Results of the on-road studies

Deliverable D3.4 – WP4 – PU

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

Mediator

Results of the on-road studies

Work package 3, Deliverable D3.4

Please refer to this report as follows:
Fiorentino, A., Ahlström C., Anund A., Andersson, S., Bakker B., Borgarello, L., Busiello, M., Cleij, D., Farah H., Jansen, R., Knauss A., Larsson, K., Mazza, M., Ponnambalam, C. T., Rauh N., Thalya P., Toffetti A., de Winkel, K. N., Yuan Y. (2023). Results of the on-road studies, Deliverable D3.4 of the H2020 project MEDIATOR.

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

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

Lead contractor for this deliverable:
Anita Fiorentino – FCA Italy

Report Author(s):	
	Fiorentino, A. (FCA Italy), Italy Toffetti, A., Borgarello, L., Mazza, M., Busiello, M. (CRF), Italy Ahlström, C., Larsson, K., Anund A. (VTI National Road and Transport Research Institute), Sweden Thalya P., Andersson, S. (Veoneer Sweden AB), Sweden Bakker B. (Cygnify BV), The Netherlands Knauss A. (Zenseact AB), Sweden Farah H., Yuan Y. (TU Delft), The Netherlands Rauh N. (TU Chemnitz), Germany Cleij, D., Jansen, R., de Winkel, K. N., Ponnambalam, C. T., (SWOV), The Netherlands

Revision history

Date	Version	Reviewer	Description
17/03/2023	Preliminary draft	Anita Fiorentino – FCA, Italy	Task internal review
04/04/2023	Draft for review	Rolf Zon – NLR, The Netherlands	Internal QA review
30/04/2023	Final draft	Ernst Verschragen (QA officer) – SWOV Institute for Road Safety Research, The Netherlands	Final editing and checks
30/04/2023	Final deliverable	Michiel Christoph (Technical project coordinator) – SWOV Institute for Road Safety Research, The Netherlands	Approved

Legal Disclaimer

Table of contents

List of Abbreviations and acronyms

ADAS	Advanced Driver Assistance System
ADS	Automated Driving System
CAV	Connected Automated Vehicles
CM	Continuous Mediation (automation level)
DMS	Driving Monitoring System
DP	Driving Piloted
Dx.x	Mediator Deliverable number x.x
EPS	Electric Power Steering
GDPR	General Data Protection Regulation
HF	Human Factors
HMI	Human Machine Interface
KPI	Key Performance Indicator
KSS	Karolinska Sleepiness Scale
LED	Light Emitting Diode
NDRT	Non-Driving Related Tasks
ODD	Operational Design Domain
RQ	Research Question
SB	StandBy (automation level)
SUS	System Usability Scale
TTAU	Time To Automation Unfitness
TTAF	Time To Automation Fitness
TI	Technical Integration
ToC	Transition of Control
TtS	Time To Sleep (automation level)
UC	Use Case
WoOz	Wizard of Oz
SWOV	Institute for road safety research
ANOVA	Analysis of Variance
VS	Versus

List of tables

List of figures

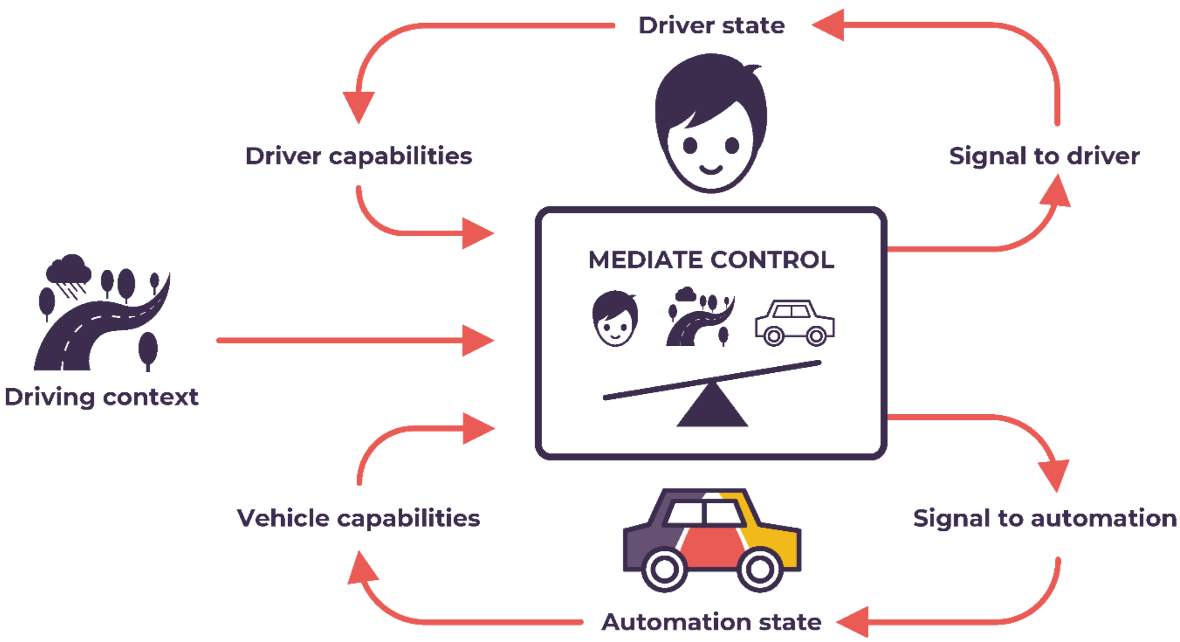
About MEDIATOR

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

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

1.1. Vision

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



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

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

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

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

https://mediatorproject.eu/

Executive summary

Within MEDIATOR the Testing & Evaluation activities aim to evaluate the functionality of the Mediator system under simulated and real road conditions with real drivers in different parts of Europe. This includes assessing the system's performance, reliability, functionality, effects on driving safety as well as the acceptance, trustworthiness, perceived safety and user-friendliness for different groups of users. This deliverable covers the on-road evaluations of the vehicle-integrated Mediator system, which consists of three studies.

This public Deliverable 3.4 describes the on-road evaluations of the vehicle-integrated Mediator system. Three real-life on-road studies were conducted to test the overall performance of the Mediator system and its effects on safety-relevant behaviours, driver reactions and driver opinions.

First Study in Italy

The first study used the basic Automated Driving System (ADS)-level prototype and focused on the functionality of the HMI solutions in different driving conditions as well as on the acceptance/trust of the users. To reach this purpose, a Wizard of Oz (WoOz) vehicle prototype (starting from a right-hand drive Jeep Renegade fully equipped, with ADS level 1) was created, integrating a non-functional double set of pedals and steering wheel on the left side and a prototype shifter to allow to simulate the driving mode change (from manual to automated and from automated to manual). This WoOz prototype was always driven by a professional driver, who was on the right seat and drove using the standard Renegade primary controls, but the Wizard of Oz methodology guaranteed participants could experience an "automated" vehicle HMI behaviour, without being in a real automated vehicle.

16 naïve participants took part to an on-road test on a 46 km scenario, during which they tested the MEDIATOR HMI solutions (visual, vocal, acoustic, luminous, haptic and cushion inflation) integrated in the WoOz vehicle cabin (e.g., centre dashboard display, participant frontal display, shifter, steering wheel, seat belt) and designed to cope with the needs of MEDIATOR use cases (e.g., handover, takeover).

This on-road user testing allowed to evaluate the MEDIATOR HMI solutions designed in the MEDIATOR project and to understand their advantages and disadvantages. Both the acceptability (before the trial) and the acceptance of automated vehicles and MEDIATOR HMI were positive. The HMI usability and the users' trust in these HMIs were positively evaluated too. Some MEDIATOR HMIs weaknesses emerged too, and they were used to understand what had to be fixed in the HMI solutions. This first study in Italy was useful to select the most appropriate HMIs to be tested during the on-road study in Sweden.

Second Study in Sweden

The second study was also conducted in the basic ADS-level prototype and focused on the functionality, validity, and reliability of the Technical Integration vehicle prototype under different degraded driver performance conditions, including conditions of degraded automation.

An on-road study with 50 naïve participants, using the same Wizard of Oz setup as in the first study in Italy, compared the full MEDIATOR HMI with a baseline HMI that was based on existing HMI designs (i.e., mainly using simple icons and sounds for interaction). Several novel features in the Mediator system showed great potential. The participants appreciated the time budget which provided information on when a change in automation level (i.e., responsibility) would take place. The combination of warnings and continuous mode and time budget information decreased the total duration in which participants were distracted and the maximum uninterrupted period of distraction. Especially the latter can have a significant effect on road safety, as long continuous periods of being distracted severely reduces situation awareness. Also, the elaborate Mediator takeover ritual, which included reasons for why a takeover was happening, was more appreciated and understood than the takeover ritual of the Baseline system. For fatigue, there were no differences in the development of fatigue between the MEDIATOR and Baseline HMIs. However, the results indicate that corrective alert-messages used in the transfer of control ritual somewhat reduced task-related fatigue, but not sleep-related fatigue. Finally, the Mediator system included an active proposal to increase the level of automation which was much appreciated by participants. They also strongly agreed that they would increase automation use if such a feature was available in their car. If automation is indeed safer than manual driving, this feature could therefore potentially improve road safety. Overall, the Mediator system was preferred over the Baseline system, but the results indicate a need for increased cohesion in the HMI and improved clarity of icons and warnings.

Third Study in Sweden

In the third study, the Mediator system’s performance was examined using the higher ADS-level prototype. Reliability and validity of the automation-status detection in relation to driving context, decision logic and timing of the Mediator system’s actions and messages were evaluated. Seven professional drivers experienced different configurations of the Mediator system. Over several weeks, they drove ten times on a specific route of one hour length covering use cases and scenarios the study focused on. Both quantitative data (from the test vehicle) and qualitative data (from structured interviews) were collected and analysed. Results indicate that drivers evaluated the Mediator system and it’s HMI quite positively. Some drivers argued that a system like Mediator is more useful for higher automation levels (level 3 and above). Overall reliability was good, although the distraction warning implemented in the vehicle was too sensitive. The simplified HMI implemented in the prototype did not affect drivers’ gaze behaviour; in fact, they looked at the Mediator display less overtime. However, due to limited data collected during the study this is an exploratory study and has limited external validity.

1. Introduction and general considerations

The main aim of the Testing & Evaluation” activities is to evaluate the functionality of the Mediator system under simulated and real road conditions with real drivers in different parts of Europe. This includes assessing the system’s performance, reliability, functionality, effects on driving safety as well as the acceptance, trustworthiness, perceived safety and user-friendliness for different groups of users. A series of studies will be carried out to evaluate the Mediator system and its key components. Based on the use cases (UC), each of these evaluation studies will focus on specific UC and connected relevant variables for driving context, vehicle automation status and driver state. In addition, the studies will include a sufficiently large number of participants to allow for assessing differences in sociodemographic factors (such as age, gender, driving experience…).

This chapter describes the main choices that were made to carry out the different studies in terms of overall evaluation strategy, evaluation areas, MEDIATOR Use Cases, and different in-vehicle prototypes used for the road tests,

1.1. Overall evaluation strategy and principles

The overall testing and evaluation strategy in MEDIATOR is aligned with state-of-the-art evaluation procedures developed and applied in other EU projects such as L3Pilot (Metz et al., 2019), ADAS&ME (Pereira Cocron et al., 2019) and AdaptIVe (Rodarius et al., 2015). The evaluation process mainly follows the FESTA V-process methodology (FOT-Net and CARTRE, 2018), developed for planning and executing Field Operational Trials. Next to an overarching consideration of ethical and legal issues, the FESTA-V process consists of the three phases 1) preparation, 2) data acquisition and 3) analysis. The preparation phase includes the definition of functions and UC, development of research questions and hypotheses, definition of performance indicators and measures, preparation of sensors and the study design. The present deliverable aims and summarizes this preparation work for each MEDIATOR evaluation study. The second step of data acquisition involves pre-tests next to the actual collection of data for the analyses. The analysis part mainly involves storage of data, transformation of measures into performance indicators, testing hypotheses and answering research questions as well as the impact analysis.

General **ethical, legal and safety procedures** for all evaluation studies were already defined within MEDIATOR project and the procedures are specified in each study plan of this deliverable. In general, all studies involving human participants will respect all ethical, legal and safety procedures.

The MEDIATOR testing and evaluation strategy consequently considers the specific features of the different evaluation methods (computer simulation study, driving simulator studies, on-road studies). This deliverable covers the on-road evaluations whereas the results from the computer simulation are reported in D3.2 (Athmer et al., 2022) and the results from the driving simulator experiments are reported in D3.3 (Borowsky et al., 2023).

A **complementary approach** is followed, taking advantage of the capabilities of the two in-vehicles prototypes that have been developed within MEDIATOR. Different UCs are thus addressed in the different prototypes. Each on-road evaluation study is planned by the respective study team, considering study-specific UCs, research questions and components of the Mediator system. The complementary approach is also reflected in the main chapter structure of this deliverable, presenting each study plan by the respective study team.

Comparability of results: In parallel to the complementary approach, maximum comparability of study results shall be achieved. One pillar for ensuring comparability is the UCs defined in Cleij et al. (2020), which ensure a common set of driving contexts, driver states, automation states and Mediator functionalities across the studies. A second important aspect of comparability are the Key Performance Indicators (KPIs) defined in Cleij et al. (2020) as well as the common design concept for the HMI. Finally, common questionnaires and interview guidelines as dependent variables of the user evaluations allow for direct comparison of results across the studies.

1.2. Evaluation areas

The MEDIATOR evaluation will focus on the two main evaluation areas 1) technical evaluation and 2) user evaluation. This approach is in line with previous EU projects and the areas will cover different aspects:

Technical evaluation covers research on the technical performance, reliability, functionality, and effects on driving safety of the Mediator system and system components. The aim of this analysis is to understand the technological readiness of the systems, identify areas for technical improvements and provide input data for further analysis in the impact evaluation. Potential KPIs for technical evaluations are listed in D1.4 (Cleij et al., 2020) and in each study chapter. For example, main technical KPIs include true/false positive/negatives of driver state detection components, prediction performance and horizons of the automation module, number, and duration of safety critical events, etc.

User evaluation aims at assessing acceptance, trustworthiness, perceived safety and user-friendliness of the Mediator system and system components for different users. User groups consist of naïve and professional users, different age groups as well as gender-balanced samples. For assessing impacts of the Mediator system, changes in user behaviour while interacting with the system are essential. In addition, user need to accept the function and show willingness to use it, to allow any potential system impacts to be realized. Therefore, a good understanding of the users, their interactions with the system over time and their opinions about it plays a central role in the user evaluation process. Potential KPIs for user evaluations are listed in D1.4 (Cleij et al., 2020) and each study chapter. Main user KPIs comprise for example acceptance, trust, usability of the Mediator system / components as well as general attitudes towards vehicle automation.

1.3. Levels of automation in MEDIATOR

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

MEDIATOR addresses automation on SAE levels 0 – 4, using the terminology defined in Table 1.1. A key point within MEDIATOR has been to adopt a user perspective on automation. Where SAE automation levels align with technical possibilities of automation, MEDIATOR automation levels are based on the driver's responsibilities and affordances. To illustrate, whereas SAE level 4 represents a level of automation that allows a driver to be out of the loop and that also ensures safe handling of situations where the automation cannot adequately perform the driving task, it does not consider how long one can be out of the loop. In MEDIATOR, the Time-to-Sleep mode is defined from a user perspective: it considers whether the driver can stay out of the loop for a short while or for a long time.

Table 1.1: Automation levels addressed in MEDIATOR (OEDR: Object and Event Detection and Response).

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

1.4. Summary of use cases

In Cleij et al. (2020) a total of ten UCs were developed to define the scope of the MEDIATOR project. As the on-road evaluations refer to these UC, a summary is given below:

1. Mediator system initiates takeover (human to automation): Degraded human fitness, caused by either drowsiness (a) or distraction (b), is detected by the Mediator system. The system reacts by initiating a takeover to automation.

2. Driver takes back control: The driver uses the HMI to indicate a desire to take the control back. The Mediator system reacts by confirming that the driver is fit enough to drive and guides the takeover.
3. Comfort takeover (human to automation): Either the driver (a) or the Mediator system (b) initiates a takeover from human to automation.
   a) The driver indicates via the HMI that he/she is not motivated to drive. The Mediator system reacts by confirming the automation fitness and guiding the takeover.
   b) The Mediator system detects an event, such as receiving a text message or an upcoming traffic jam, from which it concludes that the driver comfort could be improved. The system reacts by suggesting a takeover to automation.
4. Corrective Action: While driving in StandBy (SB) the human driver becomes drowsy, the Mediator system reacts by initiating an action to improve the driver fitness and monitors the effect.
5. Mediator initiated takeover (automation to human): A planned (a) or an unplanned (b) takeover from automation to human is initiated by the Mediator system.
   a) The automation indicates that the current route leads to automation unfitness as it will leave its Operational Design Domain (ODD). The Mediator system reacts by preparing the driver for and guiding the driver through a non-urgent takeover.
   b) The automation indicates that its fitness is rapidly degrading and can soon no longer perform the driving task. The Mediator system reacts by informing the human driver and guiding the urgent takeover.
6. Comfort switches on: Either the driver (a) or the Mediator system (b) switches on driving in Continuous Mediation (CM).
   a) The driver indicates via the HMI that he/she is not motivated to drive. The Mediator system reacts by confirming the automation fitness and switches on CM.
   b) The Mediator system detects sufficient fitness for driving in CM from which it concludes that the driver comfort could be improved and reacts by suggesting switching on CM.
7. Preventive Action: While driving in CM, the driver is supported by the Mediator system in performing the monitoring task. The system does this by trying to prevent underload and keeping the driver in the loop
8. Corrective Action: While driving in CM, degraded driver fitness is detected by the Mediator system. The system reacts by initiating a corrective action to improve driver fitness.
9. CM shuts off instantly: While driving in CM, the automation fitness degrades, and automation can no longer perform its driving task. The Mediator system reacts by communicating to the driver that CM is switching off.
10. Smooth transition from Time to Sleep (TtS) to SB: while driving in TtS the driver is fully disengaged from the driving task when the automation indicates that the current route will leave the ODD. The Mediator system detects sufficient automation fitness for driving in SB and reacts by informing the driver that SB will be switched on and subsequently monitors the required driver fitness.

The UC provide a description of the functional range and the desired behaviour of the Mediator system, whereas “scenarios” provide the concrete implementation of the UC in each evaluation study (e.g., road type, parameters, manoeuvres… etc.). Each on-road study covers a subset of these UCs and addresses them via a set of research questions (sections 2.1, 3.1, 4.2).

1.5. The different in-vehicle platforms used for the road tests

The on-road evaluations were carried out in the two in-vehicle prototypes which were developed in MEDIATOR.

The first vehicle prototype (Figure 1.1) focuses on the human factors side and realises and explores the most sophisticated driver state estimation technology and the most sophisticated HMI. This vehicle has a basic level of ADS sophistication and relies on a Wizard of Oz-like set-up to simulate vehicle automation. The focus is on presenting to the participants in the driver seat of this prototype vehicle the full concept HMI experience and recording their behaviour, response and experiences while driving on a real-world realistic route, with simulated high levels of vehicle automation. The simulated high level of vehicle automation is realised by having a Wizard driver sitting in the actual driving position of a UK-oriented right hand steering position vehicle, who was instructed to avoid interacting with the participant, with whom only the examiner/experimenter could interact.

In the first Mediator in-vehicle prototype the focus is on the HMI. For driver state, the full sensor suite of cameras and physiological sensors are used, and data from that are recorded for post-hoc analysis; more details are given in Fiorentino et al. (2022). Next to recording, some real-time driver state functionality is also realised, by using a commercial eye tracker system for distraction detection. Decision logic is largely simulated, focusing on going through pre-programmed scenarios; but there is some real-time functionality for driver corrective actions based on real-time distraction and fatigue detection.

Figure 1.1 The Mediator Human Factors prototype vehicle.

The second Mediator in-vehicle prototype (Figure 1.2) has real (Level 2, “Continuous Mediation”, “Pilot Assist”) automation. It focuses on the automation and automation state side of the Mediator system in conjunction with the driver state side and driving context side, allowing demonstration and evaluation of the real complete MEDIATOR Decision Logic component. However, this prototype simplifies heavily on the side of the HMI, using only a limited MEDIATOR HMI, allowing for more technical Mediator logic evaluation and not the full envisioned user experience. Compared to the first Mediator in-vehicle prototype, this vehicle has a basic HMI, while Driver State, Automation State, Driving Context, and Decision Logic are all real and real-time MEDIATOR software main components, as envisioned in the concept, providing inputs on the fly to the HMI, thus demonstrating the MEDIATOR concept(s) on real roads.

Figure 1.2 The Mediator TI in-vehicle prototype.

In Fiorentino et al. (2022), the two in-vehicle prototypes are described in detail; as the on-road evaluations refer to these in-vehicle prototypes a summary, in terms of type and level of sophistication of features related to each vehicle prototype, is given in the Table 1.2.

Table 1.2 MEDIATOR main software components - Prototype Platform “Match” table (Fiorentino et al. 2022).

MEDIATOR main SW components	HF in-vehicle prototype	TI in-vehicle prototype
Driver state – distraction & fatigue	Real-time distraction & fatigue Recording for post-hoc full system;	Real-time, but only camera-based system (no physiological sensing)
Driver state – comfort	Recording, interviews	Simple, table/situation-based, use case focus
Driving context	Simplified: Some real & real-time functionality fed by vehicle/route	Real and real-time, fed by vehicle/automation systems, map, other data.
Automation State	Limited simulation, using pre-programmed route / triggers based on the start and end of each use case included in the automated sequence of the predefined trip in Italy and Sweden	Real and real-time, use case focus, development focus
Decision Logic	High-level decisions simulated/triggered; with real and real-time low-level decision logic for specific use cases	Real and real-time; but focused on simple, robust version
HMI	Real and real-time, fullest version of HMI MEDIATOR development The HMI integrates several commercial and prototypes HMI components in line with the HMI holistic design concepts	Basic, in terms of limited multi-modality (visual and auditory only, one screen instead of two) but in line with MEDIATOR HMI holistic design concepts

2. HMI evaluation in different driving conditions

This chapter describes the Italy on-road study aimed to HMI evaluation in different driving condition with the Human Factors (HF) in-vehicle prototype.

The automated vehicles HMI is a crucial aspect to allow a positive users' acceptance of this important innovation. To design automated vehicles with a positive users' acceptance is indispensable that their HMI has a high usability. To achieve this goal, it is of paramount importance to adopt the User-Centred Design process (ISO 9241-210, 2010) during the HMI design. A fundamental phase of this process is the evaluation involving real users, but when the HMI is developed in parallel with the technical development, involve users in the experimentation can be very difficult, then the Wizard of Oz (WoOz) research approach was developed. It was ideated in the Human Computer Interaction domain, and it consists of making users interact with a system that they believe to be real, but which is controlled, completely or partially, by a hidden human being, the wizard (Kelley, 1984), Then, the Wizard of Oz paradigm enables this parallel HMI research.

According to a recent paper (Bengler et al., 2020), within the automotive research community, WoOz vehicles are started to be used as a method for analysing the effects of "intelligent" probabilistic systems, as automated vehicles are, even if not yet fully developed yet.

The Mediator in-vehicle prototype used for the on-road study in Italy is the HF in-vehicle protype; the in-vehicle prototype has basic level of ADS sophistication and relies on a Wizard of Oz-like set-up to simulate vehicle automation. This solution, even if the participant does not have any real control of the vehicle, has the advantage that it is possible to run experiments on automated driving HMI, simulating higher levels of automation that aren't available yet. For the technical description of the HF in-vehicle prototype, please, refer to Fiorentino et al. (2022).

2.1. Objectives, Research questions and covered use cases

The purpose of the Italian study was to evaluate the usability, the acceptance, and the perceived trust of the MEDIATOR HMI solutions (for level CM and SB) designed during the previous tasks of the project, in an ecological way on a public road with the involvement of a sample of naïve users. Moreover, the study had to select the most appropriate HMIs and following the Human-Centred Design process, modify them if needed, so to test them again in the on-road study in Sweden.

The Research Questions (RQ) of the study were the following:

1. Are there any differences in the users' acceptance of Automated Vehicle and MEDIATOR HMI before and after the trial?
2. Do people consider MEDIATOR HMI usable, and do they trust it?
3. With the MEDIATOR HMI, is the handover request (from human to automation), by driver's input, usable (effective, efficient, and satisfying) and acceptable?

4. With the MEDIATOR HMI, is the handover request (from human to automation), caused by distraction, usable and acceptable?
5. With the MEDIATOR HMI, is the handover request (from human to automation), caused by a detected event, usable and acceptable?
6. With the MEDIATOR HMI, is the handover request (from human to automation), caused by drowsiness, usable and acceptable?
7. With the MEDIATOR HMI, is the takeover request (from automation to human), caused by driver's desire and input, usable and acceptable?
8. With the MEDIATOR HMI, is the planned takeover request (from automation to human) usable and acceptable?
9. With the MEDIATOR HMI, is the improvement of driver fitness during an automated driving mode, caused by drowsiness, usable and acceptable?

During the Italian trials the following seven use cases were considered and they were repeated at least twice (when feasible) during each trial[1]:

- Use case 1a: Mediator system initiates the handover (from manual to automated driving mode) caused by simulated drowsiness
- Use case 1b: Mediator system initiates the handover (from manual to automated driving mode) caused by distraction
- Use case 2: The driver desires to take back the control, and indicates it using the Mediator shifter
- Use case 3a: The driver is no more motivated to drive and indicates it via the HMI, using the Mediator shifter
- Use case 3b: The Mediator system detects a simulated event (text message), and it suggests the handover (from manual to automated driving mode)
- Use case 4: The Mediator system reacts to the simulated status of driver drowsiness by initiating an action to improve the driver fitness (while driving in automated driving mode)
- Use case 5a: Mediator system initiates a planned takeover (from automated to manual driving mode).

2.2. Methodology

2.2.1. Participants

Tests were conducted involving naïve participants recruited by an external company to avoid any bias in the evaluations.

All participants were recruited with the following characteristics:

- At least 8/10 (even with correction) of visual acuity
- No colour blindness
- Valid driving license

Sixteen participants (56% male, 44% female) with an average age of M = 39.9 years (SD=14.1, range 23 - 60 years) and split in two age classes (8 participants of 23-30 years and 8 of 50-60 years) took part in the study. 56% of the participants had a high school diploma, 44% of them had a university degree. Participants drove an average of M=18097 kilometres per year (km/y) (SD =

1 Sometimes the use case has been split in 2 parts: part a and part b; UCxa is related to fatigue and UCxb is related to distraction.

10330, range 7000-40000 km/y) split in two ranges (8 participants ≤10000 km/year and 8 ≥10000 km/year) and in mixed types of roads.

The Advanced Driver Assistance Systems usage was 100% for Adaptive Cruise Control, Lane Departure warning, Forward collision Warning, 81% for the Blind spot, 75% for the Lane Keep Assistance, 56% for the Driver Alert and 6% for the Self-Parking assist system.

2.2.2. Procedure and design

2.2.2.1. Procedure

Before agreeing to take part into the test, the participants were informed about the main goals and procedure of the experiment, as requested from the MEDIATOR Ethics Committee. When a participant arrived for the test, he/she was welcomed, and he/she had to sign privacy modules, according to General Data Protection Regulation (GDPR) requirements and the study description forms as required by MEDIATOR Ethics Committee. The objective of the test was explained, underlining that the experimental campaign was part of a European research project. The experimenter explained that the aim was to evaluate the usability, acceptance, and perceived trustworthiness of the MEDIATOR HMI installed in the vehicle.

This was followed by the administration of questionnaires before the on-road test and by the instructions on the Wizard of Oz prototype and on the MEDIATOR HMI. The experimenter explained that:

- the participant is on a right-hand drive vehicle, seated on the left seat and accompanied by a driving wizard/a professional driver (on the right seat), who always controls the vehicle and supervised by an experimenter (on the back seat), with whom the participant has to interact
- the vehicle is a prototype, with non-functional double pedals and steering wheel, to allow test participants to have an “automated” vehicle experience even if it has no real automated functionalities

It was explained that the Mediator system

- allows to pass from manual to automated driving and vice versa. When the Mediator system or the experimenter asks the participant to change the driving mode (automated or manual), he/she can simulate this change by using the Mediator shifter, to evaluate the HMI elements involved in this process
- provides information about:
  - the current driving modality (automated or manual)
  - the transition from the automated to the manual driving and vice versa

and that this information is given on the frontal display, on the head unit display and on the steering wheel, seat, belt through visual interface, vocal messages, acoustic, haptic, and luminous feedback.

