. (2015)[157] to enhance situation awareness by showing spatial information of approaching vehicles through directional vibration on the seatback. The seat was evaluated and showed its effectiveness on drivers' takeover performance.

An HMI displaying the current location of the vehicle on the route to the destination (i.e. navigation) may enhance the strategic level of drivers' situation awareness. Such an HMI may support drivers performing an NDRA with a level 3 system to focus on manoeuvring in transition by allocating minimum cognitive resources to the route search[158][82].

A.2.6 HMIs for operating state/mode

An HMI showing the system's operating state/mode is important to prevent drivers' misunderstanding or confusion about the dynamic state of the system. Misunderstanding and confusion can be critical when the driver believes the operating mode is at a higher level than it actually is. The HMI becomes especially important when multiple levels of automated driving exist within the system. An HMI with a combination of colouring and semantic text information was found to be effective by Foster et al. (2016)[159]. An ambient HMI (i.e. LED array surrounding the driver) was tested and found to be potentially effective[160]. Beattie et al. (2014)[161] tested spatialized auditory feedback for intended actions by the system. The results showed the feedback's effectiveness in the driver's sense of being in-control and having a reduced workload. Yang et al. (2018b) [169] investigated an HMI consisting of a LED array located closed to the windscreen. The LED array conveyed information about the state and the intention of the system, potential hazards and an RtI with different colors, blinking frequencies, lighting positions and animations. The HMI was found to improve the takeover quality.

A.2.7 HMIs for detected objects and events

The driver is expected to detect objects or events which are not detectable by the system and is expected to respond to them for level 2 systems (i.e. OEDR and human-initiated transition). An HMI showing objects and events being detected by the system may be important to raise drivers' awareness of the system state and improve takeover performance[162]. Kitazaki et al. (2019)[163] found that visual information about detected objects did not improve drivers' performance of a human-initiated takeover for an object undetected by the system. However, visual information about a combination of detected objects and a system's planned action in response to the objects (e.g. braking, changing lane) significantly improved performance of the human-initiated takeover. Beattie et al. (2014)[161] tested spatialized auditory feedback for intended actions by the system. The results showed the feedback's effectiveness in drivers' sense of being in-control and having a reduced workload.

Higher HMI expectations are that HMI information of detected objects and planned actions by the system may support driver calibration of trust and the formation of an appropriate mental model of system functionality and limitations (i.e. what can be detected and what cannot be) over long-term use of the system[83]. The appropriate mental model is expected to induce a quicker response to undetected objects through the rule-based response rather than the knowledge-based response[164][165].

A.2.8 HMIs for system certainty/uncertainty

An HMI showing the dynamic state of the system certainty/uncertainty for detected environment (i.e. confidence for detected objects) has been reported to positively influence drivers' takeover performance by some researchers. When the driver receives information that the system is detecting an object but certainty is lower than the ideal performance, driver vigilance to the environment may be raised before the system finally loses the object(s), resulting in better takeover performance in both system-initiated and human-initiated transitions[166][167][168].

Bibliography

[1] SAE J3016:2016, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicle*

[2] ISO/TR 21959-1, *Road vehicles — Human performance and state in the context of automated driving — Part 1: Common underlying concepts*

[3] Itoh M., Zhou H., Kitazaki S., What May Happen or What You Should Do? Effects of Knowledge Representation Regarding Necessity of Intervention on Driver Performance under Level 2 Automated Driving. In ICPS'18: Proceedings of the 2018 IEEE Industrial Cyber-Physical Systems (pp. 621–626). Saint Petersburg, Russia, (2018)

[4] Liu T., Zhou H., Itoh M., Kitazaki S., The impact of Explanation on Possibility of Hazard Detection Failure on Driver Intervention under Partial Driving Automation. In IV2018: Proceedings of the 29th IEEE intelligent Vehicles Symposium. Chang Shu, China, (2018)

[5] Zhou H., Itoh M., Kitazaki S., How Does Knowledge about System Limitations Contribute to Interventions into Partial Automation Among Elderly Drivers? To appear in the 2018 IEEE international Conference on Systems, Man, and Cybernetics (SMC2018). October, Miyazaki, Japan, (2018)

[6] Victor T. W., Tivesten E., Gustavsson P., Johansson J., Sangberg F., Ljung Aust M., Automation expectation mismatch: Incorrect prediction despite eyes on threat and hands on wheel. *Human Factors*, **60**(8), 1095–1116, (2018)

[7] Boelhouwer A., van den Beukel A. P., van der Voort M. C., Martens M. H., Should I take over? Does system knowledge help drivers in making take-over decisions while driving a partially automated car? *Transportation Research Part F: Traffic Psychology and Behaviour*, **60**, 669–684. https://doi.org/10.1016/j.trf.2018.11.016, (2019)