Then, the on-road session, in which all naïve participants tested the MEDIATOR HMI solutions through the defined use cases) on the predefined scenario, could start. During the on-road session, the participant was asked to pretend to do different non-driving-related tasks (such as get distracted or receiving a text message) and requested to follow the Thinking Aloud protocol and to answer to specific questions about the HMI. The experimenter took note of any user's comments and answers. At the end, after on-road test questionnaires were administered, in-depth evaluation of MEDIATOR HMI solutions usability and final comments collection followed. Thanks to the participant ended each test, that lasted 3 hours. Then, the preparation for the next test started with the sanitisation of the vehicle and of the test environment.

2.2.2.2. Scenario

The Italian MEDIATOR study was conducted on public roads in the Turin area (Figure 2.1), which was around 46 km length (10 km in urban and extra urban roads and 36 km on the highway).

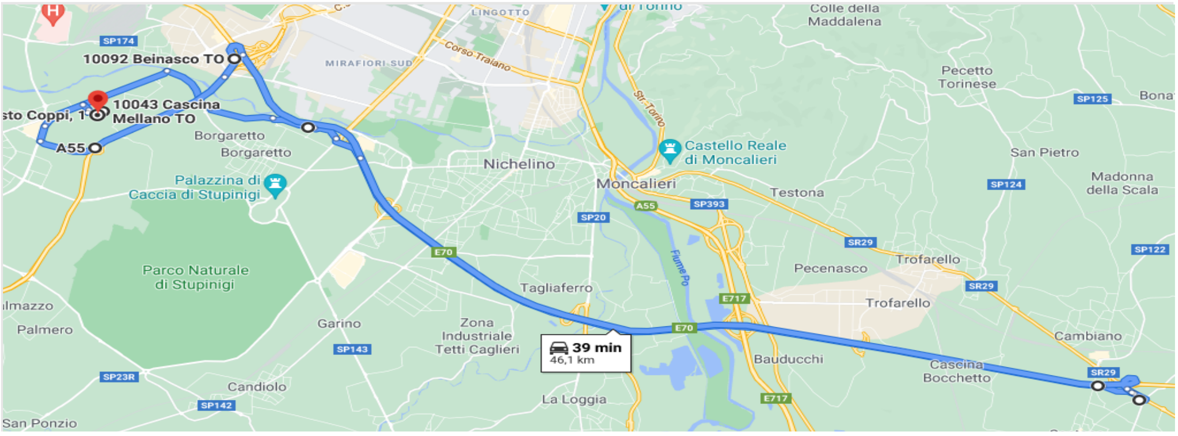


Figure 2.1 On - road Italian trials scenario (Source: Maps data ©2022 Google).

2.2.2.3. MEDIATOR HMI devices in the HF in-vehicle prototype

Grondelle et al. (2021) reports the design of a holistic HMI (visual, audio, haptic), based on the development of new devices (properly developed within MEDIATOR project) and several on-market devices (like LED strips, displays for cluster and head unit to be used by the naïve participants for feedback, infotainment, navigation purposes, for the sound system and the ambient lighting) that have been used in an integrated way with reference to the specific use case.

Fiorentino et al (2022) describes the Mediator HF in-vehicle prototype detailing with reference to HMI prototypal devices and HMI SW components.

For Italian study, the MEDIATOR HMI prototypal devices that were tested were the following:
- a cluster obtained through a PC monitor, positioned behind the steering wheel, where different icons and messages were shown depending on use cases and Mediator driving mode, which were automated (called piloted on the HMI) and manual (Figure 2.2)
- a head unit interface obtained by a monitor in the center of dashboard where navigation map and different messages were shown depending on the use cases (Figure 2.2)
- a series of Light Emitting Diode (LED) on the participant not functional steering wheel: they changed color depending on Mediator driving mode (Figure 2.2)
- some vocal messages that give advice depending on the use case and Mediator driving mode
- some acoustic feedback with different duration (short and long)
- a cushion that was inflated depending on the use case
- a seat belt giving naïve drivers haptic feedback, depending on the use case.

Figure 2.2 HMI stimuli on cluster, head unit and LEDs on steering wheel.

2.2.2.4. Use Cases and related Tasks

During the on-road test, participants experienced the seven different use cases previously described and doing consequently the tasks described inTable 2.1:

Table 2.1 Participants tasks.

USE CASE	TASK
Use case 1a	To image to get drowsy and to move the shifter to automated driving mode
Use case 1b	To simulate distraction (not to look at the road) and to move the shifter to automated driving mode
Use case 2	To image to desire to take back the control, and to move the shifter to manual mode
Use case 3a	To image not to be motivated to drive and to move the shifter to automated driving mode
Use case 3b	To imagine receiving a text message on the mobile phone and to move the shifter to automated driving mode
Use case 4	To image to be drowsy and evaluate the Mediator system action (remaining in automated driving mode)
Use case 5	To move the shifter to manual mode (after Mediator system request due to planned takeover)

2.2.3. Measurements

2.2.3.1. Vehicle data and automation state

The data related to vehicle and automation state were acquired by the prototype to verify the feasibility of this step, but they were not used to respond to the research questions. In fact:

- These data were mainly related to the professional driver, who was always the same and doing a standard driving, so the automated vehicle driving style didn't have impact on the automated Driving subjective evaluation
- As agreed in the project, participants were recruited not sleep deprived, then they were not tired, so the related use case was simulated. This was decided because otherwise the replicability of the use cases could not be guaranteed.
- During the handover/takeover phase, objective data were not completely representative of the driver behaviour because the aim of the trial was on the HMI evaluation and then the participants were asked to see all HMI escalation sequences, in the conditions in which this was feasible (e.g., traffic condition, prototype instability)

- The Mediator shift lever first prototype (used to pass from manual to automated mode and vice versa) mounted on the vehicle had an unstable behaviour and, in these situations, it didn't acquire promptly the action of the participant (and the experimenter had to move it before entering in the next use case).

2.2.3.2. Physiological data

Physiological data were not acquired because participants were recruited not tired and the drowsiness/fatigue use cases were simulated.

2.2.3.3. Eye tracking data

They eye tracking data, as agreed in the project, were acquired just to be used by the Mediator system to evaluate driver simulated distraction during the specific use cases on driver distraction, following the request of the experimenter.

2.2.3.4. HMI interactions

The HMI interactions were mainly acquired using a Thinking Aloud protocol and participants comments. Before the on-road test, participants were instructed to think aloud during driving and comment what they saw and felt about automated driving and about feedback/messages/elements of the MEDIATOR HMI. Before each use case, they were introduced by the experimenter in the specific use case context, without any explanation on the HMI. Any observation, comment on the MEDIATOR HMI solutions and on the automated driving were reported by the experimenter in a specific observational grid on an Excel file. After the on-road test, an in-depth interview on each use case was executed, using photos of the HMI to help participants to remember what they tested, asking participants evaluations and comments on each experienced aspect, highlighting advantages and disadvantages of the HMIs in each use case.

2.2.3.5. Subjective ratings

The participants evaluation on each HMI solution in each use case was asked to participants through a 7-points scale questionnaire (Figure 2.3) on the different HMIs (e.g., frontal HMI on the cluster, HMI on the head unit display, steering wheel LEDs...), shared with partners before the tests.

Figure 2.3 HMI evaluation scale.

In addition, the experimenter asked the participant to list the most important HMI solutions for each investigated use case and to comment on them.

2.2.3.6. Questionnaires

Different standard questionnaires were administered to participants before and at the end of the on-road test session. Before each driving session, participants were asked to fill the following questionnaires:

- ADAS owned and used questionnaire in L3Pilot project (Metz et al., 2019) and ADAS&ME project (Pereira Cocron et al., 2019)
- Affinity for technology (Franke et al., 2019)
- Attitudes towards vehicle automation (Pereira Cocron et al., 2019)
- SUaaVE Acceptance of Automated Vehicles (Post et al., 2020)
- Van der Laan questionnaire (Van der Laan et al., 1997).

At the end of driving sessions, before the in deep interview on each use case, the following questionnaires were administered:

- SUaaVE Acceptance of Automated Vehicles (Post et al., 2020)
- Van der Laan questionnaire (Van der Laan et al., 1997)
- Trust in Mediator system (Jian et al., 2000)
- Usability of Mediator system (Brooke, 1996)
- Takeover questionnaire, adapted from L3Pilot questionnaire.

2.2.4. Data pre-processing and statistical analysis

A preliminary analysis was done on subjective evaluations, to verify if some participants had to be considered as outliers, due to anomalous usage of ratings, considering all the HMI ratings. In particular, the analysis verified if there was a good variability in each participant's scores and at least a minimum coherence between each participant's score and the average scores of all other participants. Nobody had to be excluded due to this outlier analysis. All standard questionnaires were analysed following the standard procedures specific to each one of them (e.g., Van der Laan and SUaaVE questionnaires requires Cronbach's α analysis and calculation of new indicators as average of row data). The comparison among pre and post-test results was done using paired T-tests with 90% of confidence as threshold (but also exact probabilities).

HMI ratings were analysed considering average values, confidence intervals and ANOVA and multi comparing tests. Similarly for percentage of participants that consider relevant each HMI, considering confidence intervals on percentages. Duncan multiple comparison test and 90%[2] of confidence level were used. The graphs on ratings, reported in next paragraphs, show average evaluation, confidence interval on average with 90% of confidence limit, result of Duncan multiple comparison test and samples. An example of graph on average ratings is shown in Figure 2.4.

2 In this test the sample had to be small because had to evaluate the first HMI versions, so a 90% confidence limit was used to highlight minor significant difference.

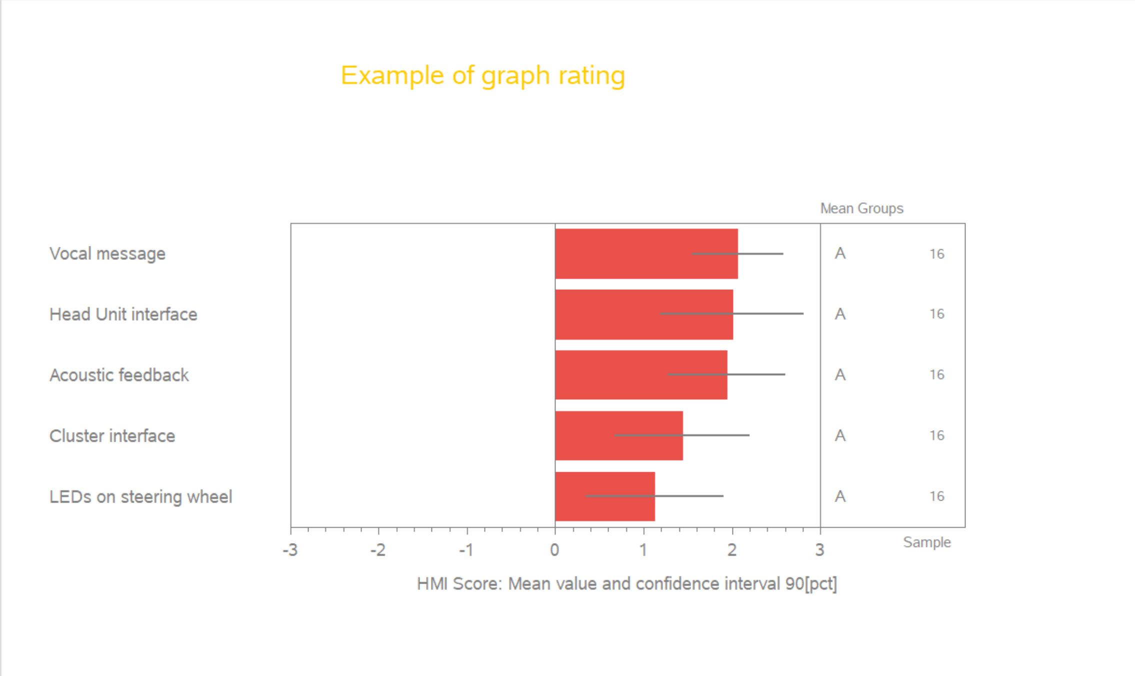


Figure 2.4 Example of graph on ratings.

Similarly, the graph on relevant HMIs shows percentage of participants that consider relevant each HMI, confidence interval on percentage with 90% of confidence limit, result of Duncan multiple comparison test and samples. An example of graph on percentage (of relevant HMIs) average ratings is shown in Figure 2.5.

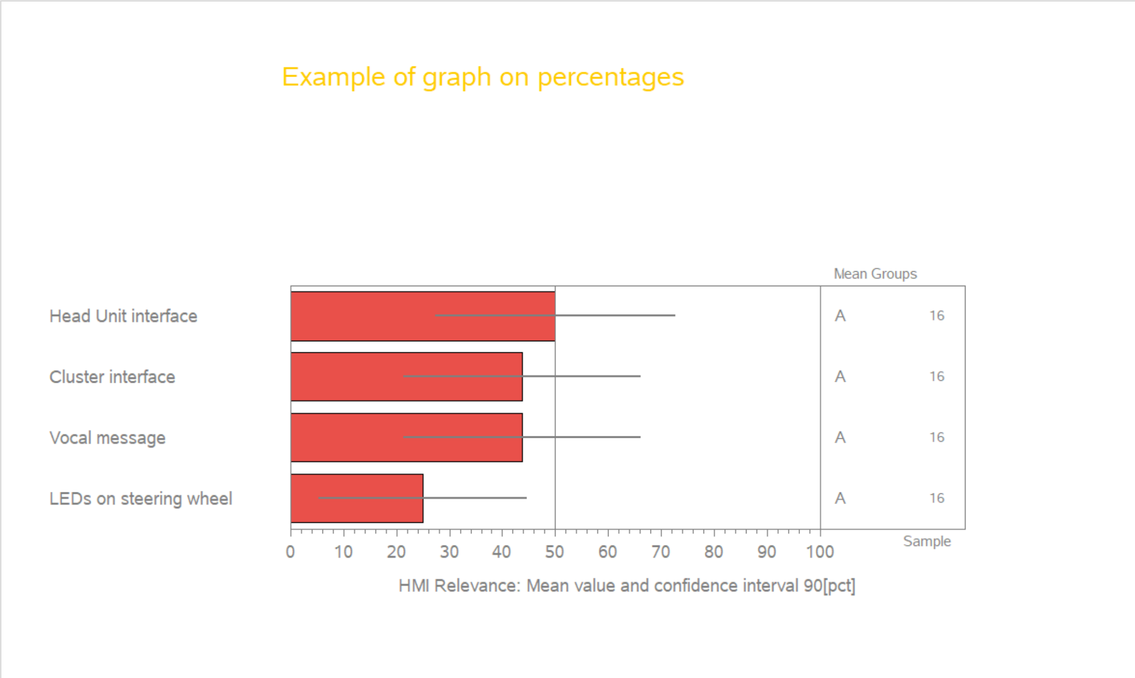


Figure 2.5 Example of graph on percentages on relevant HMIs.

2.3. Results

2.3.1. RQ 1: Are there any differences in the users' acceptance of Automated Vehicle and MEDIATOR HMI before and after the trial?

Both SUaaVE (Post et al., 2020) and Van der Laan questionnaires (Van der Laan et al.,1997) have average positive scores, before and after the on-road user test. In SUaaVE questionnaire, some significant improvements can be observed after the test with MEDIATOR HMI solutions on:

- Acceptance of automated vehicles
- Trust in automated vehicle technology

This means that, in a sample of users with positive and high expectations on automated vehicles, the MEDIATOR HMI and the experience on MEDIATOR project vehicle had a positive impact on participants, increasing the acceptance of Automated Vehicles and the trust in their technology (Table 2.2).

Table 2.2 Paired T-test results.

Item	t-value	p-value
Acceptability/acceptance (SUaaVE)	$t(15)=2.27$	$p=0.04$
Trust in CV technology (SUaaVE)	$t(15)=2.37$	$p=0.03$
Perceived Convenience (SUaaVE)	$t(15)=-1.29$	$p=0.21$
Perceived Safety (SUaaVE)	$t(15)=1.19$	$p=0.25$

Figure 2.6 and Figure 2.7 show the average evaluations and related confidence intervals.

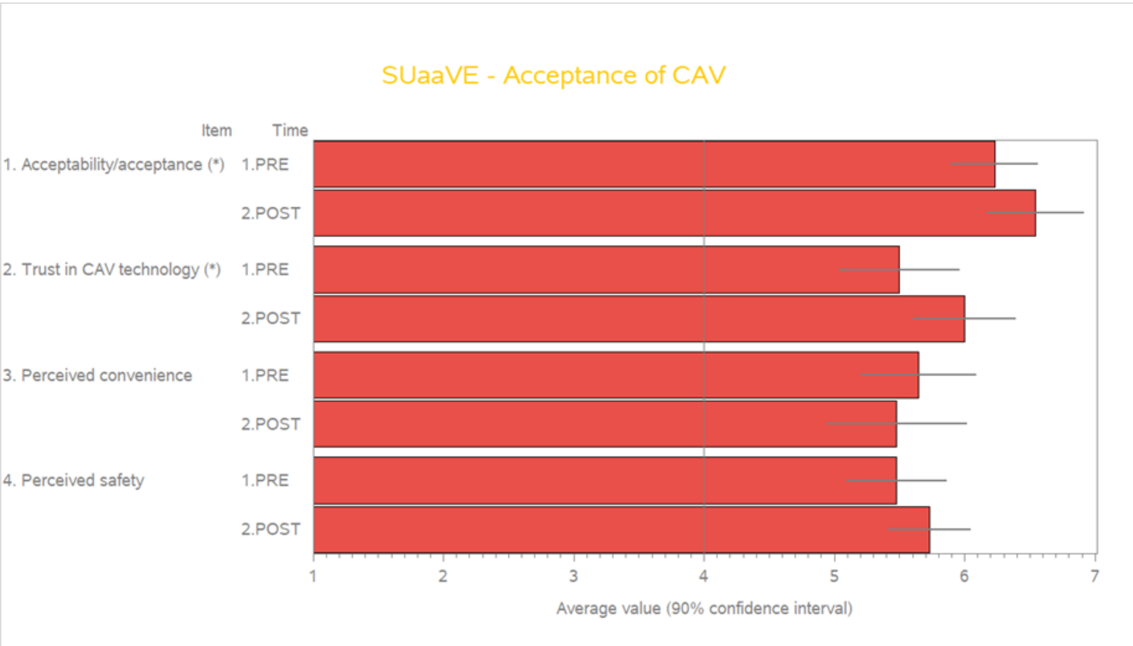


Figure 2.6 SUaaVE questionnaire evaluation.

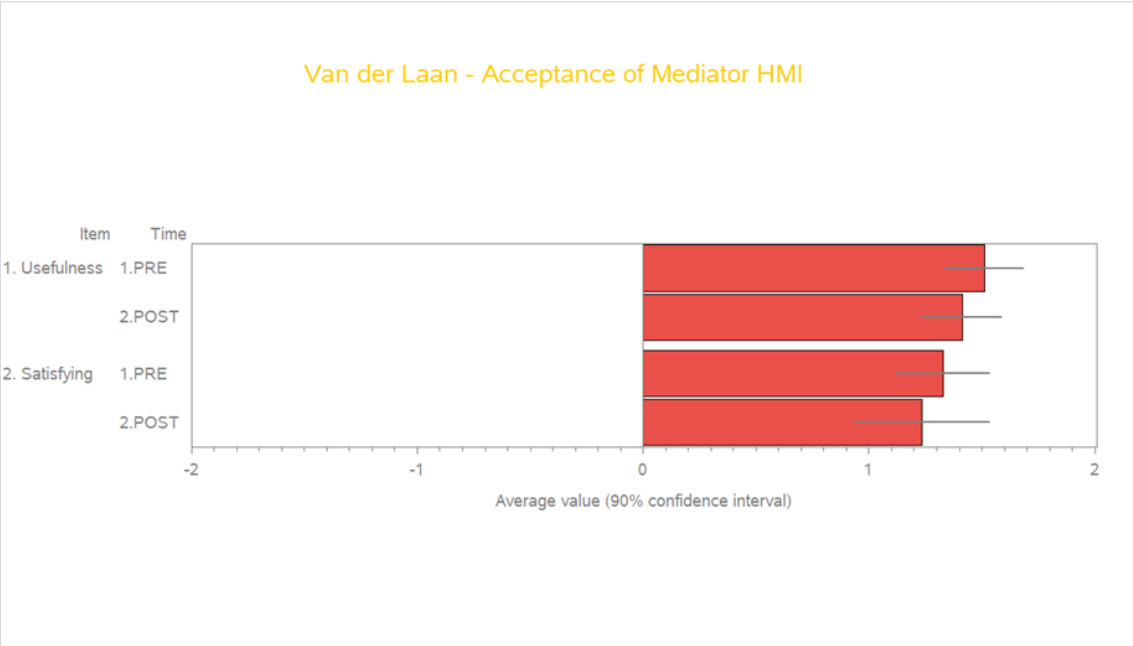


Figure 2.7 Van der Laan acceptance evaluation

2.3.2. RQ 2: Do people consider MEDIATOR HMI usable, and do they trust it?

Both on Usability and Trust all average evaluations are significantly positive and without any negative evaluations

2.3.2.1. Usability

The average score of System Usability Scale (SUS) analysis for MEDIATOR HMI is 76 (Figure 2.8), then the score can be classified as good.

Figure 2.8 MEDIATOR HMI average score on System Usability Scale.

The graph (Figure 2.9) shows the complete distribution of System Usability Scale (SUS) indicator for each person in the sample; all the ratings are in the range between marginal and acceptable.

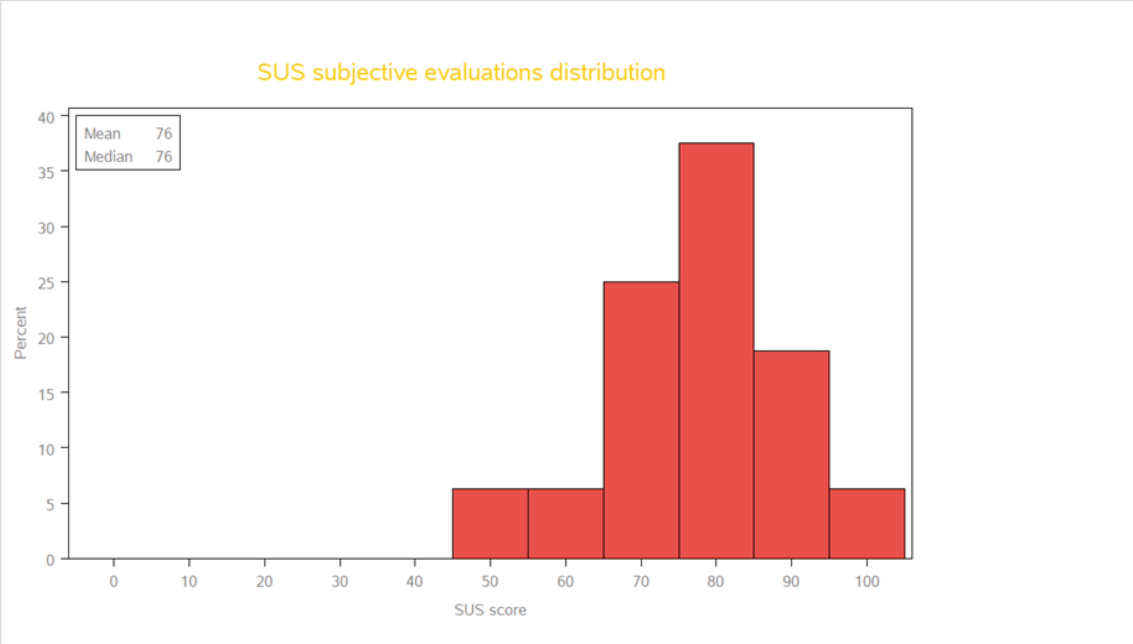


Figure 2.9 SUS subjective evaluations distribution.

2.3.2.2. Trust

All the evaluations on trust in Mediator system (Jian et al., 2000) are positive, the minimum evaluation is 5 in a 7-point scale and average evaluation is 5.9 (Figure 2.10).

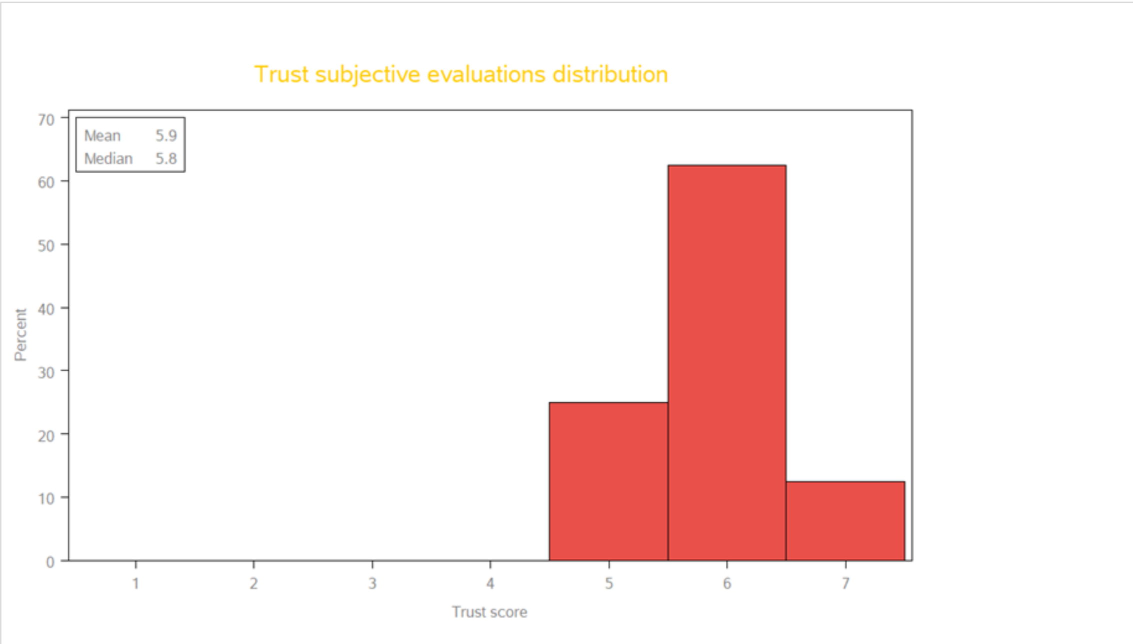


Figure 2.10 Trust subjective evaluations distribution.

2.3.3. RQ 3: With the MEDIATOR HMI, is the handover request (from human to automation), by driver input, usable (effective, efficient and satisfying) and acceptable?

This research question was answered through the evaluations of HMIs used on use case 3a: the driver is no more motivated to drive and indicates it via the HMI, by using the Mediator shifter.
In general, the HMI evaluation is positive. All the single HMI solutions have a significantly positive evaluation, a part the LEDs on the steering wheel, and are not statistically different from the neutral point. Also, differences between the different HMI solutions are not statistically significant (Figure 2.11).

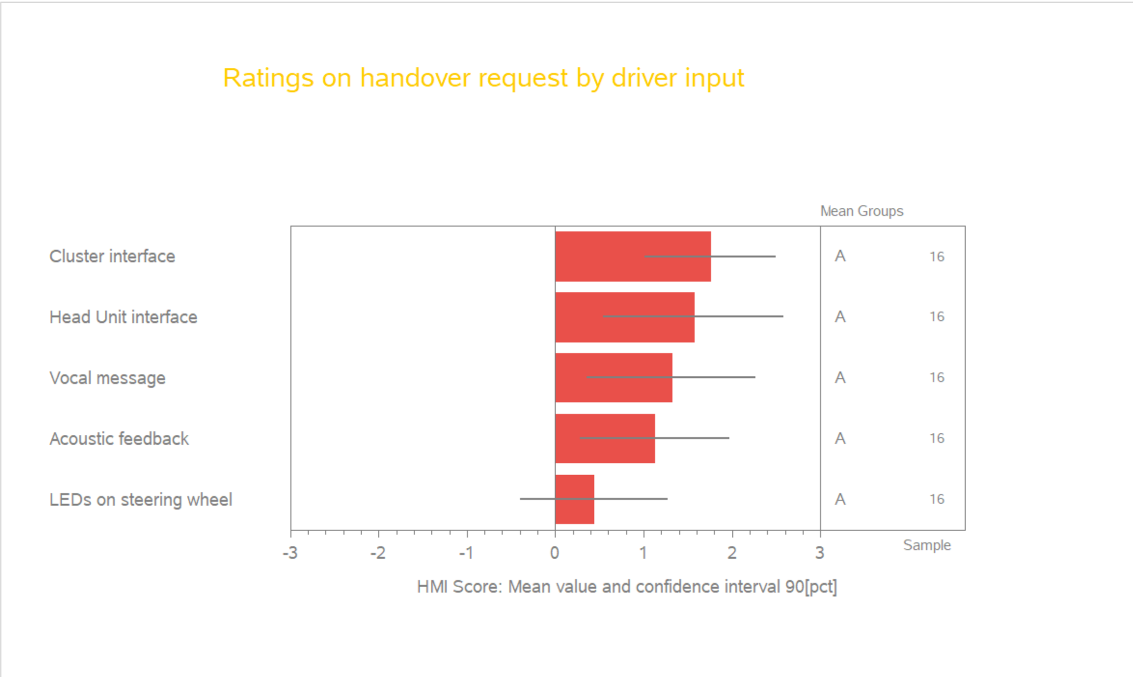


Figure 2.11 Ratings on HMIs in use case 3a.

No statistical difference among HMIs was found in use case 3a (Figure 2.12). Acoustic feedback is not present in the graph because has not been mentioned as one of the relevant HMI, but its percentage (0%) is not significantly lower respect to percentage of mentions for LEDs on the steering wheel.

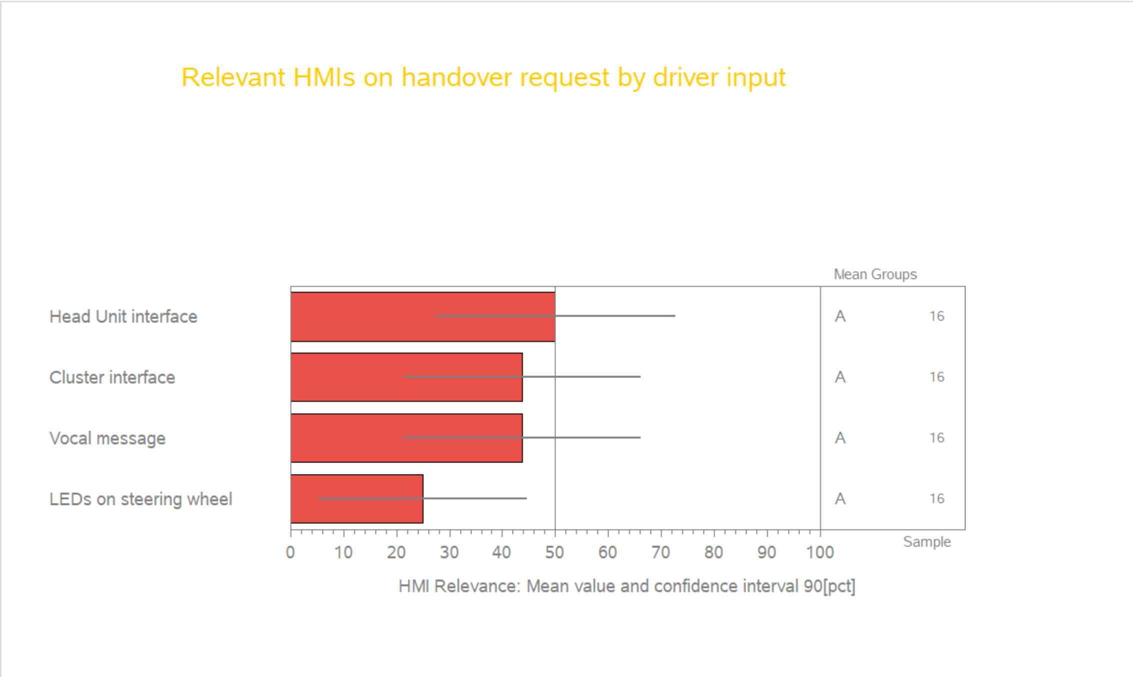


Figure 2.12 Percentage of relevant HMIs in use case 3a.

Participants, during the on-road session, gave comments and suggestions on the HMI used in this use case:

- Some participants would like to be informed by the HMI, if and why the automated driving mode is not available
- Some participants did not understand immediately the meaning of the steering wheel icon with only one hand (Figure 2.13), which represented the automated driving mode (translate as Driving Piloted, DP, activated).
-

Figure 2.13 Automated driving mode icon.

- Participants would like to have a lower latency time

- Participants considered this use case 3.a (driver not motivated to drive asks for automated driving) more critical than the opposite situation (use case 2: driver desires to drive) and then, they would have the certainty that the automated driving mode is active through an even more salient/comprehensible HMI.

During the interview after on-road session participants gave some comments:
- There were some problems regarding the labels on the cluster: they were considered barely legible because they were too small
- Instead, the road colour change was appreciated because it was visible and comprehensible
- Participants stated in a positive way that the HMI on the head unit is considered to have a good visibility
- Some participants did not understand the meaning of the steering wheel icon with only one hand (Figure 2.13), because it was not representative for the automated driving mode
- Participants considered the vocal message comprehensible but not so useful because it was a choice of the driver to do the handover
- Participants considered the LEDs less useful because not very visible during daylight and possibly annoying during the night
- In case of driver request for automated driving, the participants considered the acoustic feedback to be not very useful, because it is a voluntary action of the driver to do the handover
- Participants considered the DP (Driving Piloted) abbreviation (Figure 2.13) not comprehensible, because it is in the English language.

2.3.4. RQ 4: With the MEDIATOR HMI, is the handover request (from human to automation), caused by distraction, usable and acceptable?

This research question was answered through the evaluations of the HMI used in use case 1b: The Mediator system initiates the handover (from manual to automated driving mode) caused by distraction [The participant was asked not to look at the road in front of him to simulate being distracted]. All the single HMIs have a significantly positive evaluation, and there is no statistical difference among these single evaluations (Figure 2.14).

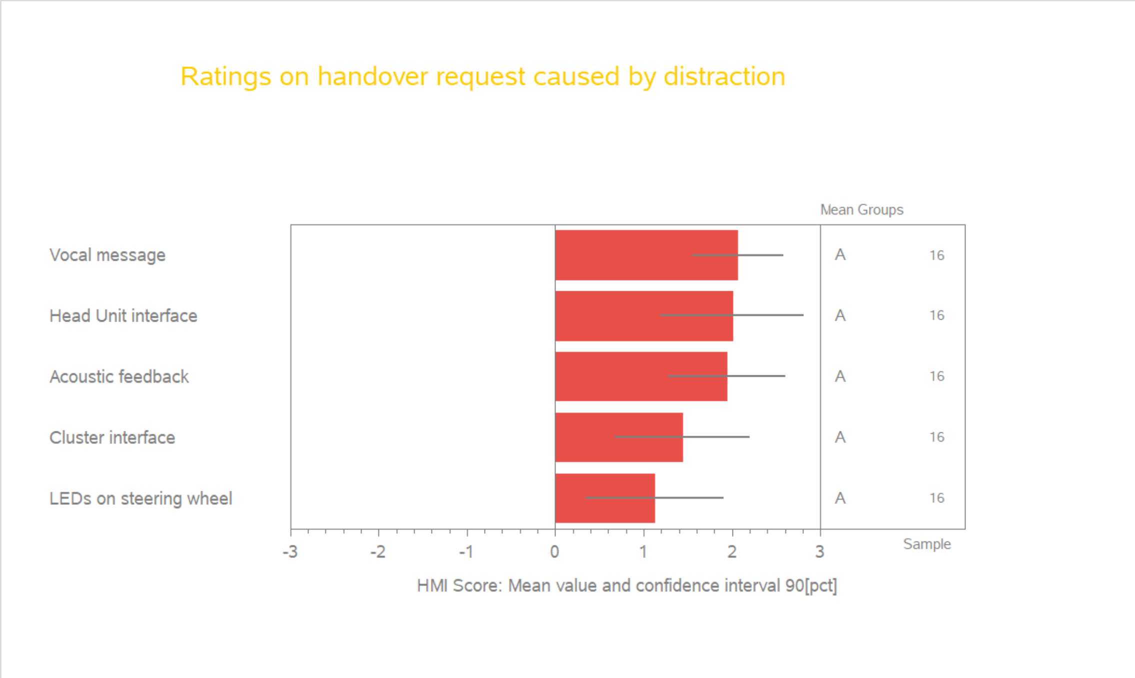


Figure 2.14 Ratings on HMIs in use case 1b.

In term of relevance of the different HMIs for the use case 1b, even if differences are small, Head Unit message is considered slightly more impacting, together with the acoustic feedback (Figure 2.15). In particular, the percentage of participants that consider the head unit relevant is significantly higher respect to all other HMIs a part the acoustic feedback, but acoustic feedback is significantly more relevant only respect to LEDs on steering wheel and to cluster HMI. In the use case 1b, the Head Unit message is evaluated more relevant than all the other (a apart acoustic feedback, which is in the same mean group) and the acoustic feedback is evaluated more relevant respect to LEDs on steering wheel and HMI on cluster.

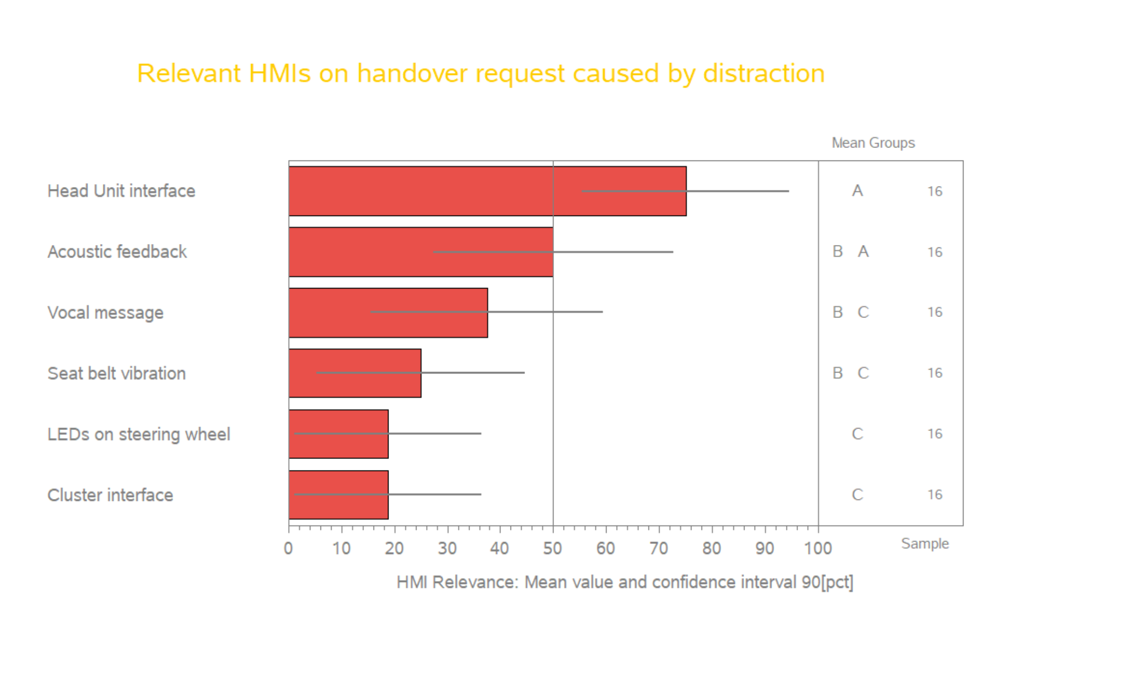


Figure 2.15 Percentage of relevant HMIs in use case 1b.

Participants, during the on-road session, gave comments and suggestions on the HMI used in use case 1b:

- Participants would prefer to have a quicker HMI escalation because the distraction was considered a dangerous situation: the HMI should faster propose the transition to automated driving
- Some participants would like the HMI says directly what to do, e.g., move the shifter
- The acoustic feedback for the distraction should be different from the other ones to recognize it immediately
- Some participants appreciated the information on how long automated driving mode is available because it informs how long the driver is free to do Non-Driving Related Tasks (but the majority did not notice it)
- Participants would like to have a lower latency time in a handover situation[3].

During the interview after on-road session participants gave some comments:

- There were some problems regarding the red labels on the cluster: they were considered difficult to read (too small) and barely visible because written in red colour
- Participants stated that this information on the cluster did not capture the driver attention
- Participants described this HMI on the head unit as simple to understand
- Participants stated the LEDs were not visible with daylight. Furthermore, participants imagined that the LEDs could be annoying during nightlight conditions
- The colour change of the LEDs was considered not very comprehensible
- Participants stated the Seat belt haptic feedback was desirable, even if they consider it a little annoying because the effect was too strong
- Participants appreciated the acoustic feedback because it captures the user's attention.

3 Participants experienced too longer latency time when handover and takeover are needed.

2.3.5. RQ 5: With the MEDIATOR HMI, is the handover request (from human to automation), caused by an event detected, usable and acceptable?

This research question was answered through the evaluations of HMI used on use cases 3b: the Mediator system detects a simulated event (text message) and it suggests the handover (from manual to automated driving mode). [The participant is asked to imagine receiving a text message on the mobile phone]. In general, the HMI evaluation is positive. All the different HMIs have a significantly positive evaluation excluding the LEDs on the steering wheel, whose evaluations are not statistically different from the neutral point (Figure 2.16). However, differences among the HMIs are not statistically significant.

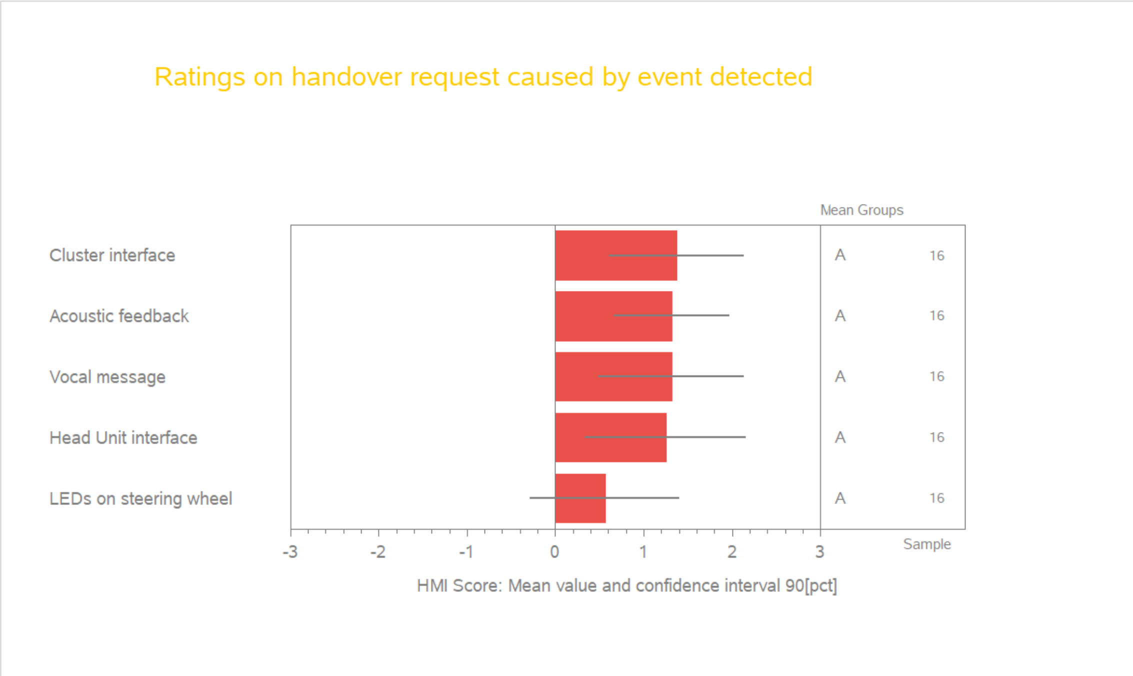


Figure 2.16 Ratings on HMIs in use case 3b.

In term of relevance of HMIs, vocal messages, head unit and cluster are considered more impacting respect to the other HMIs (Figure 2.17). Percentage of citation for LEDs on steering wheel and Acoustic feedback are not statistically different between them and from 0.

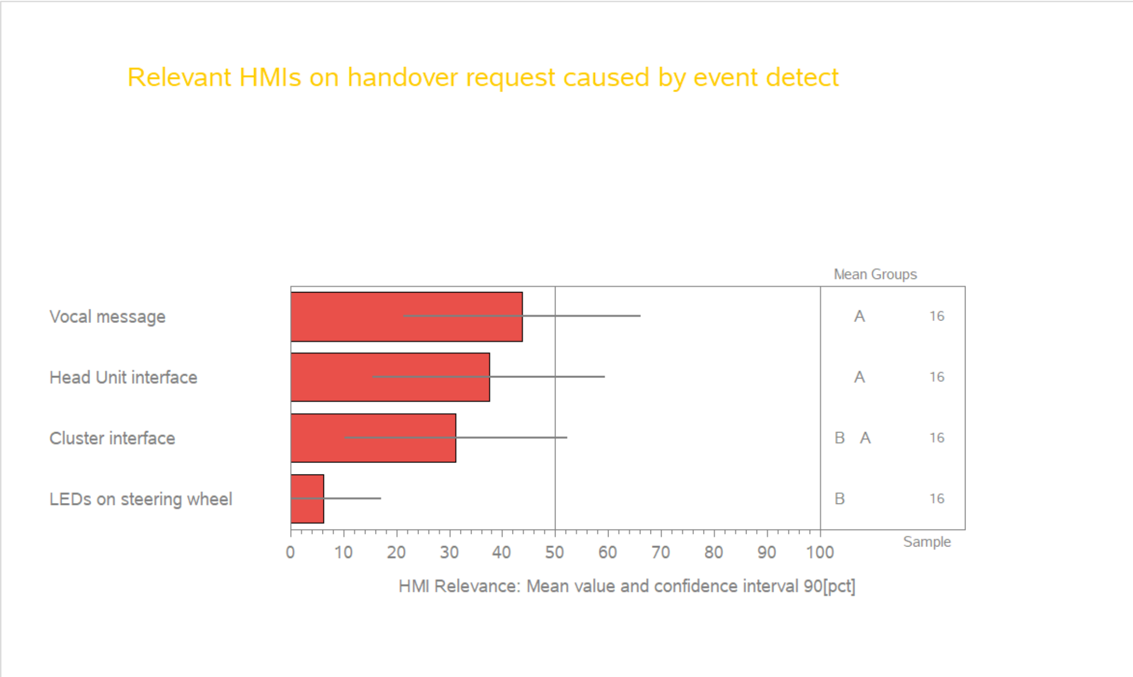


Figure 2.17 Percentage of relevant HMIs in use case 3b.

Participants, during the on-road session, gave comments and suggestions on the HMI used in the use case 3b:

- Participants would prefer to have a lower latency time
- Some participants said that if the Mediator system offers the automated driving mode whenever a WhatsApp/SMS message arrives, this will happen too often
- Participants did not appreciate that on the cluster HMI there was no feedback on the text message presence. Having the automated driving availability information only was not considered sufficient to understand the situation

During the interview after on-road session participants gave some comments:

- Some participants evaluated the HMI on the cluster comprehensible even if some participants did not appreciate that on the cluster HMI there was no feedback on the text message presence (e.g., icon with the letter).
- Most of the participants considered the labels too small to read
- Some participants did not understand the meaning of the steering wheel icon with only one hand (Figure 2.13) on the head unit, because it is not representative of the automated driving mode
- Participants appreciated the presence of vocal messages, but they would like a vocal message on to the text message presence
- Participants considered the LEDs useless and not visible in daylight
- Some participants appreciated the acoustic feedback because it captured the attention of the driver, while other didn’t perceived it or considered it not useful.

2.3.6. RQ 6: With the MEDIATOR HMI, is the handover request (from human to automation), caused by drowsiness, usable and acceptable?

This research question was answered through the evaluations of HMI used on use cases 1a: Mediator system initiates the handover (from manual to automated driving mode) caused by simulated drowsiness. [The participant is asked to imagine getting drowsy]. All the single HMIs have a significantly positive evaluation and are not statistically different one to the other (Figure 2.18).

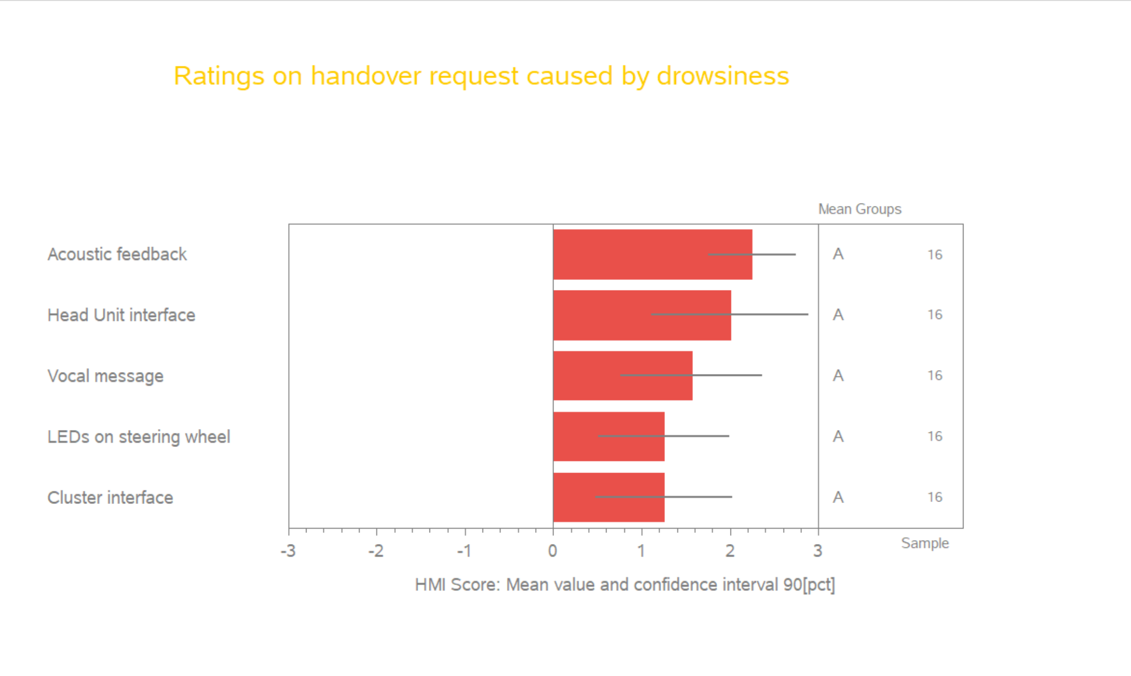


Figure 2.18 Ratings on HMIs in use case 1a.

Asking about relevance of HMIs, head unit and acoustic feedback are considered slightly more impacting respect to the other HMIs (Figure 2.18). In particular, the percentage of participants that consider the head unit HMI and the acoustic feedback relevant is significantly higher respect to all other HMIs excluded vocal message and seat belt haptic feedback, but also versus these HMIs the difference is at limit of statistical significance.

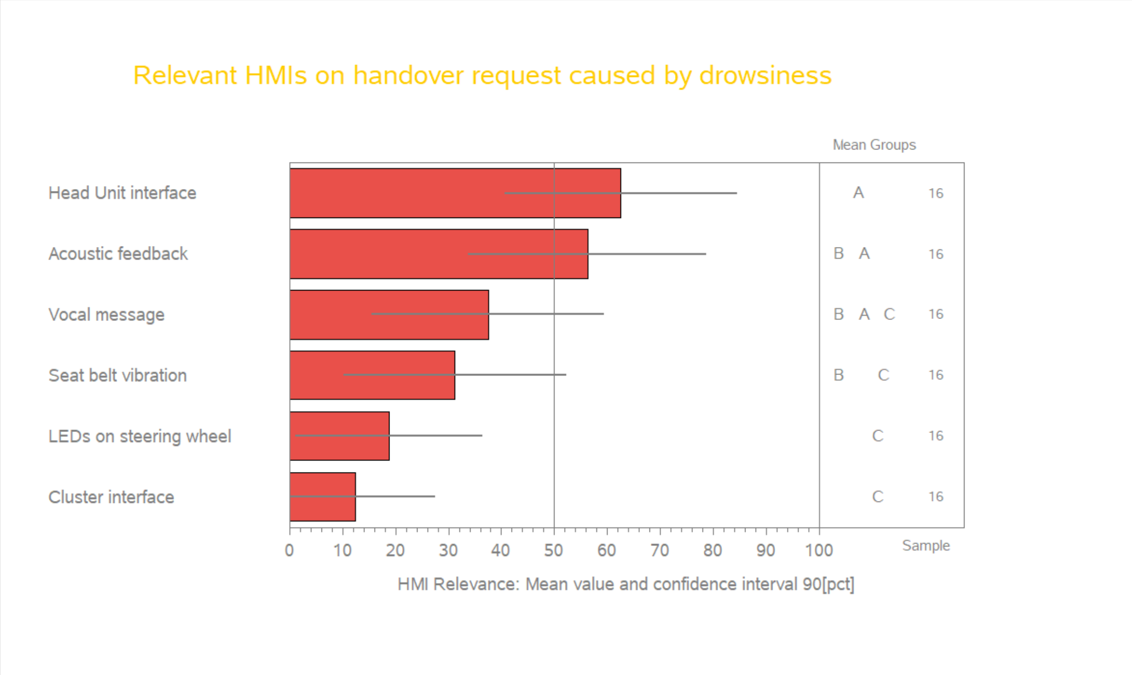


Figure 2.19 Percentage of relevant HMIs in use case 1a.

Participants, during the on-road session, gave comments and suggestions on the HMI used in the use case 1a. According to participants, this was the most dangerous situation because the fatigue in the manual mode driving was considered, even if simulated. For this reason:

- They would have preferred to have a quicker HMI escalation and a lower latency time
- They appreciated the more alerting HMI like the seat belt haptic feedback

Test had the aim of highlighting possible kind of issues on the designed HMI, to define guidelines for the redesign before the next user test. During the interview after the on-road session participants gave some comments:

- There were some problems regarding the labels on the Cluster: they were considered difficult to read (because too small) and barely visible because written in red colour
- Participants considered visible and legible the HMI on the Head Unit. They considered the icon of the eye, showing a degraded condition ("degradata" as in Figure 2.20). Some issues were related to who/what the word "degradata" was associated to (the user or the system).

Figure 2.20 HMI in use case 1a.

- Participants considered desirable the seat belt haptic feedback, even if they considered it a little annoying because the effect was too strong
- Participants appreciated the acoustic feedback because it captures the attention of the user.

2.3.7. RQ 7: With the MEDIATOR HMI, is the takeover request (from automation to human), caused by driver desire and input, usable and acceptable?

This research question was answered through the evaluations of HMI used on use case 2: the driver desires to take back the control and indicates it by using the Mediator shifter. In general, the HMI evaluation is positive. All the different HMIs have a significantly positive evaluation excluding the LEDs on the steering wheel scores that are not statistically different from the neutral point (Figure 2.21). However, differences among HMIs were not statistically significant.