[8] Itoh M., Toward overtrust-free advanced driver assistance systems. 14, 51-60. DOI: 10.1007/s10111-011-0195-2, (2012)

[9] Wright T. J., Samuel S., Borowsky A., Zilberstein S., Fisher D. L., Experienced drivers are quicker to achieve situation awareness than inexperienced drivers in situations of transfer of control within a Level 3 autonomous environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. **60**, No. 1, pp. 270–273), (2016, September)

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

[11] Payre W., Cestac J., Dang N. T., Vienne F., Delhomme P., Impact of training and in-vehicle task performance on manual control recovery in an automated car. *Transportation research part F: Traffic psychology and behaviour*, **46**, 216–227, (2017)

[12] Maltz M., Shinar D., Imperfect in-vehicle collision avoidance warning systems can aid distracted drivers. *Transportation research part F: Traffic psychology and behaviour*, **10**(4), 345–357, (2007)

[13] Körber M., Baseler E., Bengler K., Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, **66**, 18–31, (2018)

[14] Lee J., Moray N., Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, **35**, 1243–1270, (1992)

[15] Owsley C., McGwin G.Jr, Vision impairment and driving. *Survey of ophthalmology*, **43**(6), 535-550, (1999)

[16] Young K. L., Koppel S., Charlton J. L., Toward best practice in Human Machine Interface design for older drivers: A review of current design guidelines. *Accident Analysis & Prevention*, **106**, 460–467, (2017)

[17] Ponds R. W., Brouwer W. H., Van Wolffelaar P. C., Age differences in divided attention in a simulated driving task. *Journal of Gerontology*, **43**(6), 151–156, (1988)

[18] Clark H., Feng J., Age differences in the takeover of vehicle control and engagement in non-driving-related activities in simulated driving with conditional automation. *Accident Analysis & Prevention*, **106**, 468–479, (2017)

[19] Molnar L. J, Pradhan A. K., Eby D. W., Ryan L. H., St. Louis R.M., Zakrajsek J., Ross B., Lin B. T., Liang C., Zalewski B., Zhang L., Age-related differences in driver behavior associated with automated vehicles and the transfer of control between automated and manual control: A simulator evaluation. University of Michigan, Transportation Research Institute, Institute for Social Research, UMTRI 2017-4, May 2017

[20] Li S., Blythe P., Guo W., Namdeo A., Investigation of older driver's takeover performance in highly automated vehicles in adverse weather conditions. *IET Intelligent Transport Systems*, **12**(9), 1157–1165, (2018)

[21] Wu Y., Kihara K., Takeda Y., Sato T., Akamatsu M., Kitazaki S., Effects of scheduled manual driving on drowsiness and response to take-over request: A simulator study towards understanding drivers in automated driving. *Accident Analysis & Prevention*, Vol. **124**, March 2019, pp. 202–209, (2019)

[22] Körber M., Gold C., Lechner D., Bengler K., The influence of age on the take-over of vehicle control in highly automated driving. *Transportation research part F: Traffic psychology and behaviour*, **39**, 19–32, (2016)

[23] Zhang B., de Winter J., Varotto S., Happee R., Martens M., Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, **64**, 285–307, (2019)

[24] Braunagel C., Kasneci E., Stolzmann W., Rosenstiel W., Driver-activity recognition in the context of conditionally autonomous driving. In 2015 IEEE 18th International Conference on Intelligent Transportation Systems (pp. 1652–1657). IEEE, (2015)

[25] Zeeb K., Buchner A., Schrauf M., What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident Analysis & Prevention*, **78**, 212–221, (2015)

[26] Cabrall C., Janssen N., Goncalves J., Morando A., Sassman M., de Winter J., Eye-based driver state monitor of distraction, drowsiness, and cognitive load for transitions of control in automated driving. In 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 001981-001982). IEEE, (2016)

[27] Herzberger N.D., Voss G.M., Schwalm M., Identification of criteria for drivers' state detection. In International Conference on Applied Human Factors and Ergonomics (pp. 798–806). Springer, Cham, (2017)

[28] Marberger C., Mielenz H., Naujoks F., Radlmayr J., Bengler K., Wandtner B. Understanding and Applying the Concept of "Driver Availability" in Automated Driving. In: Stanton N., (eds) *Advances in Human Aspects of Transportation. AHFE 2017. Advances in Intelligent Systems and Computing, 597, 595–605.* Springer, (2018)

[29] Yang Y., Gerlicher M., Bengler K., How does relaxing posture influence take-over performance in an automated vehicle? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. **62**, No. 1, pp. 696–700). Sage CA: Los Angeles, CA: SAGE Publications, (2018a)