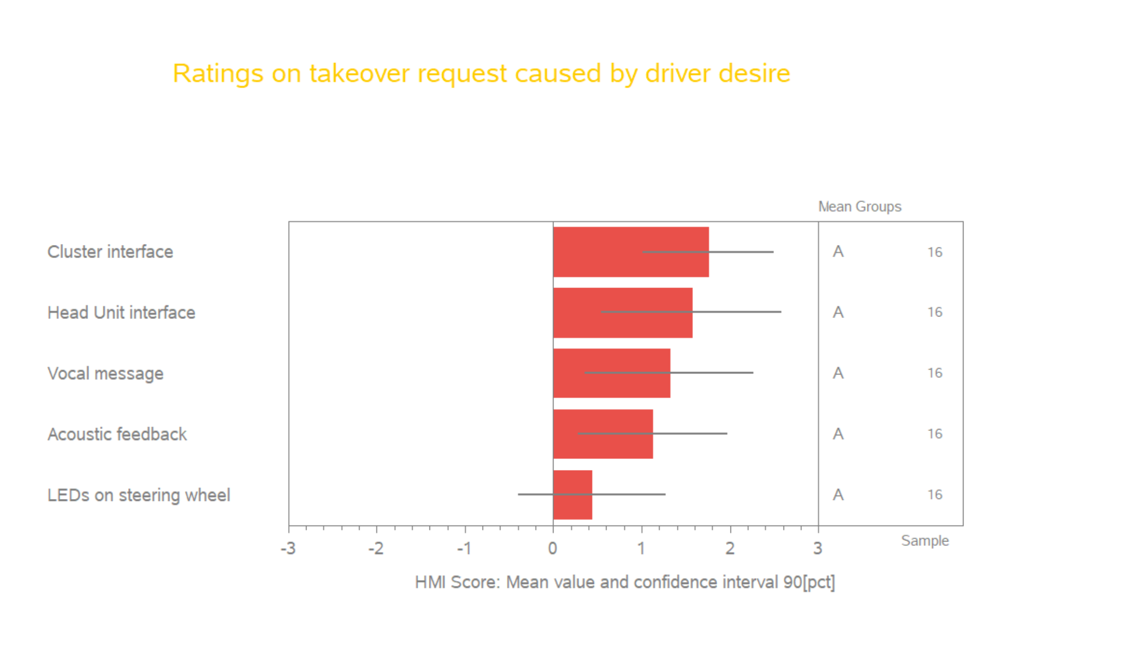


Figure 2.21 Ratings on HMIs in use case 2.

When asking about the most relevant HMIs also during the handover use case, no significant difference can be observed among them (Figure 2.22). Only acoustic feedback was not mentioned (and it was considered not very useful because in this use case the participant did a voluntary action).

Figure 2.22 Percentage of relevant HMIs in use case 2.

Participants, during the on-road session, gave comments and suggestions on the HMI used in the use case 2:

- Some participants complained on the fact that the MEDIATOR HMI covered the map of the navigator indications
- Participants would like to have an HMI, showing the transition from automated to manual driving mode with a lower latency time. For some participants, the MEDIATOR HMI did not inform the driver clearly (with a label) when the vehicle is in manual driving mode except during the transition. For most of the sample, the route colour change is sufficient to inform the driver regarding the current driving mode.

During the interview after on-road session participants gave some comments:

- Participants evaluated the HMI on the cluster as comprehensible even if some participants complained on the fact that the HMI is partially occluded by the steering wheel and that the labels are too small to be read
- Participants stated the HMI on the Head Unit is easy to understand and visible. They appreciated particularly the icon with both hands on the steering wheel
- Participants considered the vocal message comprehensible and useful because it confirms the choice of the user
- Participants stated the LEDs are not very useful and not visible during the daylight
- Participants did not consider the acoustic feedback very useful, because, in this use case, there was a voluntary action of the driver.

2.3.8. RQ 8: With the MEDIATOR HMI, is the planned takeover request (from automation to human) usable and acceptable?

This research question was answered through the evaluations of HMI used on use cases 5a: Mediator system initiates a planned takeover (from automated to manual driving mode). All the single HMIs have a significantly positive evaluation and are not statistically different among them (Figure 2.23). A difference at limit of significance can be observed among vocal message and LEDs on steering wheel, while all the other HMIs have intermediate scores not statistically different from both vocal message and LEDs.

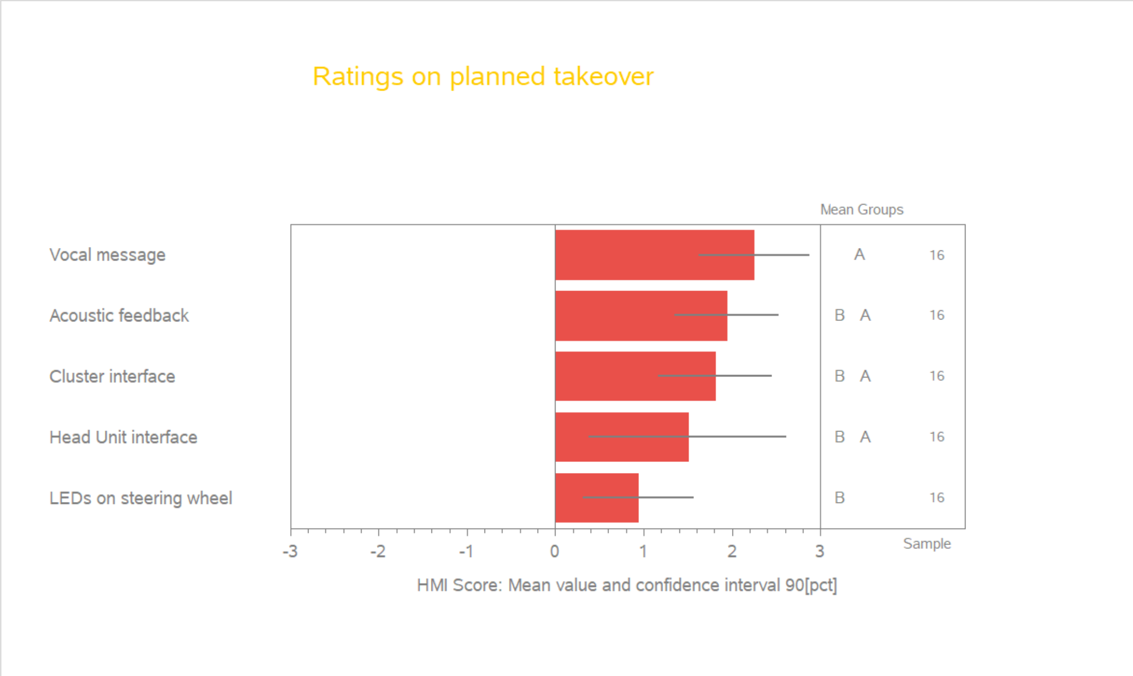


Figure 2.23 Ratings on HMIs in use case 5a.

Asking about relevance of HMIs, Vocal message and Head Unit interface are considered more impacting respect to the other HMIs (Figure 2.24).

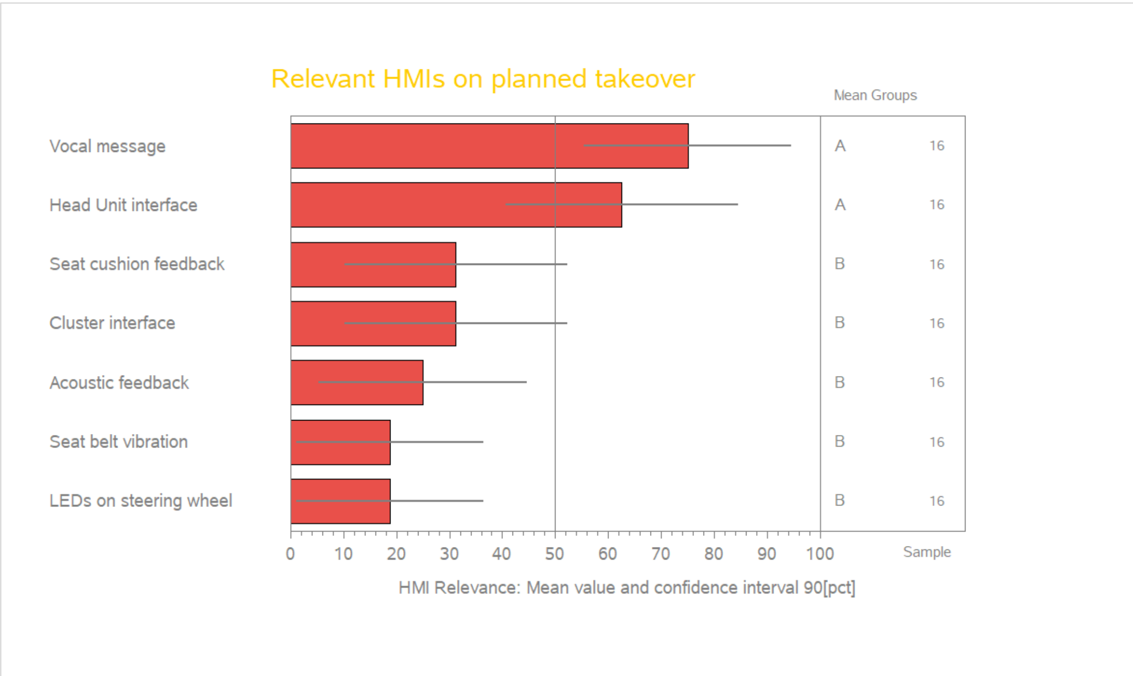


Figure 2.24 Percentage of relevant HMIs in use case 5a.

Participants, during the on-road session, gave comments and suggestions on the HMI used in the use case 5a:

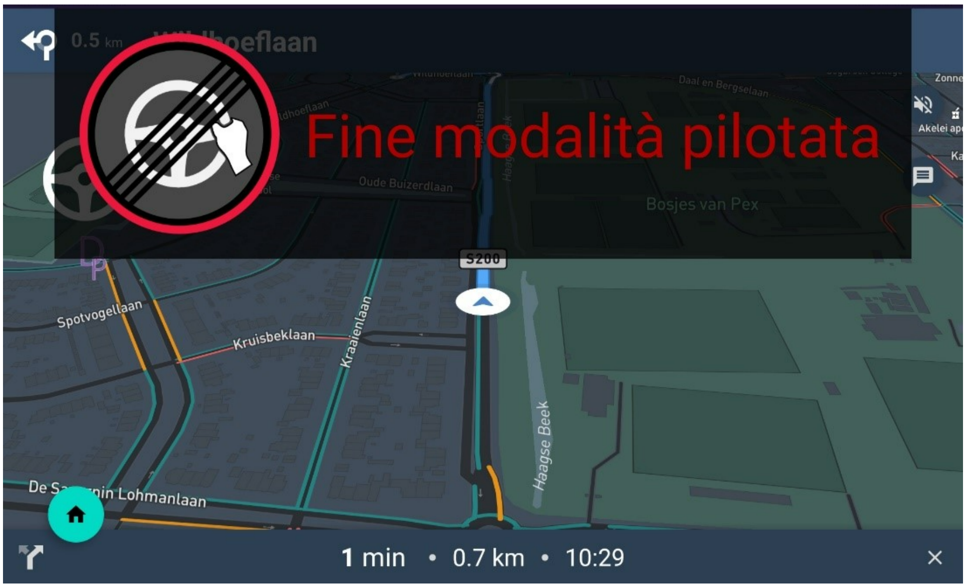
- To have enough time well in advance to resume the control of the vehicle was important for the participants
- Participants would like to have an HMI with lower latency time
- Some participants considered the cushion inflation and the acoustic feedback as redundant in this use case
- Some participants considered the messages volume as too low
- The route colour change on the cluster was considered comprehensible
- Some participants did not understand the meaning of the strikethrough steering wheel icon with only one hand (Figure 2.25) where the text can be translated as "Driving Piloted ends"
- Some participants complained because the MEDIATOR HMI covers the map of the Navigator.
- 

Figure 2.25 Strikethrough steering wheel icon.

During the interview after on-road session participants gave some comments:
- Participants considered the HMI on the cluster too small to be read
- Some participants did not understand the meaning of the numbers on the road representation, which represented, when in automated driving, the time till the active automated mode ends ("disponibile guida manuale" as in Figure 2.26) and, when in manual driving, the time till the automated mode can be available ("disponibile guida pilotata" as in Figure 2.27).

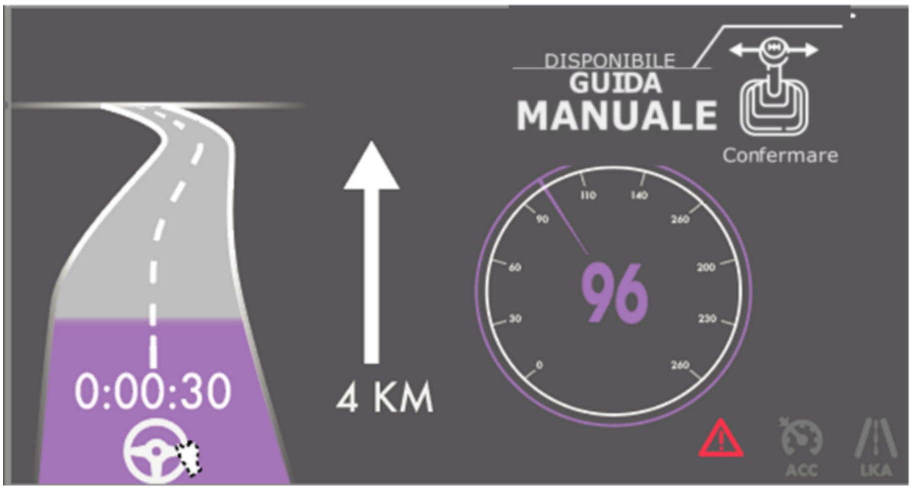


Figure 2.26 Representation on time till the active automated mode ends.

Figure 2.27 Representation on time till the automated mode can be available.

- Participants evaluated the labels on the Head Unit visible and comprehensible, except for the strikethrough steering wheel icon with only one hand (Figure 2.25)
- Participants considered the vocal messages to be comprehensible
- Regarding the LEDs, some participants appreciated the red colour and the flashing mode because the feedback more visible. Having LEDs always on was not appreciated
- Participants stated the seat belt haptic feedback was useful
- Participants considered the seat cushion that inflates useful even if it not well positioned (too towards the seat anterior part) and due to this it was difficult to feel it with the legs
- Participants agreed that there was too much information in this use case.

2.3.9. RQ 9: With the MEDIATOR HMI, is the improvement of driver fitness during an automated driving mode, caused by drowsiness, usable and acceptable?

This research question was answered through the evaluations of HMIs used on use cases 4: The Mediator system reacts to the simulated status of driver drowsiness by initiating an action to improve the driver fitness (while driving in automated driving mode). All the different HMIs have a significantly positive evaluation and are not statistically different among them (Figure 2.28). A difference at limit of significance can be observed respect to LEDs on steering wheel that have also an average value at limit to be considered neutral.

Figure 2.28 Rating on HMIs in use case 4.

When asking about the most relevant HMIs during the handover use case too, no significant difference can be observed among them (Figure 2.29).

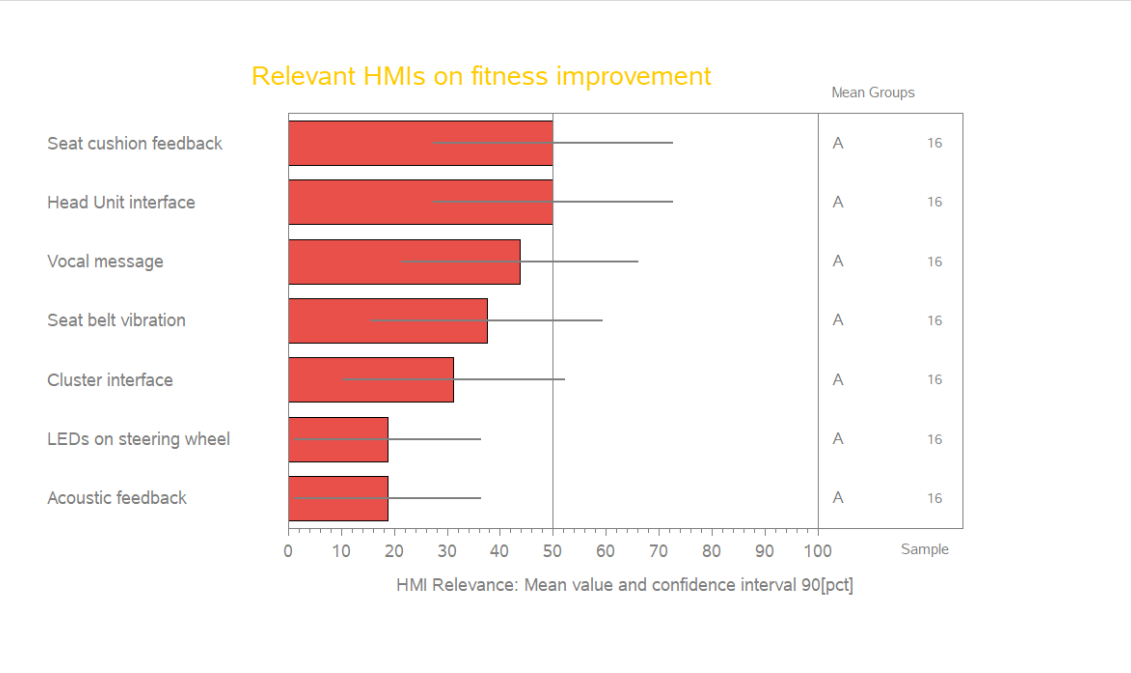


Figure 2.29 Percentage of relevant HMIs in use case 4.

Participants, during the on-road session, gave comments and suggestions on the HMI used in the use case 4. This was considered a dangerous situation, then participants would prefer to have:

- Quicker HMI escalation
- Louder acoustic feedback
- Vibration in the cushion instead of its inflation because it would be more perceivable
- More vocal information and fewer labels.

During the interview after the on-road session participants gave some useful comments:

- There were some problems regarding the labels on the Cluster: they were considered barely legible because too small
- Participants said the HMI on the Head Unit is comprehensible. Some participants did not appreciate the word "degradata" (Figure 2.20), because it seems not currently used and it is not clear to whom/what it is associated
- Participants considered the vocal message comprehensible
- Participants considered the seat cushion that inflates useful, even if it was not well positioned (too much in the anterior part of the seat) and this caused difficulties in feeling it with the legs
- Participants considered the LEDs not visible in daylight, but they can be useful in this use case 5a, especially in red colour
- Participants stated the seat belt haptic feedback is useful, even if some participants consider it a little annoying because the effect was too strong
- Participants appreciated the acoustic feedback because it is useful to capture the attention of the driver.

2.4. Discussion

The purpose of the study was to evaluate the usability, the acceptance, and the perceived trust of the MEDIATOR HMI solutions (for level CM and SB) designed during the previous tasks of the project, in an ecological way on a public road with the involvement of a sample of naïve users. This was the first on-road study, in which all MEDIATOR HMI solutions had to be tested to choose the most promising ones from a user perspective and give possible design hints to ameliorate them for the second round of testing. The Italy study has given the results reported in the following.

2.4.1.1. Usability

Mediator system usability, obtained through System Usability Scale (SUS) was evaluated as good by the participants. All average detailed evaluations of the MEDIATOR HMI solutions both from comments and questionnaires received positive evaluations and, in general, there were not significant differences among them, a part LEDs, which were not considered very useful because not well visible during the daylight and perceived as probably annoying in night lighting conditions. In fact, data analysis allowed to identify also points of weakness of the HMIs, beyond the positive ones. In general, visibility and legibility aspects of the HMI had a great importance and are evaluated as crucial aspects of usability. Icons and terms used in the HMI were noticed by participants, especially, those not immediately understandable or less used according to them, proving that the used method was able to capture judgments even on HMI details.

Another important aspect was the latency time, which was perceived as long during the interaction. Moreover, participants felt the importance to know the current driving mode and its duration and to know the need for a takeover manoeuvre well in advance, so to avoid any surprise effect. The user testing had a positive outcome because it allowed to highlight users were aware of the different levels of critical situations represented by the different use cases and to evaluate the different HMI solutions designed to cope with these different situations.

In the more dangerous situations like distraction or drowsiness (even if simulated), participants appreciated the most alerting warnings (e.g., acoustic feedback and seat belt haptic feedback) and would have had even a quicker HMI escalation with lower latency times. Vice versa, in situations in which participants had to decide by themself to change the driving mode, it is important that the type of the designed HMI avoids any annoyance.

2.4.1.2. User Acceptance

The acceptability (before the trial) of automated vehicles was positive as well as the acceptability of MEDIATOR HMI. The acceptance (after the trials) evaluation was even slightly better.

Thanks to the SUaaVE questionnaire (Post et al., 2020), some significant evaluation improvements can be observed after the test with MEDIATOR HMI solutions on:
- Acceptance of automated vehicles
- Trust in automated vehicle technology

This means that, in a sample of users with positive and high expectations on automated vehicles, the MEDIATOR HMI and the experience on MEDIATOR project vehicle had a positive impact on participants, increasing the acceptance of Automated Vehicles and the trust in their technology.

2.4.1.3. Perceived trust

- Regarding perceived trust, all average evaluations were significantly positive and without any negative evaluations
- This result is perfectly aligned with evaluation improvement on trust in automated vehicle technology dimension, emerged after the trial in data collected with SUaaVE questionnaire (Post et al., 2020)

Trials impact. The duration of each test was considered adequate by participants. In general, during trials the traffic on the open road was flowing, without incidents or other issues creating delays in the procedure timing done at the beginning. The chosen route was adequate to the chosen use cases, allowing to test them properly.

The procedure used during the test was working properly without annoying or tiring participants. The thinking aloud protocol allowed to collect valuable insight on the participants' mental model of the MEDIATOR HMI and the questionnaires were simple to be administered and comprehensible for participants.

The technical characteristics of the very first version of the Wizard of Oz prototype vehicle and of the installed HMI solutions, used in the Italian test, take to some unstable behaviours, without, however, affecting the collected data that allowed to reach the test objectives. In fact, great attention was put in defining the procedure and instruction to participants so to avoid participants could be influenced by technical limits. Moreover, the same great attention was dedicated to the driving wizard instructions, so to make him “invisible” to the participants and this effort was successful because participants forgot they were not in a real automated vehicle as we understood from their behaviour (e.g. they never spoke with the driving wizard or, once ended the trial, said they liked the automated vehicle trip…).

HMI impact. The study results, thanks to the use of the thinking aloud protocol, allowed to highlight reasons behind the different evaluations given by participants. Then, following the iterative approach of the Human-Centred Design process, redesign suggestions of MEDIATOR HMI solutions were defined, so to use, in the second study in Sweden, the most appropriate HMIs properly modified to overcome usability issues of their first versions.

3. Functionality, safety effects and user acceptance in degraded driver performance conditions

This chapter describes the Sweden on-road study aimed to evaluate functionality, safety effects and user acceptance in degrade driver performance condition, with the Human Factor (HF) in-vehicle prototype.

Kyriakidis et al. (2019) outlines human factors research needs relevant for a successful deployment of automated vehicles on our roads as (i) designing HMI that can inform the driver about the vehicle's capabilities and operational status (including upcoming situations that the vehicles cannot solve), (ii) defining automation functionalities that the human drivers would accept and use, (iii) designing safe interactions between the human driver and automation during transitions of control, and (iv) designing tests to determine and ensure safety while changing from automated to manual mode. The Mediator system targets several of these challenges by intelligently assessing the strengths and weaknesses of both the driver and the automation to mediate between them.

Vehicle automation and support systems can both enhance and degrade driving safety, and to some extent, the safety benefit depends on whether the role of the driver (as assumed by the system designers) match the role drivers actually adopt. The research literature is full of examples showing that this is not the case. Automating (parts of) the driving task paradoxically induce a state of cognitive underload, causing mind-wandering (Körber et al., 2015) and fatigue (Miller et al., 2015). Drivers experienced with automation spend more time with their eyes off the forward roadway, with more frequent and longer glance durations to non-driving-related tasks (Noble et al., 2021). With more sophisticated and reliable driving automation, it becomes harder for a human driver to maintain the vigilance needed to monitor both automation and roadway (Carsten, 2019). With these unwanted side-effects fresh in mind, it becomes important to investigate and evaluate the functionality, the safety effects and the user acceptance of new systems such as Mediator.

3.1. Objectives, Research questions and covered use cases

The aim of this study was to evaluate the functionality, safety effects and user acceptance of the Mediator system under different degraded driver performance conditions (distraction and fatigue), including conditions of degraded automation.

The study was conducted in the HF in-vehicle prototype; this vehicle has basic level of ADS sophistication and relies on a Wizard of Oz-like set-up to simulate vehicle automation.
The participant was seated on the left-hand side, a driving wizard controlling the vehicle was seated on the right-hand side, and an interaction wizard (experimenter) was seated in the back. This platform is limited in the sense that the participant does not have any control of the vehicle; even if it is difficult to emulate manual driving and lower levels of automation, the main advantage

is that it is possible to run experiments with higher levels of automation that are not yet available in vehicle (for further details see section 1.5 and Fiorentino et al., 2022).

The route and the scenarios were chosen to cover as many automation levels, use cases and Mediator functionalities as possible. This includes lower levels of automation, despite the limitations with the platform. The generalisability of the results from those parts of the experiments is not known.

The following main Research Questions- (RQ) were investigated in this study:
1. How was the Mediator system generally experienced?
2. What is the effectiveness of the Mediator system in mitigating mode confusion?
3. Did the Mediator system affect distraction?
4. How are transfers of control from automation to human experienced?
5. Do active proposals for handovers to higher levels of automation result in more automation usage?
6. How do corrective countermeasures affect fatigue?
7. Is it possible to alert the driver after being disengaged from the driving task?

3.2. Methodology

The method for this study is focused on collecting data from participants to evaluate the performance of functionality, safety effects and user acceptance of the Mediator system on public roads. Ethical approval was obtained from the Swedish Ethical Review Authority (Dnr 2022-01518-01).

3.2.1. Participants

In total, 50 unique drivers were recruited via a recruitment questionnaire sent out to prospective participants contacted through a participant pool as well as a Facebook advertisement. The selected participants were 40 ± 12 years old, 18 were women, 40 had experience with advanced driving assistance systems, and they had had their driving license for 21 ± 12 years. All had driven more than 50,000 km in total, whereof 3 had driven 0–9,999 km, 19 had driven 10,000 – 19,999 km, 12 had driven 20,000 – 29,999 km, 9 had driven 30,000 – 49,999 km, and 3 had driven > 50,000 km the previous year. One group of participants (n=13) were sleep deprived and came to the experiment in the morning directly after a night shift. The remaining participants did not have any sleep restrictions. Summary statistics of the participants self-ratings of daytime sleepiness on the Epworth sleepiness scale (Johns, 1991) are provided in Table 3.1.

Table 3.1 Epworth sleepiness ratings

	All	Day group	Nightshift group
Mean ± standard deviation	6.7 ± 3.7	6.5 ± 4.0	6.8 ± 3.6
0-5 (Lower Normal)	n=17 (37%)	n=5 (42%)	n=12 (35%)
6-10 (Higher Normal)	n=23 (50%)	n=6 (50%)	n=17 (50%)
11-12 (Mild Excessive)	n=2 (4%)	n=0 (0%)	n=2 (6%)

The inclusion criteria that were strived for were:

- Equal distribution amongst gender. If this turns out to be too difficult, we will accept a slightly unequal distribution.
- Age from 20–60 years of age.
- Height below 1.90 m. Very tall people will be excluded because the cameras belonging to the eye and head tracking system are mounted in fixed positions, and too tall people cannot adjust the seat enough to be fully visible in the cameras.
- No disabilities that prevent the participant from driving an ordinary car
- No problems with motion sickness
- No glasses. This criterion is to ensure better eye tracking quality.
- Possession of a valid driving licence for passenger cars
- Minimum total mileage 50,000 km, with a minimum mileage in the past 12 months of at least 10,000 km.
- Preferably experienced with advanced driving assistance systems.
- No alcohol 72 hours before the experiment.
-

3.2.2. Procedure and design

The study had a mixed counterbalanced design. All drivers participated twice, with the Mediator system and with a baseline system (within-subject design), in balanced order. A subgroup of participants (n=13) was sleep deprived while the remaining (n=37) participants had no sleep restrictions (between-subject design). The sleep deprived drivers participated in the morning after a night shift. The MEDIATOR versus baseline conditions took place on different days, thus requiring two visits to the laboratory. Each of the driving sessions required about 3h of experiment time (about 1h of driving). Three participants took part in the experiment each day (08.00–11.00, 12:00–15:00, 15:00–18:00), where the first time slot was reserved for the group of sleep deprived participants.

At each experiment session, the following procedure was used.

1. A background questionnaire on driving experience, quality of sleep etc. was filled in before arrival.
2. At the laboratory, the participant received instructions concerning the experiment, the HF in-vehicle prototype, and the system that would be used in the experimental session (Baseline or MEDIATOR). This included an instruction video demonstrating all systems.
3. Informed consent.
4. Information concerning bank accounts for payment of the incentive (1000 SEK ~ 100€).
5. Attachment of electrodes for recording of heart rate (Lead II electrocardiogram).
6. Attachment of chest band for recording of heart rate.
7. Pre-drive 3-minute Psychomotor Vigilance Test (Loh, 2004).
8. At the vehicle, the participant was shown all HMI elements and functionalities.
9. The experiment.
10. Post-drive 3-minute Psychomotor Vigilance Test.
11. Post-drive questionnaire
12. Post-drive interview.

Similarly to the Italian study (see section 2.2.2.1), participants were instructed to pretend that they were alone in the vehicle. This meant that talking and interacting with the driving wizard or with the interaction wizard was kept to a minimum. The participants were also informed that the driving wizard in the right seat was driving the vehicle. The participants were asked to act as if this was a real automated vehicle, both when interacting with the vehicle and when answering questionnaires etc.

The route that was driven is about 82 km and takes just over 1h to drive (Figure 3.1). The route led through the outskirts of Linköping, Sweden, via a straight and wide rural road to Norsholm, and then went northeast on the motorway E4 to exit 117 before heading back to Linköping via the motorway E4. The route essentially consists of three main stretches where the first is an urban road which affords driving with Continuous Mediation, the second stretch is a rural road which affords driving in Driver Standby, and the third stretch is a motorway which affords driving in Time to Sleep mode.

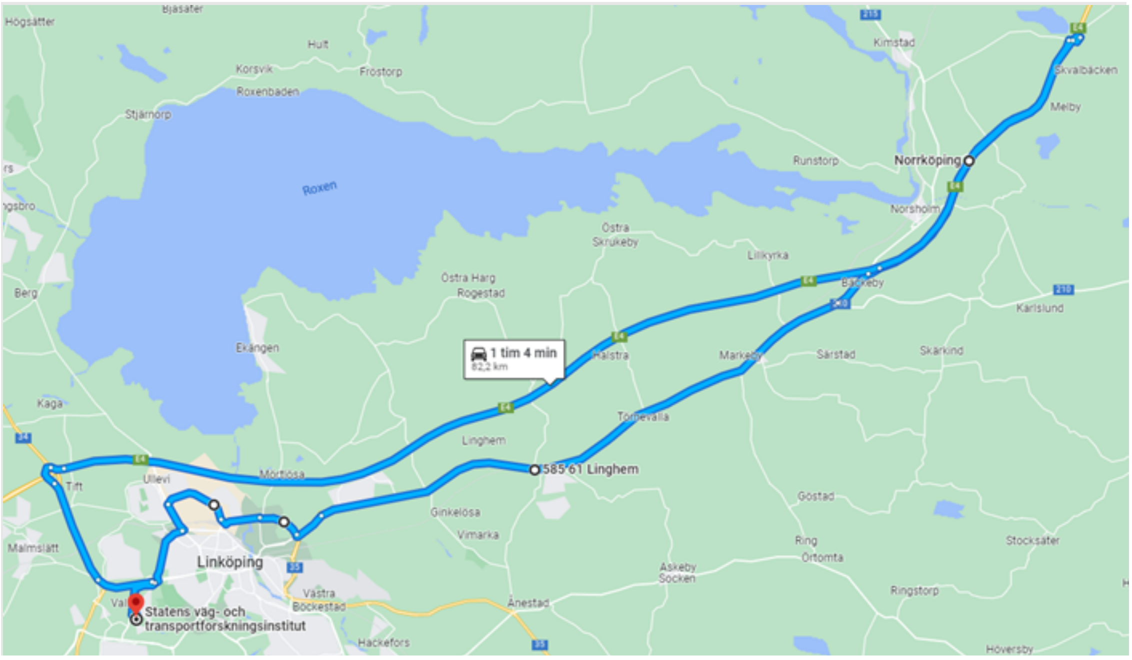


Figure 3.1 Experiment route (Source: Google maps, 2022).

3.2.3. MEDIATOR and Baseline HMI implementations

One of the key concerns in partially automated driving is mode awareness i.e., that drivers understand in what mode the automation is and display the appropriate level of responsibility in each driving mode, be it that they must constantly monitor or be able to regain control within a designated number of seconds.

The MEDIATOR HMI that was used in this study is a refined version of the HMI that was used in the HMI evaluation study described in chapter 2. The main modifications consisted of increasing the icon size, modifying the light intensity, and relocating some visual elements on the screen to make them more clearly visible.

3.2.3.1. MEDIATOR HMI

The MEDIATOR HMI strives to achieve mode awareness by continuously communicating the current as well as the upcoming automation mode. The current driving mode is indicated via a consistent colour scheme using yellow/amber for Continuous Mediation and purple/magenta for

Driver Standby and Time to Sleep. These colours are displayed on all screens but also via ambient light, and via LED strips on the dashboard and on the steering wheel, see Figure 3.2 (left). The automation mode can also be derived from a set of icons, see Figure 3.2 (right). The participant uses a gear shifter to switch between different automation modes, see Figure 3.2 (left) in the lower right corner. The gear shifter changes colour in congruence with all other HMI elements.

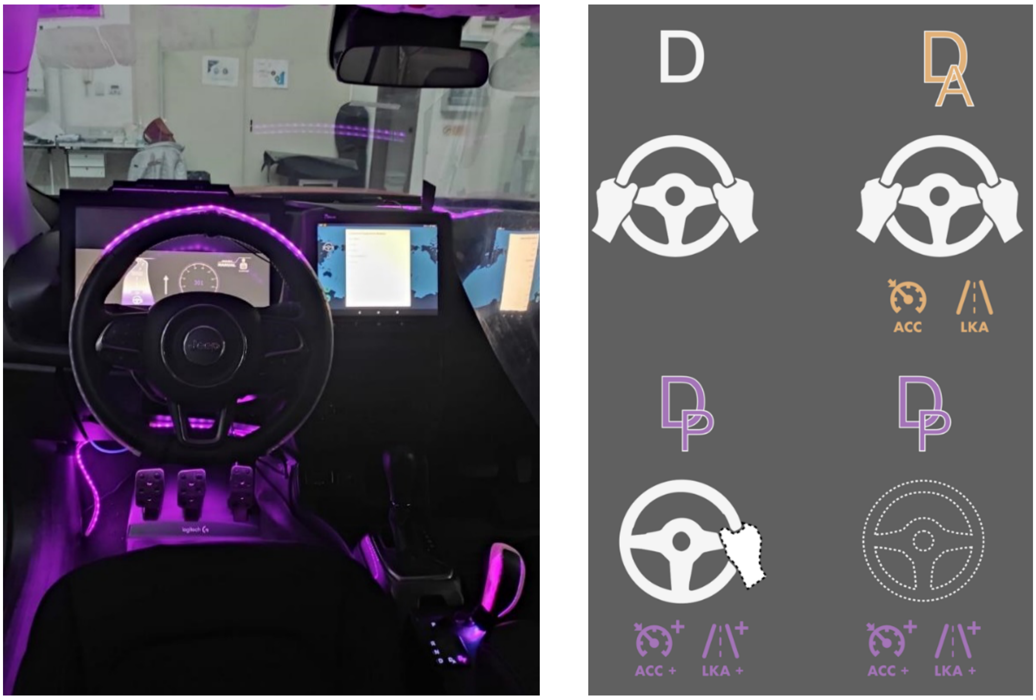


Figure 3.2 LED strips and ambient lights visualizing Driver Standby. To the left is an overview of driving mode icons: Manual (D), Continuous Mediation (Assisted driving, DA), Driver Standby (Piloted driving, DP), and Time to Sleep (Piloted driving, DP).

The current and the upcoming automation mode is conveyed to the participant via a stylized road widget (Figure 3.3). Colours, icons, and timers are used to inform the participant on the current driving mode and its time-budget (time to next) and the upcoming driving mode (available time). In manual driving, no time budget is shown, indicating that it's always available.

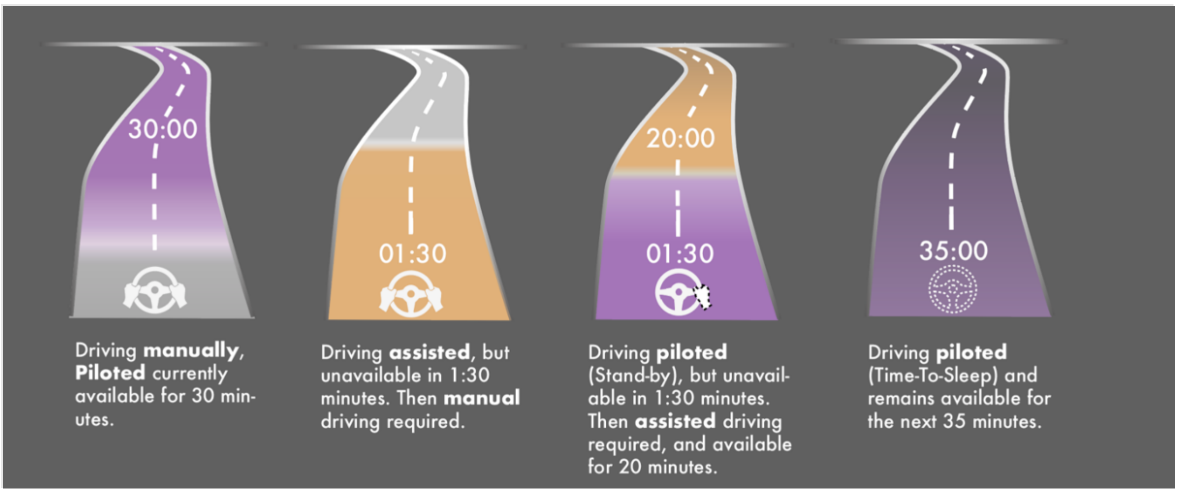


Figure 3.3 The upcoming automation mode is conveyed to the participant via a stylized road stretch providing the remaining time until the next mode change.

If fatigue is detected, a degraded fitness message is shown on the displays, a fatigue icon is lit, and an audio alert is triggered. The warning is escalated if needed, by a second notification and a stronger audio alert. If this does not help, an emergency take-over is triggered, and the LEDs starts pulsating in red colour. A similar procedure takes place if distraction is detected, but with a corresponding distraction message instead of a degraded fitness message. In cases distraction is detected and where a suitable (higher) level of automation is available, the Mediator system will suggest that the driver should switch to automated mode. This is illustrated in Figure 3.4. Note that distraction warnings are inhibited in piloted mode.

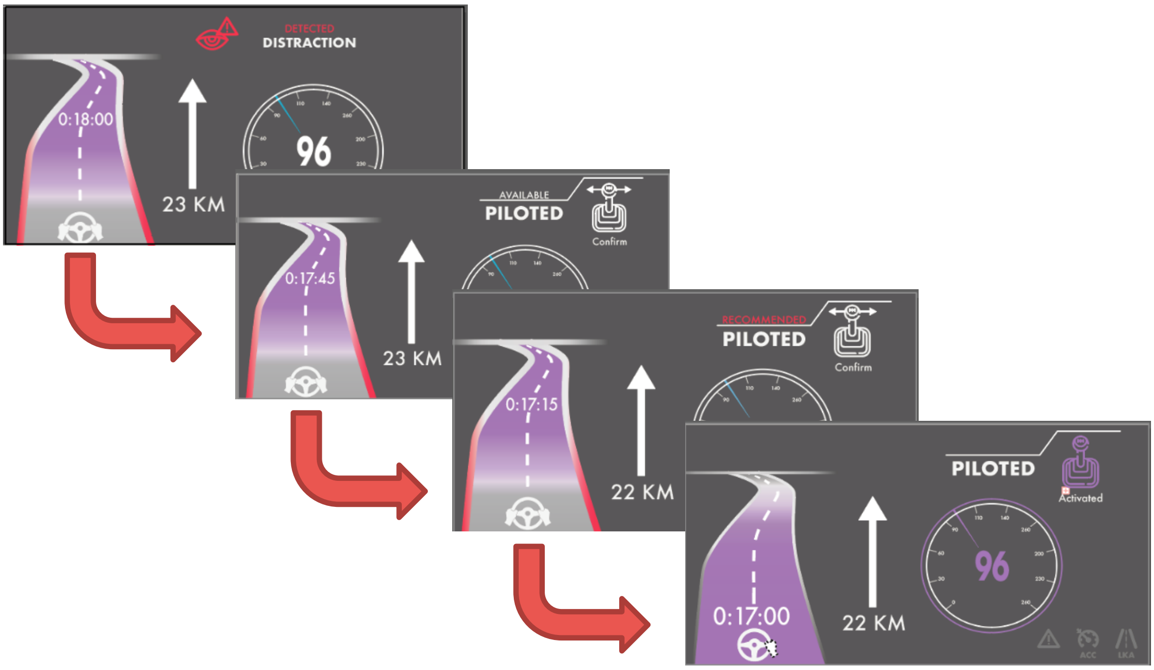


Figure 3.4 If distraction is detected and a higher level of automation is available, the Mediator system will recommend the driver to switch to automated mode.

3.2.3.2. Baseline HMI

The baseline HMI is a down-scaled version of the MEDIATOR HMI. The gear shifter is replaced by a touch screen control, the current automation mode is displayed by a simple icon (the same as in the MEDIATOR HMI), take-over routines make use of a simple sound and an accompanying icon, the distraction detection system is turned off, and the fatigue alerts consists of a coffee cup icon and a simple sound. See Figure 3.5 for examples. A general comparison between the Baseline HMI and the MEDIATOR HMI is provided in Table 3.2.

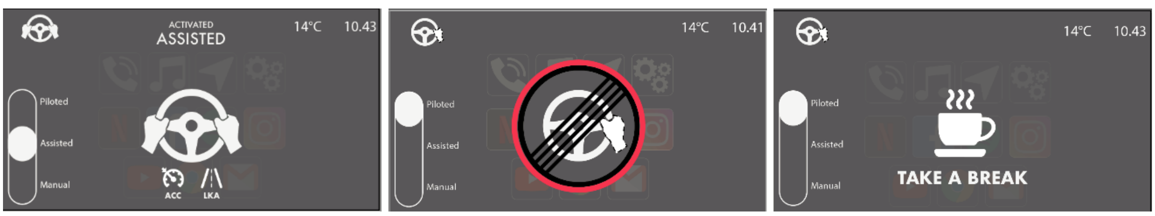
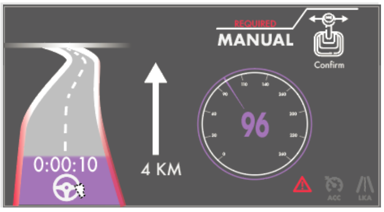
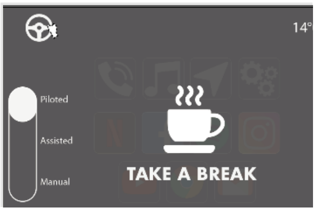
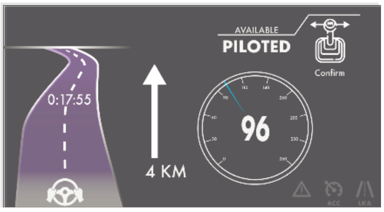
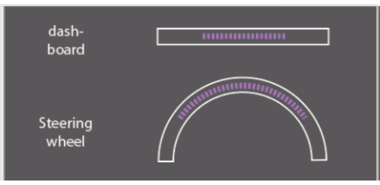
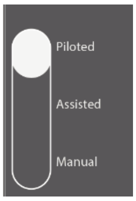


Figure 3.5 . Examples of displayed messages in the Baseline HMI. The automation mode touch control is seen to the left in all figures. Activation of assisted driving is illustrated to the left, end of driving mode in the centre, and the fatigue warning icon to the right.

Table 3.2 Comparison between the Baseline and MEDIATOR HMIs

	Event	Baseline HMI	MEDIATOR HMI
Continuous information	Upcoming driving mode	N.A.	
	Current driving mode		
Corrective actions	Distraction warnings	N.A.	
	Fatigue warnings		
Take-overs	Level up suggestion	N.A.	
	Take-over routine		

Mode change device		

3.2.4. Distraction detection and mitigation

A real-time driver distraction detection and mitigation system was implemented in the vehicle. The distraction system relies on a commercial eye tracking solution (Smart Eye AI-X, Smart Eye AB, Gothenburg, Sweden) in combination with an established distraction detection algorithm called AttenD (Ahlström et al. 2013). At the core of the algorithm is a so-called time buffer that depletes when the driver looks away from the road and fills up again when looking back to the road. If the buffer runs empty the driver is classified as distracted. Distraction detection based on the buffer approach conveniently incorporates long glances as well as visual time-sharing behaviour. This means that a driver is considered distracted both when looking away for too long with a single long glance, or when frequently looking away without sufficiently glancing back at the road in between. The size of the time buffer is 2 seconds, giving the driver 2 seconds to look away from the road before getting an alert. The driver is also granted an extra second of look away time if these glances are directed towards the instrument cluster or a mirror. If the eye tracker is unable to provide gaze direction data, the algorithm seamlessly switches over to head direction data. In this mode, the algorithm will check if the driver's head is facing forward, which is then treated as if the eyes were directed to the road. Using a buffer size of 2 seconds is loosely based on the finding that glances away from the road are rarely longer than two seconds (Horrey and Wickens, 2007), and also that glances exceeding two seconds are considered dangerous (Klauer et al., 2006).

The AttenD algorithm was modified in the MEDIATOR project to output three escalating levels of distraction warnings:

- Level 1: Triggered by the AttenD algorithm as described above.
- Level 2: Triggered after a Level 1 warning has been on for more than 15 seconds and eyes-off-road time in total is over 9 seconds.
- Level 3: Triggered after a Level 2 warning has been on more than 15 seconds and eyes-off-road time in total is over 14 seconds.

The warnings are deactivated according to:

- Deactivate Level 1: Eyes-off-road time in total is less than 9 seconds and the driver is looking at the road (AttenD buffer is increasing).
- Level 2 to Level 1: Eyes-off-road time in total is less than 14 seconds and the driver is looking at the road (AttenD buffer is increasing).

Distraction detection and mitigation is an integral part of the Mediator system. In the Baseline system, the distraction detection algorithm runs silently in the background. This means the data is logged in both conditions, but warnings are only shown to the driver in the Mediator condition.

In summary, the MEDIATOR HMI communicates a distraction event to the participant (level 1) by a visual message in the instrument cluster and in the right display (

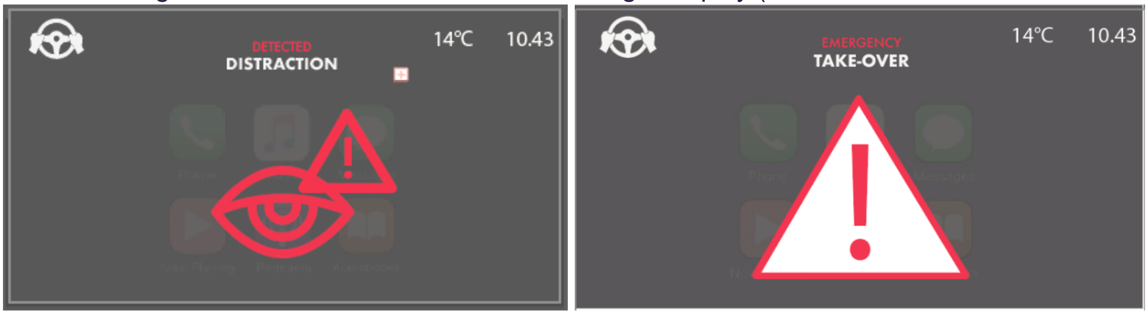


Figure 3.6). This was accompanied by an alert sound and a voice message “distraction detected”, and a gentle vibration in the seat belt. At level 2, the alert was intensified by another audio alert plus vibration, and by inflating a cushion in the seat to give the participant a more upright seating position. At level 3, an emergency takeover is initiated. Here another audio alert plus vibration is triggered, while at the same time as all lights turn red.

The distraction detection algorithm operates in real-time and thus triggers immediate alerts. However, the decision logic runs in 1Hz, which causes long delays of up to several seconds between the detection and the warning. This delay has implications on the transparency of the system, making it difficult for the participants to understand the warnings.

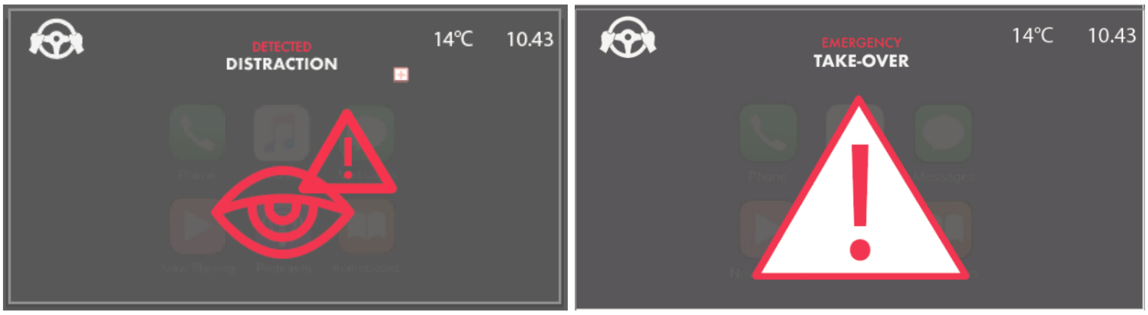


Figure 3.6 Visual warning messages shown to distracted participants at levels 1 (left) and 3 (right).

3.2.5. Fatigue detection and mitigation

The fatigue detection is based on physiology, specifically it is calculated via heart rate variability. A wearable device (Polar H10, Polar, Kempele, Finland) was used to measure the electrocardiogram from where heart rate variability parameters were extracted. The signal was transmitted via Bluetooth to a data processing device (Raspberry Pi Ltd, Cambridge, UK), where a pre-trained decision tree model predicts drivers’ Karolinska Sleepiness Scale (KSS) score every 5th minute. The physiology-based algorithm was complemented by a manual decision process based on the participants subjective sleepiness ratings. Based on this information, the actual fatigue warnings were then triggered manually by the interaction wizard in the backseat. The fatigue warnings were triggered on two levels, either as an escalation or separately. The timing of the warnings was adjusted to fit the experimental protocol, making sure that the warnings were not triggered in a way that interfered with for example a transfer of control. Fatigue warnings were active in both the MEDIATOR and Baseline conditions, with the main difference that only level 2 warnings were used in Baseline. The thresholds for the two levels below are defined based on data from Åkerstedt et al.

(2014) where it is shown that driving performance degradations as well as physiological sleepiness indicators of sleepiness starts to increase exponentially at these KSS levels.

- Level 1: Corresponding to Karolinska sleepiness ratings of 6 or 7.
- Level 2: Corresponding to Karolinska sleepiness ratings of 8 or 9.
-

If a participant fell asleep between KSS ratings this was pickup up by the interaction wizard who triggered a level 2 warning. The warning was then relayed to the driver via the decision logic system, either immediately, or when the vehicle reached a road segment that did not afford.

In summary, the MEDIATOR HMI communicated a level 1 warning by showing the message in Figure 3.7 (left). A similar message was also shown in the instrument cluster. This was accompanied by an alert sound and a voice message “fatigue detected”. A cushion in the seat was also inflated to give the participant a more upright seating position, and a gentle vibration could be felt in the seat belt. At level 2, a visual message requesting the participant to take a break while an escalated alarm sound is triggered, accompanied by the voice message “Rest required”, and another vibration in the seat belt. If the driver is still unresponsive, an emergency take-over request is triggered, with a retraction of the seat belt, an escalated warning sound, a voice emphasizing that “Rest required”, at the same time as all lights turn red.

The corresponding HMI used in the Baseline condition is a scaled down version with only level 2 warnings accompanied by an alert sound.

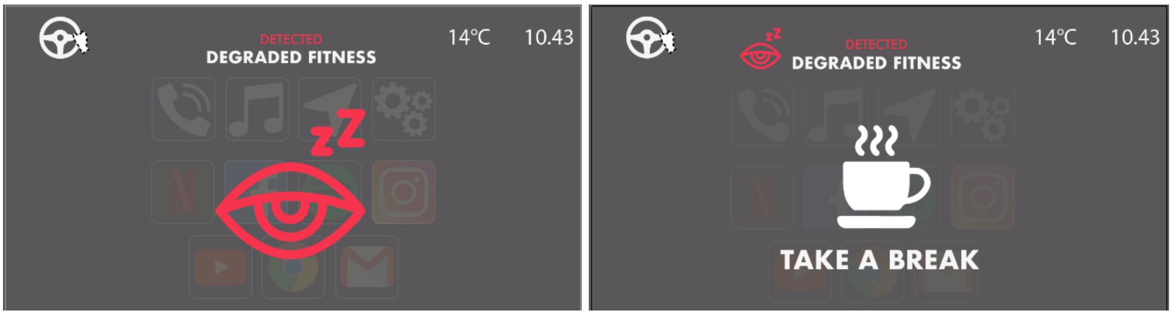


Figure 3.7 Visual warning message shown to the participant when reaching fatigue levels 1 (left) and 2 (right).

3.2.6. Non-driving related tasks

The participants were asked to perform two different non-driving related tasks in the experiments. The first task consisted of finding an arrow pointing upwards in an array of 5x5 arrows pointing left, right and downwards (Östlund et al., 2004). In 50 % of the cases an upwards-pointing arrow was present. The task was presented on a touchscreen mounted close to the centre console to the right of the steering wheel (Figure 3.8). It was self-paced in that a new task appeared when the participant had completed the previous one. The participant answered the task by touching the upwards-pointing arrow (if present) or selecting a “no”-button (if not present). The instruction was to engage frequently in the task, to the extent that felt reasonable in the present traffic situation and with the present level of automation. The idea was to assess the ability to integrate a non-driving related tasks (NDRT) with driving, and to direct the drivers’ possible spare visual capacity to the touchscreen.

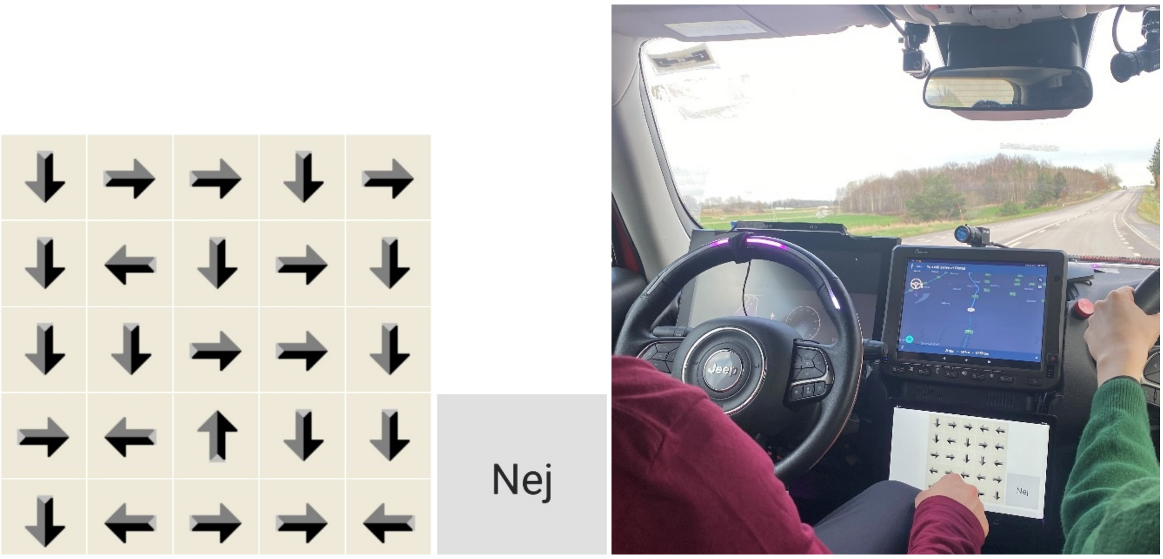


Figure 3.8 The "arrows" task used in the experiments, where the participants should determine if there is one arrow pointing upwards in the matrix. "Nej" translates to "No" in English.

The second non-driving related task was designed as a high-priority task forcing the participant glance away from the road towards the touchscreen. Six grey buttons were shown on the screen (Figure 3.9). A very annoying alarm started to sound at the same time as the six buttons became visible. The only way to turn off the alarm was to press the green button, which changed position according to a random sequence.

Figure 3.9 The high-priority task, where a loud alarm sound could only be turned off by pressing the button that turned green according to a random sequence. "Tryck på den gröna knappen" translates to "Press the green button".

3.2.7. Test protocol

In total, the route covered 10 scenarios targeting different aspect of the research questions. Some scenarios covered extensive stretches and were meant to evaluate the Mediator system in between transfers of control. Other scenarios were shorter and intended to trigger or invoke a certain transfer of control. The scenarios were controlled based on GPS coordinates, starting the scenarios at predefined locations.

3.2.7.1. Scenario 1 – Distraction detection

The participants were tasked to engage with a non-driving related activity while driving with Continuous Mediation through an urban environment. This covers the first 5 km of the route (https://goo.gl/maps/QYQYQQsaWpyuXch57). The participants could choose freely where, when and whether to engage in the non-driving related task (the task is described in section 3.2.6). The participants were encouraged to interact with the task throughout the entire 1h route, but it was especially important in this initial stretch.

Engagement with the non-driving related task will occasionally lead to long or frequent glances away from the road. When driving in the Mediator condition, this will be detected by the system and the participant will be alerted. In the Baseline condition, these warnings are inhibited. The number of distraction events can then be analysed and compared between the two systems.

3.2.7.2. Scenario 2 – Automation failure

Just before entering a roundabout (https://goo.gl/maps/synfaictsn6ZHXsHA), Continuous Mediation fails, and the participant must resume control by switching to manual driving. In the Mediator condition the participant is alerted by a visual message stating that the lane markings are degraded, accompanied with an audible beep. In the Baseline condition there is only an audible beep. In both conditions, all HMI elements switches mode from Continuous Mediation to Manual.

3.2.7.3. Scenario 3 – Recommendation to turn on Driver Standby due to distraction

While “driving” in Manual mode in the final urban stretch along the route (https://goo.gl/maps/ckTbANtXJ84LpCxN9), a high priority non-driving related activity was initiated (see section 3.2.6). This was detected by the distraction detection algorithm roughly at the same time as automated driving in Driver Standby mode became available. The Mediator system then notified, or recommended, the participant to switch to the Driver Standby to be able to pursue the high priority task. In the baseline condition, where distraction detection was unavailable, the high-priority non-driving related task was not tiggered, and the interaction wizard simply asked the participant to make the switch to Driver Standby when it became available, unless the participant did so by him/herself.

3.2.7.4. Scenario 4 – Fatigue detection

The participants drove on a rural road stretch for about 15 minutes in Driver standby (https://goo.gl/maps/uiBtLYFSxyLMFDsu9). If the driver became fatigued, which was quite probable in the sleep deprived group, the Mediator system attempted to alert the driver with escalating corrective countermeasures. Distraction warnings are disengaged since it is allowed to disengage from the driving task for short periods of time.

3.2.7.5. Scenario 5 – Transfer of control from Driver Standby to Continuous Mediation

In the end of the rural road stretch from Scenario 4 (https://goo.gl/maps/x2rkdLM4Hzz355vo7), automation reaches the boundary of its operational design domain. The Mediator system, as well as the Baseline equivalence, communicates this to the participant by requesting a mode switch from Driver Standby to Continuous Mediation.

3.2.7.6. Scenario 6 – Transfer of control from Continuous Mediation to Driver Standby

When entering the motorway (https://goo.gl/maps/DbLXHLFYS8MGmTAC7), Driver Standby becomes available. The Mediator system, and the Baseline equivalence, communicates this to the participant by suggesting a mode switch from Continuous Mediation to Driver Standby.

3.2.7.7. Scenario 7 – Fatigue detection

Driving on the motorway (https://goo.gl/maps/NUyCQ6rBbr2JNCEq9), the participants received (escalating) fatigue warnings if they became fatigued. This is identical to Scenario 4 described in section 3.2.7.4.

3.2.7.8. Scenario 8 – Driving in Time to Sleep mode

After turning around at exit 112 and when heading back towards Linköping and the starting point, the system automatically switches driving mode from Driver standby to Time to Sleep. The participant is then allowed to disengage from the driving task and do other things, including interacting with their own phone and sleeping. Since it is allowed to disengage and even sleep on this road stretch, both fatigue and distraction warnings are disengaged. The scenario lasts for about 20 minutes (https://goo.gl/maps/fBYifbHfguV8cY5x5).

3.2.7.9. Scenario 9 – Taking back control after a longer time in Time to Sleep mode

The Mediator system, and the Baseline equivalent, alerts the participant before the vehicle leaves the operational design domain of Time to Sleep (https://goo.gl/maps/TAGDs5Kx1msDF7U26). Note that the automation mode change is dome automatically and without confirmation from the participant. If the participant is fatigued, a message is shown requesting the participant to take a break. If this happens, the interaction wizard will intervene and ask the participant to proceed with the final scenario despite being too fatigued to continue in Driver Standby.

3.2.7.10. Scenario 10 – Fatigue detection

The last scenario is identical to Scenario 7 and lasts for about 5 minutes (https://goo.gl/maps/ZWy72zGDY2J9xNpGA).

3.2.8. Measurements

3.2.8.1. Vehicle data, automation state and HMI interactions

See sections 2.2.3.1 and 2.2.3.4.

3.2.8.2. Physiological data

An electrocardiogram was recorded with a Vitaport 3 bio-amplifier (Temec Instruments BV, the Netherlands). Time synchronization was done via analogue synchronization pulses sent out by the Autoliv Raspberry Pi. The Raspberry Pi computer also stored heart rate and interbeat interval data from the Autoliv steering wheel, and from a chest band (Polar H10).

3.2.8.3. Eye tracking data

The drivers' eye movements will be tracked by a Smart Eye system. The eye tracking data, as agreed in the project, were used by the Mediator system to evaluate driver distraction, but the eye movement data were not stored.

3.2.8.4. Video recordings

Sequences of time stamped compressed images were stored in the data logger. Four cameras filmed the drivers' face, the upper body, the forward roadway, and an over the shoulder view of the forward roadway.

3.2.8.5. Subjective ratings

Every fifth minute throughout each drive, participants rated their sleepiness on the nine-point Karolinska Sleepiness Scale (KSS): (1) extremely alert (2) very alert (3) alert (4) rather alert (5) neither alert nor sleepy (6) some signs of sleepiness (7) sleepy, no effort to stay awake (8) sleepy, some effort to stay awake; and (9) very sleepy, great effort to keep awake, fighting sleep (Åkerstedt and Gillberg, 1990). The participants were prompted to enter their rating on a tablet in the centre console, see Figure 3.10. The rating scale was sent home to the participants for them to practise before the experiment. Onsite, we also reviewed the scale together with the participants to make sure that they knew it by heart.

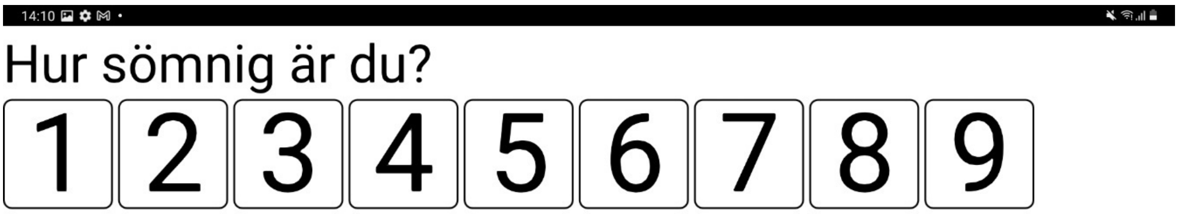


Figure 3.10 The display used to self-rate subjective sleepiness every fifth minute. "Hur sömnig är du?" translates to "How sleepy are you?".

3.2.8.6. Vigilance test

The participants performed a simple 3-minute reaction time test before and after each drive (Loh et al., 2004). The test randomly shows a digit on a screen and the participant must respond as fast as possible. The test was performed in a quiet room without any influence from others.

3.2.8.7. Questionnaires

Three different set of questionnaires were used in the study; a recruitment questionnaire, a background questionnaire, and a questionnaire answered after driving the two conditions (with the Baseline and MEDIATOR HMIs, respectively). The questionnaires were based on the standard questionnaires listed in section 2.2.3.6.

The recruitment questionnaire collected data on age, gender, height, weight, shift work, known sleep disorders, motion sickness, driving experience, sleep habits, and contact details.

The background questionnaire collected data on demographics, experience with automated systems, and attitude towards and experience of non-driving related activities and driver monitoring systems.

The post-drive questionnaires collected data on experienced motion sickness symptoms, experienced fatigue symptoms, and about the system under investigation. The latter included questions on acceptance, usefulness, trust, safety, workload, and similar entities, in relation to the system as a whole, as well as to specific sub-functions such as take-over requests and driver state warnings.

3.2.8.8. Interview

An interview was conducted after the drives. The purpose of the interview was to complement the questionnaires with more information on why the participants appreciated or disliked the system. The main questions asked during the interviews were:

1. How did you experience the drive?
2. What was your general experience with the information system?
3. Did you trust the system, given that automation works as intended?
4. The vehicle involves three automation modes: manual, assisted, and piloted. What was your experience with these modes?
5. How did you use the various elements of the information system?
6. Please have a look at these images [of HMI elements indicating distraction]. Can you describe the situation when you received these notifications? E.g., what were you doing, what happened inside and outside the car?
7. Please have a look at these images [of HMI elements indicating fatigue]. Can you describe the situation when you received these notifications? E.g., what were you doing, what happened inside and outside the car?
8. Please have a look at these images [of HMI elements indicating switching to a higher level of automation due to distraction]. Can you describe the situation when you received these notifications? E.g., what were you doing, what happened inside and outside the car?
9. Please have a look at these images [of HMI elements indicating switching down to a lower level of automation due to operational design domain constraints]. Can you describe the situation when you received these notifications? E.g., what were you doing, what happened inside and outside the car?
10. Would you use the information system in your future car?
11. After two test days you have now experienced two information systems [the Baseline HMI and the MEDIATOR HMI]. Which one do you prefer and why?

3.2.9. Data pre-processing and statistical analysis

Data were collected from 50 participants. The following inclusion criteria were followed to select participants for a specific analysis:

- Data should be available for all conditions present in the analyses. The active proposal of automation and the distraction warning were only present in the Mediator condition and for these analyses therefore only data for the Mediator condition were needed.
- The participant should have experienced the relevant scenarios for a particular analysis without experimental failures. For the analyses regarding general experiences with the system a threshold was set to at least half of the scenarios during an experimental trial to have been experienced without experimental failures.

Data loss occurred due to errors in real-time in-vehicle logging (timeseries data), but also due to missing data in the questionnaire responses. More information about how much data were available in a specific analysis is provided in the results section. To determine if system failures were present during certain scenarios, the experimenter logs were examined.

3.2.9.1. Main effects and covariates

The main effect of interest in this chapter is the difference between the MEDIATOR and the Baseline conditions. The effects of the interaction between condition and age were also analysed, by fitting regression models to the data. The exact models and tests used to determine significance of any effects found depends on the data types used and is described at the start of each sub-section in the results section 3.3. The remaining data for each analysis were checked for potential order effects that could affect the results presented in this chapter. No significant order effects were found for datasets that were not sufficiently balanced and these results are therefore left out for further reporting.

3.2.9.2. Interview analyses

The recordings were analysed by listening to all interviews while making notes about what the participants thought about the systems and the test itself. The notes covered aspects that the participants had a clear opinion about but did not cover all answers to all questions in the interview guide. The notes were then encoded, categorised, and summarized based on the predefined research questions. The number of participants expressing the same opinion were counted but note that these counts are based on the encoded material, which in turn is based on the notes of clear opinions (rather than on all answers).

Interview data form 45 participants were available from both the MEDIATOR and Baseline conditions, two participants were only interviewed after the Baseline condition (since there were severe technical issues during the Mediator drive), two participants were only interviewed after the Mediator condition (for the same reason), and one participant had no interview data for unclear reasons.

3.3. Results

Results are reported below for each of the predefined Research Questions (RQ).

3.3.1. RQ1. How was the Mediator system generally experienced?

Three types of data were used to answer how the Mediator system was experienced: interview data per condition, questionnaire data per condition, and questionnaire data at the end of the experiment. The latter regards a questionnaire where preferences for either condition are compared directly. As the data were not normally distributed, a two-sided paired Wilcoxon test, using the *wilcox.test* function in the R package stats, was used to compare medians between conditions. For a more in-depth analyses of possible interaction and random effects, regression models with random effects were used to test interaction effects between condition and age and condition and order effects.

The models were fitted to the dataset using the *lmer* function in the R package lme4 and the formula. Significance of the interactions were tested with a type III ANOVA using the *anova* function in the R package lmerTest. The categorical data on preference for either the Baseline or MEDIATOR or being undecided were instead fitted with a multinomial regression model using the function *multinom* from the R package nnet. The results for the type II ANOVA, obtained with the *anova* function in the R package lmerTest,

3.3.1.1. Is the HMI generally accepted?

To answer this question the results from questions on the Van der Laan acceptance scale were used. Scores regarding the constructs "usefulness" and "satisfying" were calculated (Van der Laan et al., 1997) on a scale from -2 to 2. Data from 41 participants were available. Both HMI designs were on average judged moderately useful (*Mdn*=0.8, *IQR*=0.6) and satisfying (*Mdn*=0.8 and *IQR*=0.7) and no overall significant difference was found.

In the covariate analyses a significant effect of condition ($p<0.001$) and of interaction between condition and age ($p<0.001$) was found. The model estimates showed that generally the Mediator condition was deemed more acceptable (Estimate condition Mediator = 0.67) and that this positive acceptability difference between conditions is reduced with age (Estimate condition Mediator: Age = -0.02).

Based on responses from the interviews, the participants found the Piloted mode to be useful when driving long distances and for daily commutes (Mediator n=16, Baseline n=17). Especially, they appreciated the opportunity to be relieved in predictable situations and on boring motorways, but also in situations with a lot of traffic, or when you are tired or distracted. The participants supported having several automation levels in the vehicle (Mediator n=10, Baseline n=14), as different levels are suitable in different (more) situations. However, several participants (n=10) expressed concerns about having several levels if people are not aware of what each level means. There is a risk that you mix up the levels and that you miss that you are responsible for something. Assisted was found to be a good complement when Pilot isn't available. In general, the participants thought that the vehicle selected an appropriate level of automation at appropriate times during the journey.

The developed MEDIATOR and Baseline prototypes were generally perceived as crude and immature, and the participants said they would have felt more at ease if the vehicle acted as one consolidated system (rather than several separate subsystems) and if the HMI components had been better integrated with the interior of the vehicle. For the Mediator system, several participants (n=7) appreciated the LED lights, especially in the steering wheel, while others (n=2) found all the lights to be annoying and distracting[4]. They were not fond of having information scattered on several displays (n=6) and would have preferred to have all information present in the instrument cluster (n=6), to have all information more centralised. Warnings and messages on the right screen were easily missed. For the Baseline HMI, the system was perceived as clear and modern (n=16), but again, some participants wanted all information to be shown in the instrument cluster rather than on several screens (n=5), mentioning that it messages on the right screen were easily missed.

3.3.1.2. Was the system usable?

4 Respect to the Italian experiment in which the subjects were less positive about LEDs, the changes to the HMI for the Swedish experiments (icon size, light intensity, location of some visual elements on the screens) where well evaluated.

To answer this question the system usability scale was used and a single construct score for “usability” was calculated (Brooke, 1996) on a scale from 0 to 100. Data from 41 participants were available. Both HMI designs were on average judged to be usable (*Mdn*=73, *IQR*=24) and no overall significant difference was found.

In the covariate analyses a significant effect of condition (p<0.001) and interaction between condition and age (p<0.001) was found. The model estimates showed that generally the Mediator condition was deemed more usable (Estimate condition Mediator = 23.4), but that the higher usability score for the Mediator system decreases with age (Estimate condition Mediator: Age = -0.63).

3.3.1.3. Was the system trusted?

To answer this question the trust in automated systems questionnaire was used and a single construct score for “trust” was calculated (Jian, 2000) on a scale from 1 to 7. Data from 41 participants were available. Both HMI designs were on average judged trustworthy (*Mdn*=5, *IQR*=1) and no overall significant difference was found.

In the covariate analyses only a significant effect of Age (p<0.05) was found. The model showed that trust increases with Age (estimate Age = 0.03).

The interviews revealed similar insights about automation and trust regardless of the system being evaluated. Some participants trusted the system and found it reliable and safe, feeling confident that the system would detect and cope with the traffic situation (Mediator n=17, Baseline n=6). The participants believed that their trust in the system would increase if they had had the opportunity to test it for a longer time (Mediator n=11, Baseline n=16), to gain a better understanding of the system’s capabilities. They also mentioned that they would have trusted the system more if the prototype had felt more mature (Mediator n=8, Baseline n=11). In general, automation was not trusted in urban environments (Mediator n=12, Baseline n=16), on rural roads (Mediator n=3, Baseline n=4), in bad weather (Mediator n=4, Baseline n=7), or when unplanned events such as road works, police interventions, wildlife, potholes and in the presence of vulnerable road users (Mediator n=7, Baseline n=8). In general, for both MEDIATOR and Baseline, the participants described it as hard to fully leave control to the vehicle (n=23), some mentioning that it was comfortable but unfamiliar, that they could never relax completely, that they would have preferred to be in control themselves, and that it should not be allowed to let go control completely. Others trusted automation more than human drivers (n=2).

3.3.1.4. Did the system increase workload?

To answer this question the raw NASA TLX was used and a single construct score for “workload” was calculated as the mean over all questions on a scale from 0 to 100 (Said et al., 2020). Data from 23 participants were available. Both HMI designs were on average judged to have a low workload (*Mdn*=20, *IQR*=26) and no overall significant difference was found.

In the covariate analyses a significant interaction effect between Age and Condition was found (p<0.05). The model showed that the workload for the Mediator system, as compared to the Baseline condition, increased with age (Estimate condition Mediator: Age = 0.90).

Only a few participants reflected on workload induced by the system during the interviews. In the Baseline condition, two participants said they were distracted by the system and that it was difficult to keep track of all information. Three participants believed there was too much information in the Mediator condition and suggested that communication between the vehicle and the driver should be simplified. Three participants found the distraction warnings to be stressful.

3.3.1.5. Does the driver imagine buying the system when it comes to market?

To answer this question two items from questionnaires used in the L3Pilot project (Metz et al., 2019) were used, regarding willingness to buy and willingness to recommend the system, which were rated on a scale from 1 to 5. Data from 41 participants were available. Participants rated the willingness to buy and the willingness to recommend either HMI design relatively high (*Mdn* buy/recommend=4, *IQR* buy/recommend=1) and no overall significant difference was found.

In the covariate analyses a significant effect of Condition ($p<0.01$) and of the interaction between Condition and Age ($p<0.01$) was found. The model showed that the willingness to buy score for the Mediator system is generally higher than that of the Baseline system (estimate condition Mediator = 0.80) but that the willingness to buy the Mediator system, as compared to the Baseline system, reduces with age (Estimate condition Mediator: Age = -0.02).

In the interviews, many participants mentioned that they would be willing to buy the system (Mediator n=17, Baseline n=7), given that it was working as envisioned and that it was not too expensive.

3.3.1.6. Which system was preferred?

At the end of the experiment, when the participant had experienced both the Baseline and the Mediator condition, they were asked to indicate their preference for either condition regarding willingness to buy (n=34), comfort (n=38), how easy the system is to learn and use (n=39), safety (n=30) and trust (n=29) in the system. In Figure 1.1 the number of participants answering MEDIATOR, Baseline or Undecided is shown per item.

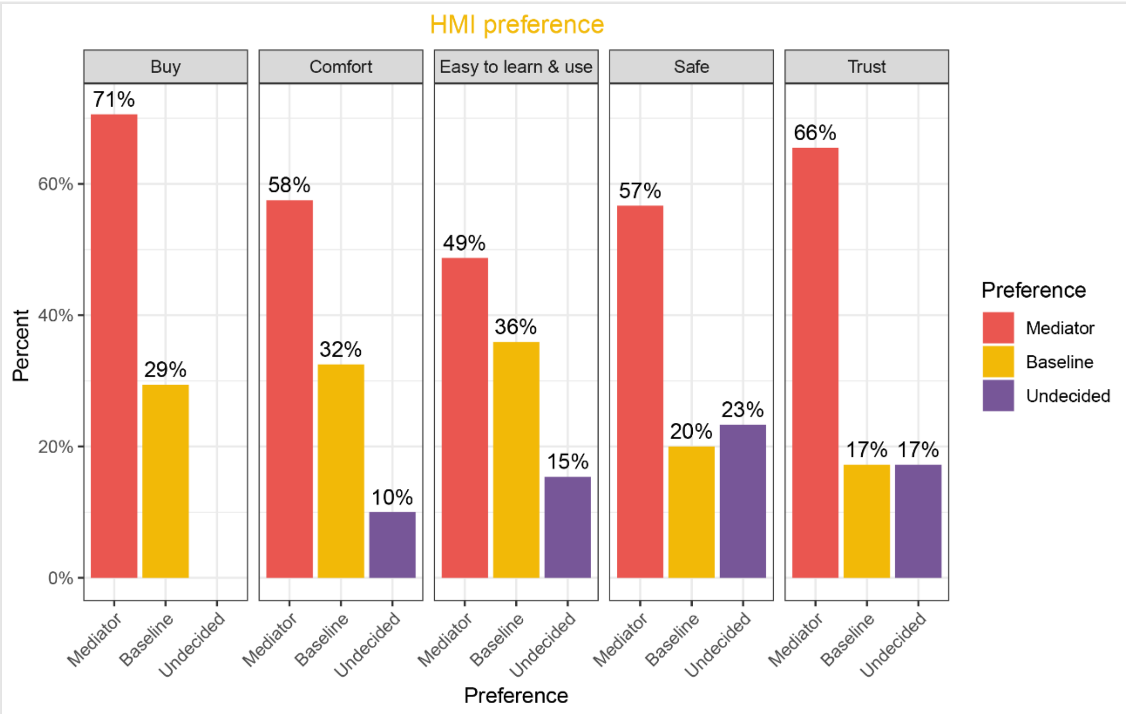


Figure 3.11 . Participant preference regarding willingness to buy (buy), comfort of the system (comfort), how easy it is to learn and use the system (easy to learn & use) and if they trusted the system (trust).

Figure 3.11 shows that most participants preferred the Mediator system on all items. 71% of the participants would prefer to buy the MEDIATOR over the Baseline system. 58% found the Mediator system more comfortable than the Baseline system, while only 32% thought the Baseline system was more comfortable and about 10 % of the participants were undecided. 49% of the participants thought the Mediator system was easier to learn and use than the Baseline system, while only 36% reported the opposite and 15% was undecided. 57% thought the Mediator system was safer than the Baseline system while 20% thought the opposite and another 23% was undecided. Finally, 66% of the participants trusted the Mediator system over the Baseline system, while 17% reported the opposite and 17% was undecided.

The covariate analyses showed a significant effect of Age ($p<0.05$). The model estimates indicate a correlation between increasing age and preferring either the Baseline system or being undecided as compared to preferring the Mediator system (estimate Baseline: Age = 0.03, estimate Undecided: Age = 0.04).

3.3.1.7. Summary

No clear difference between the Baseline and the MEDIATOR HMI was found for acceptance, usability, trust and willingness to buy when asking about these constructs right after each trial. No overall significant difference between the conditions was found and both HMI's were rated moderately useful (*Mdn*=0.8, *IQR*=0.6 on scale from -2 to 2) and satisfying (*Mdn*=0.8 and *IQR*=0.7 on scale from -2 to 2), usable (*Mdn*=73, *IQR*=24 on scale from 0 to 100), trustworthy (*Mdn*=5, *IQR*=1 on a scale from 1 to 7) and as buyable and recommendable (both items: *Mdn* =4, *IQR*=1 on a scale from 1 to 5). When asking the participant on their preference in terms of wanting to buy the systems, their comfort, how easy the systems were to learn and use, and how safe and trustworthy the systems were in a direct comparison at the end of the experiment, a clear majority of the participants preferred the Mediator system. From the interview questions more people mentioned trusting the Mediator system than the Baseline system (Mediator *n*=17, Baseline *n*=6). Also, more participants mentioned that, given that it was working as envisioned and that it was not too expensive, they would be willing to buy the Mediator system than the Baseline system (Mediator *n*=17, Baseline n=7).

There was no overall difference in workload between the conditions (both conditions: Mdn=20, *IQR*= 26), implying that the added features in the MEDIATOR HMI design did not increase workload in general.

Interaction effects between Condition and Age were found for acceptability, usability and willingness to buy and willingness to recommend, indicating that younger participants generally preferred the Mediator system, while elderly participants preferred the Baseline condition. For workload a similar effect was found, where increasing age corresponded to a higher workload for the Mediator system. The interviews further showed that the developed MEDIATOR and Baseline prototypes were generally perceived as crude and immature, and they would have preferred if both systems were better integrated into the vehicle HMI. They also wanted all information gathered in one screen rather than scattered over several screens. Warnings and messages on the right screen were, for example, easily missed and participants mentioned that they would have trusted the system more if the prototype had felt more mature.

Generally, however, the results show that the MEDIATOR HMI was preferred over the Baseline HMI, but that increasing age reduces this preference. One possible reason for this decrease is the increase in perceived workload.

3.3.2. RQ2. What is the effectiveness of the Mediator system in mitigating mode confusion?

The concepts of Mode Awareness and Mode Confusion can be operationalized as the extent to which responsibilities and affordances within different automation modes are understood by individuals. To quantify this understanding, data were collected through questionnaires and interviews.

The questionnaires contained questions where participants were asked to rate their understanding of their responsibilities *during* each drive on a four 5-point Likert-scale (from strongly disagree to strongly agree) for both system types. Data from 41 participants were available. As the data was not normally distributed, a two-sided paired Wilcoxon test, using the *wilcox.test* function in the R package stats, was used to compare medians between conditions. No interaction effects between Condition and Age were found and this part of the analysis is therefore left out in the reporting for clarity.

The questionnaires also provided information on the participants actual mode awareness in the different automation modes. Participants were asked to judge whether a number of driving subtasks were *mandatory* depending on the automation mode, and whether a number of NDRTs were *allowed*. Table 3.3 shows which subtasks were mandatory or allowed (1) or not (0) per automation mode.

Table 3.3 Mandatory and allowed subtasks per automation mode. M = Manual. CM = Continuous Mediation. SB = Standby. TtS = Time-to-Sleep.

Mandatory or allowed driving subtasksColumn title	M	CM	SB	TtS
Mandatory tasks				
Vehicle control				
• controlling the vehicle	1	0	0	0
• Being ready to control anytime	-	1	0	0
• Being ready in a few seconds	-	-	1	0
Keep eyes on road	1	1	0	0
Monitor the information system	1	1	1	0
Monitor how vehicle behaves in traffic	1	1	0	0
Allowed tasks				
Reading a book	0	0	1	1
Answering emails on my laptop	0	0	1	1
Sleeping	0	0	0	1
Adjusting general settings of the car	1	1	1	1
Making a handsfree phone call	1	1	1	1
Adjusting seat to sleeping position	0	0	0	1

For each participant their answers were summarized as correct or incorrect per automation mode (i.e., Manual, Continuous Mediation, Standby, or Time-to-Sleep) per question (i.e., mandatory or allowed) for each trial (MEDIATOR or Baseline). Data from 45 participants were available and used to fit a logistic regression model using the *logit* function from the R package stats. No significant interaction effects between Age and Condition were found and therefore left out of the further reporting. Significance of the interactions were tested with a type III ANOVA using the *anova* function in the R package lmerTest. The interview questions regarding mode awareness focused mainly on the usage and understanding of the different HMI elements in obtaining mode awareness.

3.3.2.1. Perceived mode awareness during drive

Self-assessments indicate that participants tended to be confident that they were aware of their responsibilities when using the Mediator system (*Mdn*>=3), but the difference was only significant for the third question ($p<0.05$). The items and median values of the responses for the two systems are provided in Table 3.4.

Table 3.4 Questionnaire items used to evaluate perceived mode awareness.

Question	Baseline (*Mdn*)	MEDIATOR (*Mdn*)
I was continuously aware of which parts of the driving tasks I was responsible for during the trip.	4	4
I was only aware of which parts of the driving tasks I was responsible for at certain moments during the trip.	3	3

I was continuously aware of the time until my responsibilities for parts of the driving task would change.	2	3
I was only aware of the time until my responsibilities for parts of the driving task would change at certain moments during the trip.	2	3

3.3.2.2. Understanding of automation modes

Table 3.4 was used to determine the number of correct responses per participant per automation mode, condition and question. There were no main or interaction effects of Condition. Significant effects were however found for Mode ($p<0.001$) and for Question ($p<0.001$), where the latter indicates that some questions were easier than others. Consideration of the coefficients for automation mode showed that, relative to the manual mode, the probability of a correct response was considerably lower in the Standby mode ($p<0.001$) and in the Time-to-Sleep mode ($p<0.001$), indicating that participants were unsure about their responsibilities particularly in these automation modes

3.3.2.3. Usage and understanding of HMI elements

The interviews showed that participants supported having several automation modes in the vehicle (Mediator *n*=10, Baseline *n*=14), with the motivation that different modes are suitable in different (more) situations.

The most commonly used HMI elements to understand which automation mode is currently active are the steering wheel icon on the right screen (*n*=19), the steering wheel icon in the instrument cluster (*n*=14), the D_A/D_P icons (*n*=16), the LED-strip in the steering wheel (*n*=12) and the stylized road stretch (*n*=12). In the Baseline condition, the participants recognised the automation mode via the ACC/LKA icons (*n*=27), the steering wheel icon (*n*=24) and via warning/event messages (*n*=11).

The steering wheel icons, as well as the D_A/D_P icons and ACC/LKA icons, were appreciated because icons are familiar and represent a specific/unique meaning. There were however desires/requests that the icons should be larger and more intuitive, with more discernible differences between the automation modes (i.e., a larger difference than a subscript letter, the presence/absence of a (dashed) hand, or the presence/absence of a plus-sign). The LED-strip in the steering wheel was appreciated since it effectively communicates with the driver without requiring glances away from the road, but the message conveyed by the LEDs was sometimes perceived as abstract and difficult to understand (one must learn what the colours mean). The stylized road was appreciated since it did not only convey the current automation mode but also the time to the next shift in responsibility. However, while many liked the information rich representation, some did not notice it at all, and others did not understand it.

For the MEDIATOR HMI, the participants thought that the colour schemes contributed to the clarity of the system (*n*=18), making it easier to keep track of the current automation mode. A few participants (*n*=3) said that they did not understand the colours, that they would have preferred more contrasting colours, and especially that red and purple was difficult to see against the black background.

A few participants (*n*=3) said it was easier to understand the gear shifter used in the Mediator condition, but that it could just as well have been a physical button on the steering wheel. Many (*n*=10) preferred to change automation mode on the touch screen (as in the Baseline condition) since it was easier, more pleasant, comfortable, and aesthetically pleasing. They also appreciated that the touch screen control was located higher up than the gear lever, and that they didn't have to look down as much.

3.3.2.4. Summary

Generally, drivers indicated that they understand their responsibilities during the drive. Drivers were more likely to perceive to be continuously aware of the time until a change in responsibility would occur in the MEDIATOR than in the Baseline condition. This can be explained by the information that was provided in the Mediator condition on the time budget left within the current automation mode. This difference therefore shows that this information was indeed used by the driver to estimate the time until a change in responsibility would occur. The interviews also indicated that time budget information, as communicated via the LED strips and road icon, was indeed used by participants to determine time until the next responsibility shift.

However, the MEDIATOR HMI did not change the actual understanding of the driver responsibilities and affordances in the different automation modes. For both systems, the responsibilities in driving modes Standby and Time to Sleep were harder to understand than those in the Continuous Mediation mode. Possibly because the Continuous Mediation mode is similar to cars already driving on the road.

The interview results further showed the preference of users to receive information through icons and interact through a touchscreen, which are both communication channels they are familiar with.

3.3.3. RQ3. Did the Mediator system affect distraction?

Several aspects of the different HMIs could have affected the distraction observed during the experiment. For example, the level of mode awareness and knowledge on the time until the next automation mode could have affected driver choices regarding distracting activities. During the Mediator condition only, a corrective action was initiated when distraction was detected. To assess the effect of the Mediator system on distraction several data sources were analysed.

Questionnaire data from questions regarding the distraction warning together with interview data were used to assess the understanding and appreciation of these signals. Data from 48 participants were available for the questionnaire analysis where 4 questions were answered on a 5-point Likert scale from strongly disagree to strongly agree. As the data were not normally distributed, a one sample two-sided Wilcoxon test (using *wilcox.test* from the R package stats) was used to compare the median over all participants to the “3 – neutral” answer on the Likert scale. No significant interactions between Condition and Age were found and this analysis is therefore left out of further reporting.

Additionally, timeseries data to estimate distraction were used to estimate the observed effect of Condition (Baseline vs MEDIATOR) on distraction. Data from scenario 1, where participants drove in Continuous Mediation mode for about 7 minutes and were thus expected to monitor the automation and not be distracted, were used for this analysis. Data from 27 participants were available. A total of 24 of these participants (89%) received at least one distraction warning while driving with the MEDIATOR HMI during scenario 1. Distraction was operationalized with the AttenD score measure. As explained in section 3.2.4 AttenD involves a 2 second time buffer that depletes when the driver looks away from the road and fills up again when looking back to the road. Only when the buffer is empty, the driver is deemed “distracted”. The resulting binary signal is used to calculate four distraction related measures per participant per condition.

The first measure is the *distraction uptake frequency*, which was calculated as the number of non-interrupted distraction events divided by the time in scenario 1. The second measure is the *proportion of the time distracted*, which was calculated as percentage of the time in scenario 1 the participant was distracted. The third and fourth measures related to the duration of the distraction events. Visual inspection of histograms of the duration of non-interrupted distraction events showed that the distribution of these durations within participants were often skewed (with short durations occurring more often than long durations). Therefore, median duration was chosen as one measure of distraction. Due to the significance of long duration distraction events for road safety, also the maximum duration was analysed. Therefore, the third measure was the *maximum distraction duration*, which was calculated as the longest non-interrupted distraction event in scenario 1. And finally, the fourth measure was the *median distraction duration*, which was calculated as the median duration of non-interrupted distraction events in scenario 1.
As the data were not normally distributed, a paired samples two-sided Wilcoxon test (using *wilcox.test* from the R package stats) was used to compare the median over all participants for each of these measures between the conditions.

3.3.3.1. Understanding and appreciation of corrective action for distraction

Table 3.5 shows the median response over all participants regarding the distraction warnings. No significant difference between the sample median and the “3 – neutral” answer was found for any of the questions regarding the distraction warning. Indicating that participants generally had neutral opinions about understanding the distraction warnings and appreciating their timing.

Table 3.5 Median scores for questions regarding distraction warnings.

Question	Median (0-5)
Q1: It was obvious to me why the notification about distraction occurred.	4
Q2: I would have liked more information about why notifications about distraction were triggered.	3
Q3: Timing of the notification start was very good.	2
Q4: Timing of the notification end was very good.	2

From the interviews, the drivers generally appreciated the concept with distraction warnings (*n*=10). Some mentioned that distraction warnings should be mandatory in automated cars. However, some did not understand the warnings and wanted more information as to why the warning was triggered, including suggestions on what the vehicle thought they should focus on instead (*n*=5). Some mentioned that the distraction warnings were badly timed (*n*=4), and that they got warnings when looking at traffic at the sides, in the blind spot or in the mirrors.

3.3.3.2. Effect of condition on distraction

Figure 3.12 shows the distributions of the four distraction measures, distraction uptake frequency, proportion of the time distracted, maximum distraction duration and median distraction duration, for both the MEDIATOR and the Baseline condition. As illustrated in Panel B in Figure 3.12, when driving with the Baseline HMI, the proportion of time spent distracted in scenario 1 was significantly larger (*Mdn*=1.12%, *IQR*=2.19%) than when participants were driving with the MEDIATOR HMI (*Mdn*=0.67%, *IQR*=1.01%), *V*=285, *p*=0.02. Panel C in Figure 3.12 illustrates that the maximum duration of distraction lasted significantly longer when driving with the Baseline HMI (*Mdn*=1.8 s, *IQR*=3.7) compared to the MEDIATOR HMI (*Mdn*=1.0 s, *IQR*=1.8), *V*=296, *p*=0.01. The frequency and the median distraction duration were not significantly different between conditions.

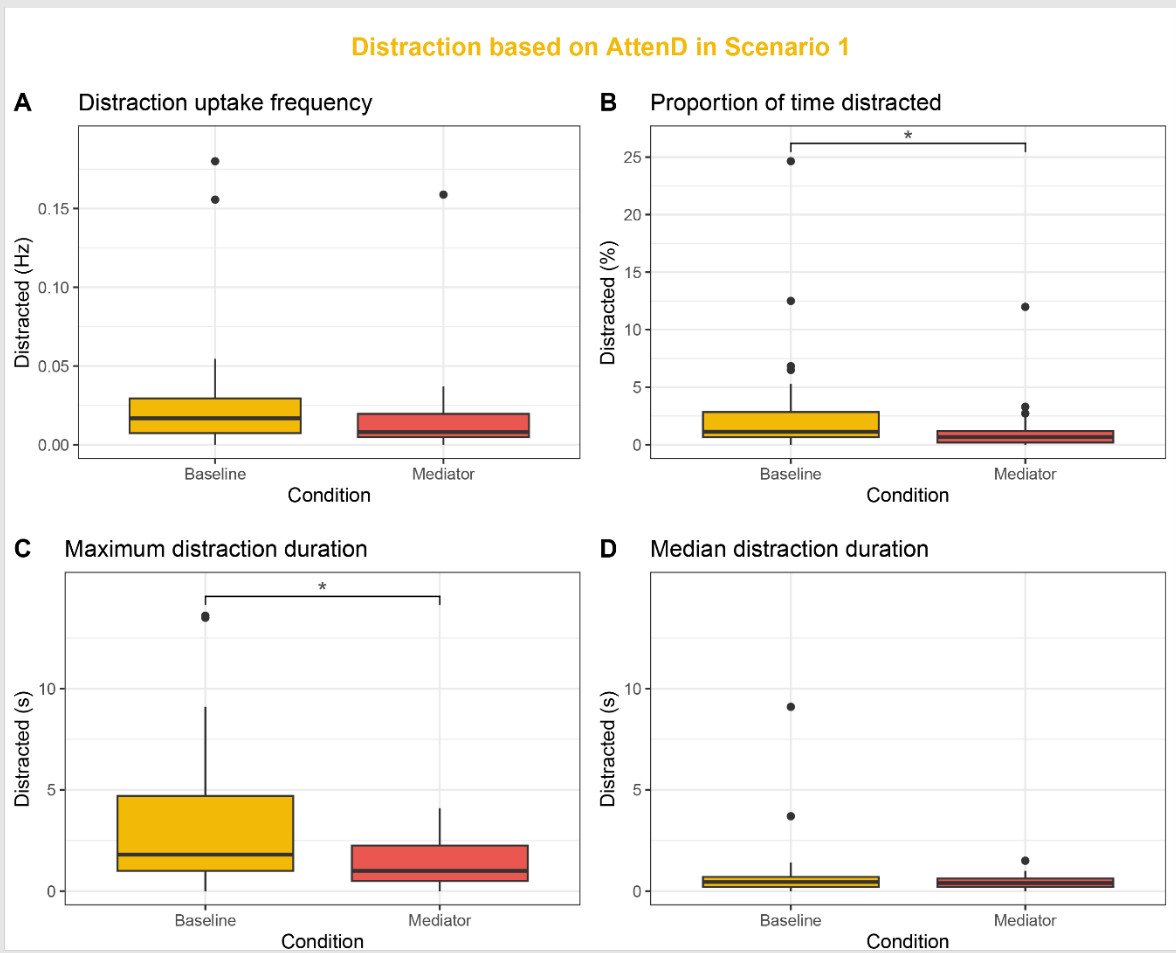


Figure 3.12 Boxplots showing the distribution of four measures relating to distraction based on AttenD in scenario 1. Differences between the Baseline HMI and the MEDIATOR HMI were tested using paired samples Wilcoxon tests (= p<.05).*

3.3.3.3. Summary

The distraction warnings were generally appreciated, but participants differed on their opinion on the clarity of the signals, indicating that the signals were obvious, but also more information was needed. The objective timeseries data shows that the MEDIATOR HMI, which also includes distraction warnings, results in significantly less distraction than the baseline HMI when driving in Continuous Mediation (Baseline: *Mdn*=1.12%; Mediator: *Mdn*=0.67%). It also shows that the maximum duration of a distraction event is much larger in the Baseline condition (*Mdn*=1.8 s) than in the Mediator condition (*Mdn*=1.0 s). Participants did not significantly increase the frequency of their distraction events, nor was the median duration of a distraction event different between conditions.

3.3.4. RQ4. How are transfers of control from automation to human experienced?

Transfer of control from the automation (partially) back to the human was experienced on several occasions. In scenario 2 the automation in Continuous Mediation mode shuts off suddenly and the driver should resume full control of the vehicle (manual mode). In scenario 5 a planned take over from Standby to Continuous Mediation mode occurs, where the driver is warned in advance. In scenario 9, after a long time of driving in Time to Sleep mode, the driver is warned about nearing the end of the Time to Sleep Operational Design Domain (ODD) and a control transfer to Standby occurs automatically. In scenarios 2 and 5, the participant is expected to change to the required automation mode by moving the gear lever (MEDIATOR) or the touchscreen slider (Baseline). If this does not happen, the interaction wizard will intervene and ask the participant to switch to the required automation mode. In scenario 9 no such action is required from the participant.

For each of the scenarios, 7 questions about the experienced takeover ritual were answered on a 5-point Likert scale from strongly disagree to strongly agree. Data from 36, 35 and 39 participants were available for the analyses for scenarios 2, 5 and 9, respectively. The log data on the interventions were available for 41 participants for scenario 2 and for 43 participants for scenario 5. Additionally, information from the interviews was gathered regarding the takeover experience. As the data were not normally distributed, a two-sided paired Wilcoxon test, using the *wilcox.test* function in the R package stats, was used to compare medians between conditions. This test was also used to determine a significant difference for each condition between the median and the neutral score “3”. No interaction effects with Age were found, so this is left out of further reporting for clarity.

3.3.4.1. Interventions

As an indication if participants needed assistance from the interaction wizard to make the switch to the new automation level in scenarios 2 and 5, the number of interventions per scenario per condition is shown in Table 3.6. In all scenarios, there were more interventions in the Baseline condition than in MEDIATOR.

Table 3.6 Number of interventions per conditions for scenarios 2 and 5. SB = Standby. CM = Continuous Mediation, M = Manual.

	Baseline	MEDIATOR
Scenario 2: SB>CM (Piloted > Assist)	21	8
Scenario 5: CM>M (Assisted > Manual)	40	32

3.3.4.2. Questionnaire

The median answer overall all participants for the questionnaire items is shown in Table 3.7. All medians (except for those which were equal to 3) were significantly different from 3. Only for question 6 in scenario 5 a significant effect ($p<0.05$) of the condition (MEDIATOR vs. Baseline) on responses to this question was found. Requests for taking back control were experienced as more obvious in the Mediator condition.

*Table 3.7 Median over all participants for questionnaire items regarding takeovers from automation to human. In scenario 2, no warning was given, but a sudden takeover was experienced. Question 5 was therefore not asked in this scenario. The symbol * indicates a significant difference between conditions*

Question	Scenario	Baseline	MEDIATOR
1. I always felt safe during the takeover procedure	2 5 9	4	4
2. The takeover procedure felt comfortable	2 5 9	4	4
3. It was easy to follow instructions to hand over control	2 5 9	4	4
4. When I was requested to take back control, I was properly warned.	2 5 9	4	4
5. I always felt safe during the takeover procedure.	2 5 9	- 4 4	- 4 5
6. It was obvious to me why the requests for taking back control occurred	2 5 9	3 3* 4.5	4 4* 4
7. I would have liked more information about why the requests for taking back control was triggered.	2 5 9	4 4 2	3 3 2

3.3.4.3. Interviews

From the interviews, more drivers were found to appreciate the way mode changes were communicated by the MEDIATOR HMI (*n*=16) compared to how they were communicated by the Baseline HMI (*n*=5). Some said that they found it easy to understand the timeliness of warnings and when it was time to change automation mode (Mediator *n*=6, Baseline *n*=6), while others found it unclear, especially when it comes to the urgency level of a takeover request, i.e., if one must change mode immediately upon receiving the message or if it is possible to wait for a while, and if so, how long (Mediator *n*=8, Baseline *n*=13).

The stylized road that was used in the MEDIATOR HMI to convey automation mode and the remaining time in the current mode was found to be helpful and informative by some (*n*=7), while others thought this important information should be communicated in a better way (*n*=7), for example with descriptive text and larger symbols/larger font size. It was also suggested that the countdown should be accompanied by a sound, especially when the timer approached zero. Information about the time to the next mode change was not present in the Baseline condition, and many participants mentioned that they lacked this kind of input (*n*=14).

In both conditions, the participants expressed a desire to improve the takeover ritual (Mediator *n*=13, Baseline *n*=15). In MEDIATOR, there was a desire to gather the information in one place instead of having it distributed across several screens. It was also mentioned that transfers of control to higher automation modes were perceived as warnings rather than as invitations/requests. Text messages were perceived as too short, and thereby ill-defined and unclear. Some indicated that they wanted to know which information the vehicle used to form its decisions, and they also wanted more information about why they needed to change automation mode. In the Baseline condition, takeover requests were found to be unclear, confusing, and hard to understand. It was difficult to understand what the icons meant. For example, does a crossed out steering wheel symbol mean that the function has already been turned off, or that you should turn it off, or that you should no longer have your hands on the steering wheel? It was also unclear which automation mode to switch to, and what would happen if you refused to make the switch.

3.3.4.4. Summary

Both systems indicated to the driver to take back control at certain instances during the trip. If this transfer of control was not performed by the participant, the experimenter would intervene and instruct the participant to perform the takeover anyway. These interventions were more often needed in the Baseline condition than in the Mediator condition, especially in the take over from Standby to Continuous Mediation. The questionnaire results indicate that during this takeover it was less obvious why the takeover was happening in the baseline than in the Mediator condition.

The interview results indicate that more participants appreciated the Mediator take over ritual compared to the Baseline ritual. Also, participants expressed a desire for information about the time to take over in the baseline condition, implying that this added information in the Mediator condition was appreciated.

Generally, the questionnaire results indicate that all takeovers were found safe, comfortable, easy to follow instructions and provide proper and timely warnings, regardless of the condition.
While the results generally indicate a preference for the MEDIATOR HMI takeover rituals, some improvements were also mentioned during the interviews. Some participants found the urgency level of the takeover request and the propositions unclear and would have preferred an additional sound during the last moments of the countdown for a takeover. Participants also would have liked more and clearer information regarding the reasons for take over.

3.3.5. RQ5. Does active proposal for handovers to higher levels of automation result in more automation usage?

In scenario 3 and scenario 6 an active proposal to handover control to automation was offered in the Mediator condition only. In scenario 3 the offer to switch from Manual to Standby mode was in response to the detected distraction of the participant. In scenario 6 the offer to switch from Continuous Mediation to Standby mode was initiated because the higher level of automation became available at that point. In both scenarios, if the participant did not accept the offer to switch to a higher level of automation, the interaction wizard would intervene and ask the participant to switch to the required automation mode before arriving in the next scenario.

The questionnaire contained 6 items about the experienced proposal to hand over control to automation. Participants could answer on a 5-point Likert scale from strongly disagree to strongly agree. Data from 45 participants were available. The log data on the interventions were available for 42 participants for scenario 3 and for 44 participants for scenario 6. Additionally, information from the interviews was gathered regarding the handover experience.

As the data were not normally distributed, a single sample two-sided paired Wilcoxon test, using the *wilcox.test* function in the R package stats, was used to determine a significant difference between the median and the neutral score “3”. No interaction effects with age were found, therefore this is left out of further reporting for clarity.

3.3.5.1. Interventions

Table 3.8 shows the number of interventions per scenario per condition. In both scenarios there were more interventions in the Baseline than in Mediator condition, and the difference was largest in scenario 6.

Table 3.8 Number of interventions needed to switch to higher level of automation per scenario per condition.

	Baseline	MEDIATOR
Scenario 3 M > SB (Manual > Piloted)	29	24
Scenario 6 CM > SB (Assist > Piloted)	22	7

These results show that participants were less inclined to switch to the available higher level of automation if they were distracted than if the offer was due to automation becoming available, as the number of interventions for both conditions was higher in scenario 3 than in scenario 6. The active proposal in the Mediator system to switch to the Standby driving mode, as compared to the simple change in availability of the Standby driving mode in the baseline condition, resulted in less need for interventions and thus more switches to the new automation mode by the participants themselves.

3.3.5.2. Questionnaire and Interviews

The median answer overall all participants for the questionnaire items is shown in Table 3.9. All medians (except for those which were equal to 3) were significantly different from 3.

Table 3.9 Median over all participants for questionnaire items regarding the active proposal to take over control.

Question		MEDIATOR
1.	**I always felt safe during the handover procedure.**	3
2.	**The timing of the proposal to hand over control was well chosen.**	3
3.	**It was easy to follow instructions to hand over control**	4
4.	**It was obvious to me why the proposal for handing over control occurred.**	4
5.	**I would have liked more information about why the requests for taking back control was triggered.**	4
6.	**I would make more use of automated functions when they are proactively proposed when they become available.**	5
7.	**I would make more use of automated functions when they are proactively proposed when I am distracted.**	5

Participants were neutral with respect to safety of the handover ritual but found it easy to follow instructions for the hand over. They understood why the handover was proposed, but still would like more information on why the handover took place. Participants strongly agreed that the active proposals, in case of either distraction or automation availability, would result in them making more use of automated functions. From the interviews participants also indicated that they liked that the vehicle suggested to level up when they were distracted (*n*=10).

3.3.5.3. Summary

Participants appreciated the active proposal to a higher level of automation in case of both distraction and upcoming automation availability. Participants strongly agreed that they would make more use of automation if such active proposals were offered in their car. In case of upcoming automation availability, the participants also indeed needed less help from the interaction wizard to switch to the higher automation level if the active proposal was present (i.e., with the Mediator system).

3.3.6. RQ6. How do corrective countermeasures affect fatigue?

The reported fatigue levels in the dataset were normally distrubuted with ratings ranging from KSS = 1, extremely alert, to KSS = 9, very sleepy, great effort to keep awake, fighting sleep, see Figure 3.13 (left). An analysis of variance with categorical factors for MEDIATOR/Baseline and for alert/sleep deprived and a continuous factor for distance driven showed that the participants in the sleep deprived group reported higher KSS ratings, $F_{(1,1262)}$=40.1, p<0.001. No significant differences in fatigue ratings were found between the MEDIATOR and the Baseline systems, Figure 3.13 (right). A clear time on task effect with increasing levels of fatigue is evident in both conditions, with no differences between the MEDIATOR and Baseline systems, $F_{(1,1262)}$=99.62, p<0.001. Occurrences of high fatigue levels were evenly distributed throughout the route (Figure 3.14).

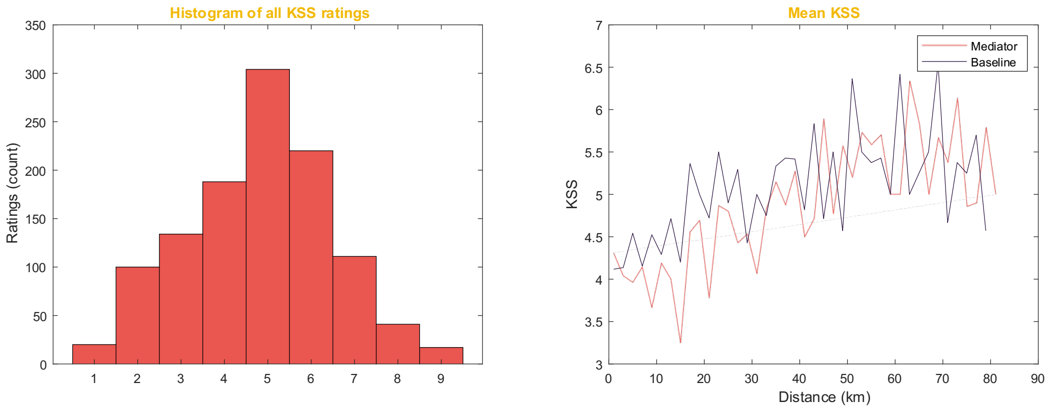


Figure 3.13 Histogram of KSS ratings across all participants in the study (left), and the mean KSS ratings as a function of distance driven (right).

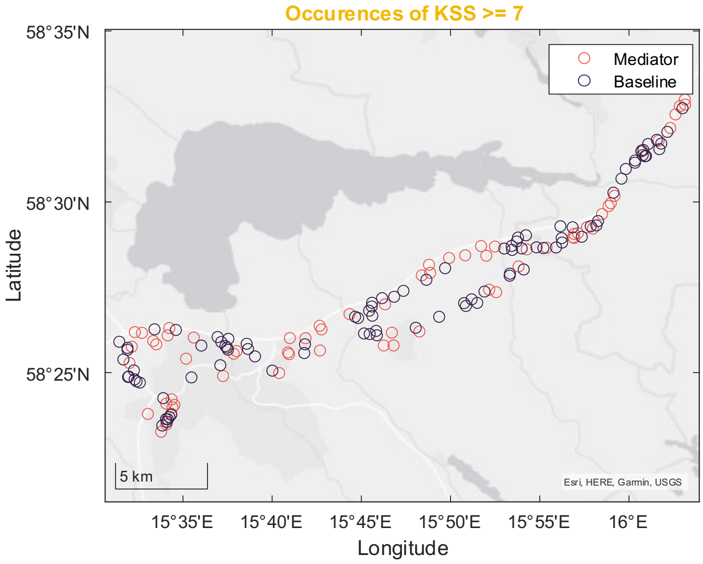


Figure 3.14 Occurrences of high levels of fatigue along the route.

As described previously, in both the MEDIATOR and the Baseline condition fatigue warnings were triggered based on the subjective rating of fatigue provided by the participant. To investigate the effectiveness of the warnings the last KSS rating before each fatigue warning were compared with the first KSS rating after the warning. An analysis of variance, with categorical factors for MEDIATOR/Baseline and for alert/sleep deprived, showed no significant differences in KSS levels before versus after the warnings. Neither were there any significant differences in the two fatigue indicators heart rate and root mean square of successive differences before versus after the warnings. The two heart rate related metrics were calculated in 60-second time windows before and after the warning.

Data from four items in the questionnaires referred to the fatigue warning signal and were all answered on a 5-point Likert scale (It was obvious to me why the notification about degraded fitness occurred, I would have liked more information about why notifications about degraded fitness were triggered, Timing of the notification start was very good, and Timing of the notification end was very good). Data from 44 participants were available for the analysis. As the data were not normally distributed, a one sample two-sided Wilcoxon test (using *wilcox.test* from the R package stats) was used to compare the medians over all participants between conditions. No significant differences between conditions were found. A more in-depth analysis was performed on possible interaction effects between Condition and Age by fitting an ordered logistic regression model (using the R function *polr* from the R package MASS) to the data. Significance of the effects was tested with a type II anova test (using the R function *Anova* from the R package rstatix). No significant effects were found.

From the interviews, the participants generally appreciated the fatigue warnings (Mediator *n*=7, Baseline *n*=12). They understood why they got the warnings and found them to be timely and helpful. The seatbelt vibrations that were used to alert drivers in the Mediator condition was found to be effective (*n*=16), and the participants liked that it was a physical warning communicated via a channel that is not used for anything else. Some found the seatbelt warnings to be uncomfortable, causing slight panic reactions (*n*=5). The inflatable seat was however not appreciated (*n*=9), especially as it felt strange and as the seatbelt warning was found to be more effective.

3.3.7. RQ7. Is it possible to alert the driver after being disengaged from the driving task?

Data from 86 drives could be used analyse the effect of the smooth transition from Time to Sleep to Standby. Fourteen drives were excluded due to data logger issues or system errors occurring before the participants reached the point where the transfer of control took place. The fatigue level of the drivers, plotted as a function of distance driven, is illustrated in Figure 3.15. The first vertical line represent when the transfer of control is initiated. Data for fatigued drivers (KSS ≥ 7) who received a wake-up notification as part of the transfer of control ritual are visualised in the right subplot. It can be seen that the reported fatigue levels steadily increase with distance driven but starts to decrease already before the alert-message is triggered. This decrease in fatigue takes place when the vehicle approaches the city of Linköping, where it is soon time to exit the motorway, and where traffic becomes denser. An analysis of variance showed that there was a significant effect of the alert-message ($F_{(1,78)}$=4.42, p=0.04) on the reported KSS level, with a mean reduction of 0.57 KSS units after the warning. The reduction was smaller in the group of shift workers (0.26 KSS units) compared to the other group (0.68 KSS units). Similar results were found for the other two fatigue indicators, with slightly lower heart rate amongst shift-workers ($F_{(1,72)}$=8.28, p<0.01) and after the alert-message ($F_{(1,72)}$=6.35, p=0.01), and higher root mean square of successive differences amongst the shift-workers ($F_{(1,72)}$=7.31, p<0.01) and after the alert-message ($F_{(1,72)}$=5.42, p=0.02). There were no significant differences in any of the fatigue indicators between the MEDIATOR HMI compared to the Baseline HMI, nor were there any interaction effects. All in all, the results indicate that the corrective alert-message used in the transfer of control ritual somewhat reduced task-related fatige but not sleep-related fatigue.

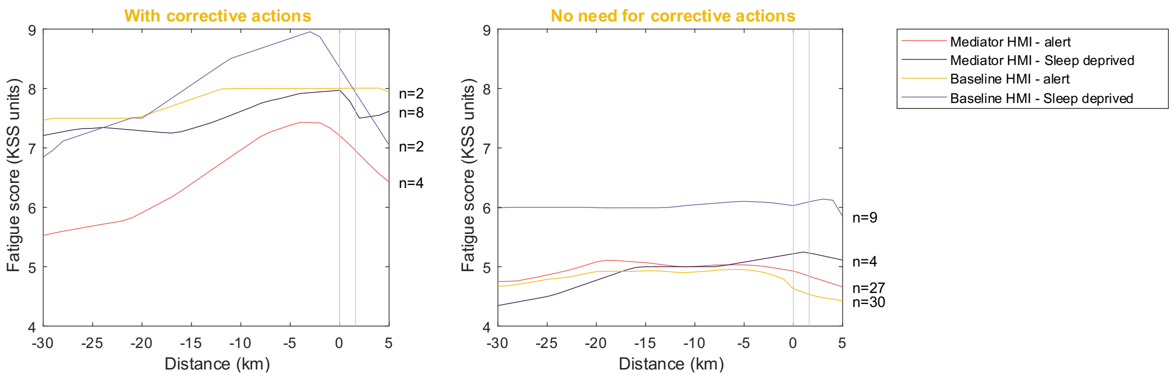


Figure 3.15 Mean fatigue level as a function of distance driven in the drives where a corrective alert-message was triggered (left) versus when no corrective measure was needed (right). The vertical lines represent the locations where the transfer of control procedure was initiated/ended. The number of drives included in the mean calculation is accounted for to the right of each curve.

From the interviews, some participants noted that it felt unpleasant to get sleepy when driving in Piloted mode (n=2), and they expressed a concern that it was easy to become too relaxed and that they got more tired from not doing anything active while driving. Some also mentioned that the fatigue warnings should be more invigorating in cases where you are really about to fall asleep (Mediator n=4, Baseline n=4). There was a request that you should interact more with the system, e.g., confirm that you have seen the warnings and that you know you need to be more careful, for example when driving off the motorway after a transfer of control when you have been sleepy in the automated preceding road stretch.

3.4. Discussion

The aim of the study was to evaluate the functionality, safety effects and user acceptance of the Mediator system under different degraded driver performance conditions (distraction and fatigue), including conditions of degraded automation.

3.4.1.1. User acceptance

For insight in the general experiences of participants with both the MEDIATOR and the Baseline systems, participants were asked about their acceptance of the systems, their usability, trust in the system and willingness to buy the system and the workload due to interaction with the system after each trial. No significant differences between conditions were found in these questionnaires. However, when asking questions on similar subjects in a direct comparison between conditions at the end of the trial a clear preference for the Mediator system was reported. The participants' preference for the Mediator system in terms of wanting to buy the systems, their comfort, how easy the systems were to learn and use, and how safe and trustworthy the system was ranged from 49% to 71%, while the preference for the baseline system ranged from 17% to 36%. During the interviews also there were 2.5 times more participants that mentioned trusting and being willing to, given that it was working as envisioned and that it was not too expensive, buy the Mediator system than there were participants stating this about the Baseline system. One possible explanation for the different results is that drivers preferred the Mediator system, but that, due to the low sample size in combination with high interpersonal variability, the effect size of this preference was not large enough to be measured with the questionnaires at the end of each trial.

In a more in-depth analysis, a significant interaction effect between condition and age was found. Younger participants generally preferred the Mediator system, while elderly participants preferred the Baseline system. For workload a similar interaction effect was found, where increasing age corresponded to a higher reported workload for the Mediator system. One explanation could be that age-related cognitive decline (Callaghan et al., 2017) made it more difficult for elderly participants to cope with the novel information channels (such as LED bars and the novel road icon) that were also spatially scattered. This could have increased workload and, with that, decreased the appreciation of the Mediator system. In future research it is advised to take this into account and investigate other characteristics, next to age, that effect the appreciation of these systems, so that they can be designed and adjusted to the differing needs. For example, the interview results indicate that participants, regardless of age, did not appreciate that information was scattered over different screens. This should be adjusted in future versions of the Mediator system.

3.4.1.2. Functionality

Generally, drivers understood their responsibilities during the drive, but this understanding was lower for the novel automation modes Standby and Time-to-Sleep than for the Manual and Continuous Mediation modes, which are similar to automation options already in vehicles nowadays. It is likely that the understanding of the novel modes would increase with increasing real-world experience. Participants did indicate that their continuous awareness of the time until a change in responsibility would occur was higher in the MEDIATOR than in the Baseline condition. As also the interviews showed that drivers used the time budget information that was communicated via the LED bars and the road icon to obtain the awareness of any upcoming responsibility switch. Participants also indicated during the interviews after the Baseline condition that they missed information about the time until their responsibilities would change. These findings suggest that communicating such time budget information, which is currently not standard present in vehicles on the road, could be a valuable addition to future vehicle HMIs. However, during the interviews drivers also showed a preference to receive information through icons and interact through a touchscreen, which are both communication channels they were familiar with. Touchscreens, however, can have a detrimental effect on driver attention (Ferris et al., 2016) due to, amongst others, the lack of haptic feedback and might thus not be the ideal interface. Nonetheless, for future HMI designs it is advised to take these learned affordances into account, such as adding novel communication devices in new cars only stepwise and not all at once.

3.4.1.3. Safety effects

The Mediator system supports the driver in their monitoring task by providing continuous feedback on the current driving mode and the time until upcoming switches, as well as warning the participant when they fail to perform their monitoring task due to distraction. The results of this experiment show that drivers are indeed distracted a smaller proportion of the time in the MEDIATOR than in the Baseline condition. It is possible that the combination of continuous time budget information and the corrective warnings made drivers more aware of and/or more willing to comply with their responsibilities. Not only the total duration, but also the maximum duration during which participants were distracted was lower in the MEDIATOR than the Baseline condition. This result might have the most beneficial effect on road safety, as in Simons-Morton et al. (2014), the single longest glance was shown to provide a more consistent estimate of crash risk than total time eyes off the forward roadway. Furthermore, the distraction warnings were generally appreciated. However, participants differed on the clarity of the signals, indicating that the signals were obvious, but also more information was needed. In future research the clarity of these signals should therefore be improved. As in this experiment only the combination of proactive (time budget and continuous mode information) and corrective (distraction warnings) was experienced by the participant, it is unclear which of these items were of most influence on the participant glance behaviour and/or their appreciation of the warnings. It is, for example, possible that *supporting* features such as continuous communication of driving mode and anticipatory information on future responsibility changes, affected the appreciation of and adherence to the *corrective* feature, i.e., the warnings. In future research it would be interesting to investigate the effect of each of these features on both distraction and appreciation of the HMI separately as well as in combination.

When a takeover to a lower level of automation was needed, both the Baseline and the Mediator system guided the participant through the takeover. In the Baseline condition the interaction wizard needed to intervene more often than in the Mediator condition, indicating that the Mediator takeover ritual was somewhat better understood. The interview results also showed a higher appreciation of the takeover ritual in the MEDIATOR than in the Baseline condition. The questionnaire results additionally showed that in the Baseline condition it was less obvious why the takeover occurred. It is hypothesized that not understanding why a takeover occurs could have a negative effect on the overall understanding and appreciation of the takeover ritual. While the results generally indicate a preference for the Mediator system takeover rituals, some improvements were also mentioned during the interviews. Some participants found the urgency level of the takeover request unclear and would have preferred an additional sound during the last moments of the countdown for takeover. Participants also would have liked even more and clearer information regarding the reasons why a takeover was initiated. To improve appreciation and understanding of the takeover ritual for future research and design, it is therefore advised to make the takeover warnings clearer by both including a sound signal and informing the participant why the takeover is happening. As indicated in a literature study on takeover requests by Jansen et al. (2022), additional research is required to establish the optimal implementation of such signals for safe takeovers.

In the Mediator condition a novel feature where the system proposes to increase the level of automation was implemented. This proposal occurred if a higher level of automation just became available or if the participant was distracted and a higher level of automation already was available. Participants appreciated these proposals and strongly agreed that they would make more use of automation if these active proposals were offered in their car. This effect was also observed during the study, where participants more often switched to the higher level of automation by themselves after an active proposal to increase automation (MEDIATOR) than when simply a new icon would be shown (Baseline). If driving with automation can be made safer than driving manually, this feature could thus potentially improve road safety.

In both the MEDIATOR and the Baseline condition the driver was warned if they were too fatigued. The different fatigue warning designs did not affect the level of fatigue. However, the results indicate that the corrective alert-message used in the transfer of control ritual somewhat reduced task-related fatige but not sleep-related fatigue. This finding is in line with previous research, stating that task-related fatigue can be countered by doing something else for a while (such as taking over control of a moving vehicle) while sleep-related fatige can only be countered by actual sleep (May and Baldwin, 2009).

4. Functionality, validity, and reliability of the full Mediator System with the Technical Integration in-vehicle prototype

This chapter describes the evaluation of the integrated Mediator system as realised in one key Mediator prototype, the Technical Integration (TI) in-vehicle prototype.

In terms of full integration of all actual (not simulated or heavily simplified) Mediator software components, this was the most complete prototype in the MEDIATOR project. All components were integrated, interacted in real-time with each other, and were tested in a real-world vehicle with real test drivers. This, therefore, serves as an example of how the project realised the fully integrated Mediator system and that, in general, such a Mediator-like system is actually *feasible*. This “Proof of Concept” of the Mediator concept was, for this prototype and this study, as important as the more detailed individual research questions (and their answers) described below.

Being an actual road legal vehicle, this TI prototype vehicle was limited, in terms of SAE automation levels, to SAE Level 2 (in MEDIATOR terms: Continuous Mediation, CM). Another limitation, or simplification, compared to the ‘full’ MEDIATOR concept and to several other of the Mediator prototypes (esp. the “Human Factors” or HF-oriented vehicle described in a previous chapters), was that the HMI component was somewhat simplified in terms of visual, auditory, tactile, and control elements – while still keeping the essentials, and actually being sophisticated in terms of real-time software and vehicle integration.

4.1. Objectives, Research questions, and covered use cases

This study aims to evaluate the functionality and user acceptance of the fully realised technical Mediator Proof-of-Concept system, as implemented in the Technical Integration prototype vehicle. Special focus is given to the functionality of the Automation State component and the Decision Logic component, as this prototype was the only one where those components were fully technically realised. However, also the Driver State component and the HMI were part of the system, and (therefore) part of the overall evaluation.

In the study, the Automation State module under different driving contexts, including the assessment of how well the system can predict bad automation performance as defined in D1.3 (Mano et al., 2021), is covered. The study aims to investigate research questions concerning the Automation State component and the user perspective of the Mediator system. Furthermore, the decisions and recommendations of the central Decision Logic are evaluated in terms of appropriateness and user acceptance.

4.2. Research Questions

The Research Questions (RQ) for the study are noted as follows:

1. How reliably does the automation state component calculate automation fitness and unfitness compared to the GPS locations of known static Operational Design Domain (ODD) changes?
2. What is the driver's perception of the reliability of the Mediator system after repeated exposures to the same static ODD change?
3. What timings are most appreciated by the drivers in repeated exposures to the same static ODD change?
4. How do drivers evaluate the Mediator system with its functions and the HMI after repeated use?

4.2.1. Use cases/scenarios covered

The selected route consisted of and covered five scenarios. The scenarios have been designed to evaluate the following functionality of the Mediator in-vehicle prototype:

- Driver initiated CM on (UC6, scenario 1)
- Mediator initiated CM on (UC6, scenario 2)
- Low automation performance, CM will shut off soon (adapted UC9, modified scenario 3)
- Shorter-term planned Mediator initiated takeover from CM (adapted UC 5, scenario 4)
- Longer-term planned Mediator initiated takeover from CM (adapted UC 5, scenario 5).

4.3. Methodology

The method for this study is focused on collecting data from professional test drivers to evaluate the performance of the automation state module on public roads. Ethical approval was obtained from the MEDIATOR ethical review board.

4.3.1. Participants

The study includes seven professional test drivers. The test drivers have a history of working with safety system development and are used to driving prototype vehicles. Drivers were between 27 and 58 years old (M = 42); one driver was female and the other six were male.

4.3.2. Procedure and design

The participant drove the TI in-vehicle prototype on a specified route. Each drive consisted of a pre-defined route with a duration of 1 hour. The pre-defined route (see Figure 4.1) contains certain stretches known to affect automation performance so that the driver will probably wish to deactivate assisted driving, or that automation will reach its limitations and deactivate itself. The choice of route was made to ensure long stretches of safe automation that fulfilled the use cases/scenario requirements for the MEDIATOR study, along with stretches that did not have automation availability. To ensure the safety of the drivers the route chosen did not have many road works. The route consisted of short inner-city road stretches, long motorways, and the long rural road stretches with lane markings. Drivers performed several trials consisting of two drives, hence, they drove the specified route twice in a row (2 h driving). On average, drivers performed five trials on different days or different times of day (morning vs. afternoon).

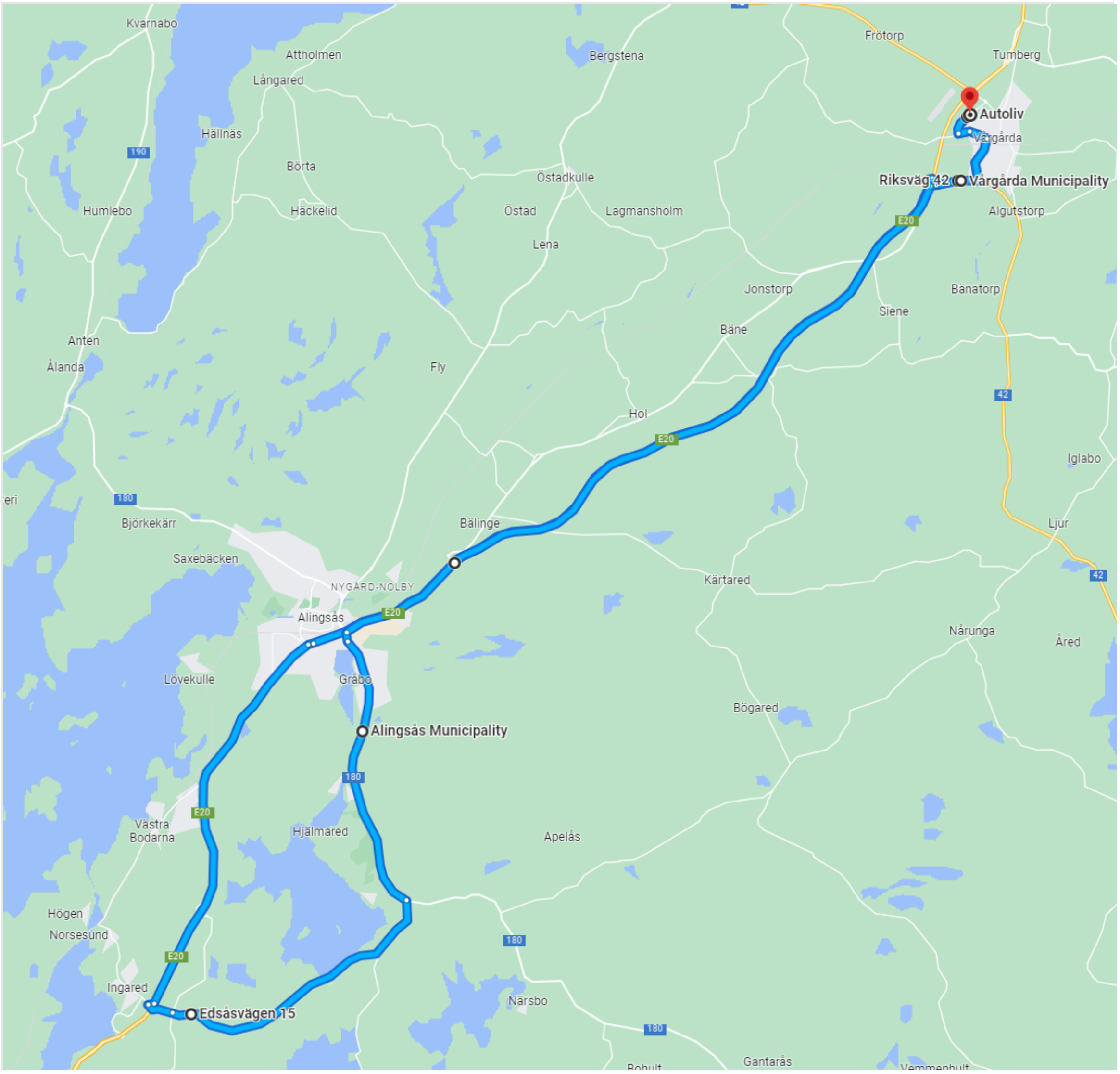


Figure 4.1 The pre-defined route for the on-road study.

Test drivers were not familiar with the chosen route; hence, the test leader joined on their first drive (Drive 1) and a smartphone, along with paper map with route information was available in the test vehicle for subsequent drives. Sufficient breaks were provided between each drive to ensure the safety of the drivers. Before the first drive, participants were informed about the study, and signed a consent form (see Appendix A.1). Afterwards, they were introduced to the TI in-vehicle prototype. The study was carried out over several weeks. **Fout! Verwijzingsbron niet gevonden.** shows the s cheme of how the study was carried out. Each driving session consisted of driving the selected route twice in a row (approximately two hours driving time in total). For the first two blocks, drivers performed four drives with the complete Mediator system (e.g., time budgets, animated colours on displayed route, proactive proposals to change driving mode). In the following drives, participants experienced different configurations of the Mediator system (Block 2, 3 and 4):

- Block 2: Drivers experienced a basic system without proactive requests by Mediator to change driving mode and no information about time budget

- Block 3: Drivers experienced Mediator recommendations to activate assisted driving also for very short parts resulting in short time budgets and frequent take-over suggestions
- Block 4: Drivers experienced again the complete Mediator system but with added distraction detection functionality.

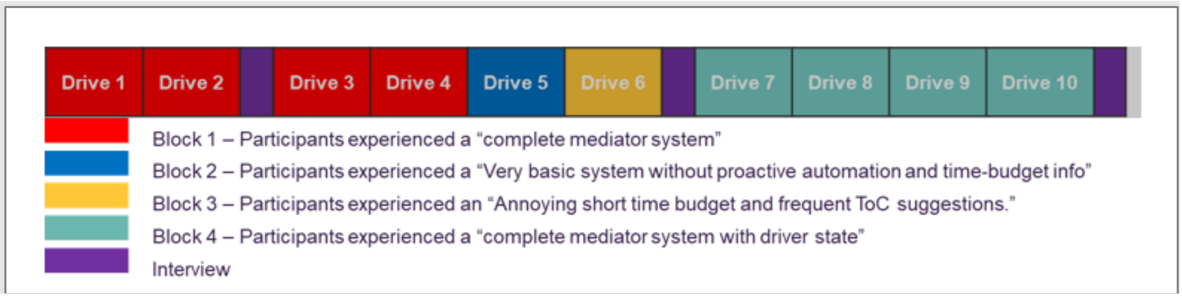


Figure 4.2 The experiment design and different configurations followed during the study.

In the subsequent drive, any change in the configuration of the TI in-vehicle prototype was made aware to the participant. The drivers filled in questionnaires after each drive. Interviews were conducted after Drive 2, Drive 6, and Drive 10. The questionnaires are not validated and are single-item measurements used to receive quantitative data to compare the different drives. No randomization of the drives was performed between the test drivers because:

1. Only seven drivers took part in the study. To be able to compare their experiences on a qualitative level, the whole study was designed to be as comparable as possible between the drivers.
2. Drivers should experience the complete Mediator system first and repeatedly to get familiar with the system and its functionalities to be able to evaluate the system properly.
3. Afterwards, drivers experienced the basic configuration as kind of a baseline giving them the opportunity to evaluate which Mediator-specific functionalities they missed the most.
4. Drivers experienced the configuration with shorter time budgets and more frequent take-over suggestions after the baseline to evaluate other settings of Mediator but also to experience one possibility of adaptation to user needs.
5. At the end of the study, drivers experienced the full Mediator system again but with added driver distraction detection demonstrating Mediator's potential to react to driver states.

Most important for the study were the in-depth questions during the interviews.

4.3.3. Measurements

4.3.3.1. Subjective ratings

After each drive, drivers were asked to fill in a short questionnaire regarding their impression of the system and the timings. Forms used to collect subjective ratings are attached in the Appendix A.2.

4.3.3.2. Online interviews

Structured online interviews (as shown in **Fout! Verwijzingsbron niet gevonden.**) were c onducted after Drive 2, Drive 6, and Drive 10. Detailed interview questions can be found in the Appendix A.3.

- The first interview aimed to examine test drivers first impression of the Mediator system (e.g., trust, acceptance, perceived safety, comfort, evaluation of the timings, preferences regarding timings and transparency, and understanding of HMI messages).
- A second interview was done to examine test drivers experience of the different configurations. Further, the interviews focus on the question of the higher familiarity with the Mediator system changes the drivers' perception of the system (e.g., trust, acceptance, perceived safety comfort, preferences).
- The final interview was conducted focusing on a final evaluation of the Mediator system with its functions and the HMI.

4.3.3.3. HMI interactions

The test drivers during the study experienced different MEDIATOR HMI based on the configuration selected. An example of HMI during the drives is shown in the Figure 4.3 and Figure 4.4.

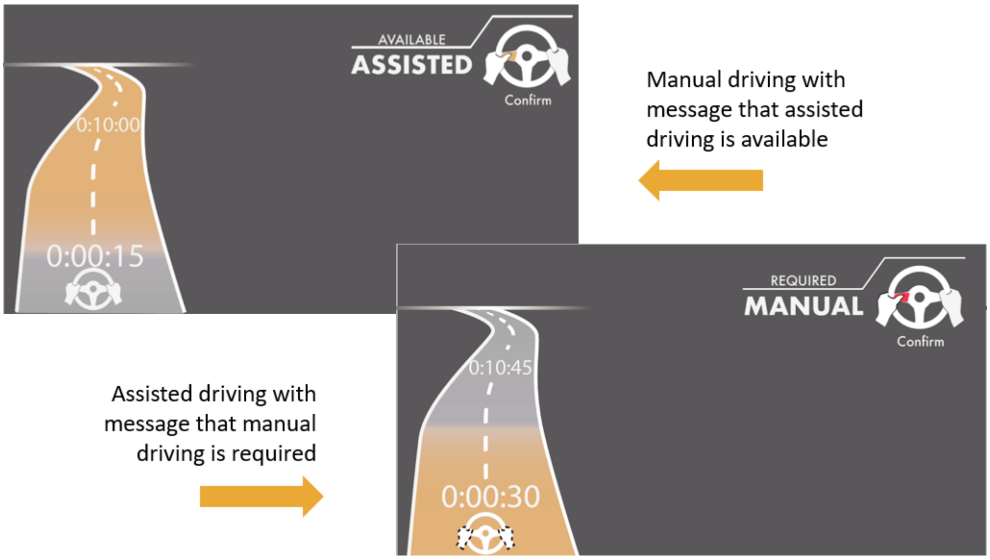


Figure 4.3 HMI example shown in the centre stack of the test vehicle.

Figure 4.4 LED strips used inside the cabin of the test vehicle to show the driving mode of the TI in-vehicle prototype.

4.3.4. Data pre-processing and statistical analysis

Data pre-processing and statistical analysis are built upon the dataset collected from the TI in-vehicle prototype field test. The analysis of the collected data has been performed at three main levels, namely the component level (e.g., automation state component, driver state components, driving content component, decision logic component, and HMI), the Mediator system technical level, and the user behaviour level, to see how the components and overall Mediator system perform, relative to what was intended and expected, and to see how the users use the system, relative to what we expected from them.

The variables of interest that were provided in the database that were relevant to analysis are linked to time constraints and are related to the following set:

- timestamp of each report record
- geographical information of the vehicles (longitude and latitude positions)
- time-budget provided by the Mediator system (together with the decision logic to determine take-over time and switch back time)
- the average attention (*AttenD*) and distraction values provided
- the decision logic to distinguish different automation levels/status (L0, L1, or L2)
- the gaze area of drivers during the test driving (associated with its duration, frequency, alternation).

For each test drive, there is a unique index to distinguish the specific drive scenario and the driver under the corresponding driving and automation conditions. For each of the concerned variables related to the research question, repeated measures ANOVA (followed by Post hoc analysis) analysis was conducted to test if there is a statistically significant difference thereof between the different clusters of drives and drivers. In total, that data set included 7 professional drivers with 10 drives per driver. Besides, the analysis was conducted for the highway-only section and for the entire field-test route. The detailed analyses are further elaborated in Section 4.3.

The following four major research questions (with a few related sub-questions) were investigated to get an idea of the performance of the Mediator system and its components, and insights into the user behaviour:

1. How reliably does the automation state component calculate automation fitness and unfitness compared to the GPS locations of known static ODD changes? Regarding time-budget accuracy.
2. What is the driver's perception of the reliability of the Mediator system after repeated exposures to the same static ODD change?
   Analyses focused on gaze behaviour, automation usage and distraction.
   - Gaze behaviour:
     - Do drivers look at the screen less frequently over time?
     - Do drivers look at the screen less in drive 5 (No Mediator)?
     - Does driver gaze behaviour change over time (i.e., number of gaze direction changes)?
     - Are there clusters of drivers in terms of gaze behaviour on the Mediator display?
     - Was the gaze behaviour different (i.e., gaze changes, gaze length) in drives (1-4), special cases (5, 6) or distraction-system-active (7-10)?
   - Automation usage:
     - Is there a difference in mean take-over time (i.e., switching to manual driving) between drivers and drives, after a suggestion from Mediator?

- If there's no time budget (Drive 5), is automation turned off later?
- Is there a difference in take-over time (i.e., switching to manual driving) between different conditions?

- Distraction:
  - Is there a difference in Distraction when distraction warnings were active versus inactive?
  - Is there a difference in visual Distraction when Mediator is AVAILABLE versus UNAVAILABLE?
  - Is there a difference in Distraction when pilot assist is ON versus OFF?

3. What timings are most appreciated by the drivers in repeated exposures to the same static ODD change?
   A few sub questions:
   - Did some drivers switch back to manual driving later than others?
   - Did some drivers switch on automation when automation becomes available much later than others do?
   - Did some drivers switch on automation much more than others (i.e., number of times)?
   - Is there a difference in gaze behaviour before and after take-over request?
4. What is the driver view of the Mediator system after repeated use? And what did drivers say about this based on the conducted interviews.

4.4. Results

The analysis results are presented for highway-only driving and for the entire route (Figure 4.5). This would help in understanding the Mediator prototype with pilot assist usage and compare it with the rest of the route or the entire route. The data analysis results related to the vehicle logs are split into their respective research questions.

Figure 4.5 The left figure shows the entire route and the right figure focuses on the highway section of the route.

4.4.1. RQ1: How reliably does the automation state component calculate automation fitness and unfitness compared to the GPS locations of known static ODD changes?

For this research question, the time budget accuracy was investigated. A comparison between the elapsed times between the moments the first "time budgets" were computed to the realized time until that zone is reached. The first TTAU (Time To Automation Unfitness) or TTAF (Time To Automation Fitness) expected time budget values were computed after each change in the automation availability zone. The analysis was conducted per driver and the results are illustrated in the Figure 4.6.

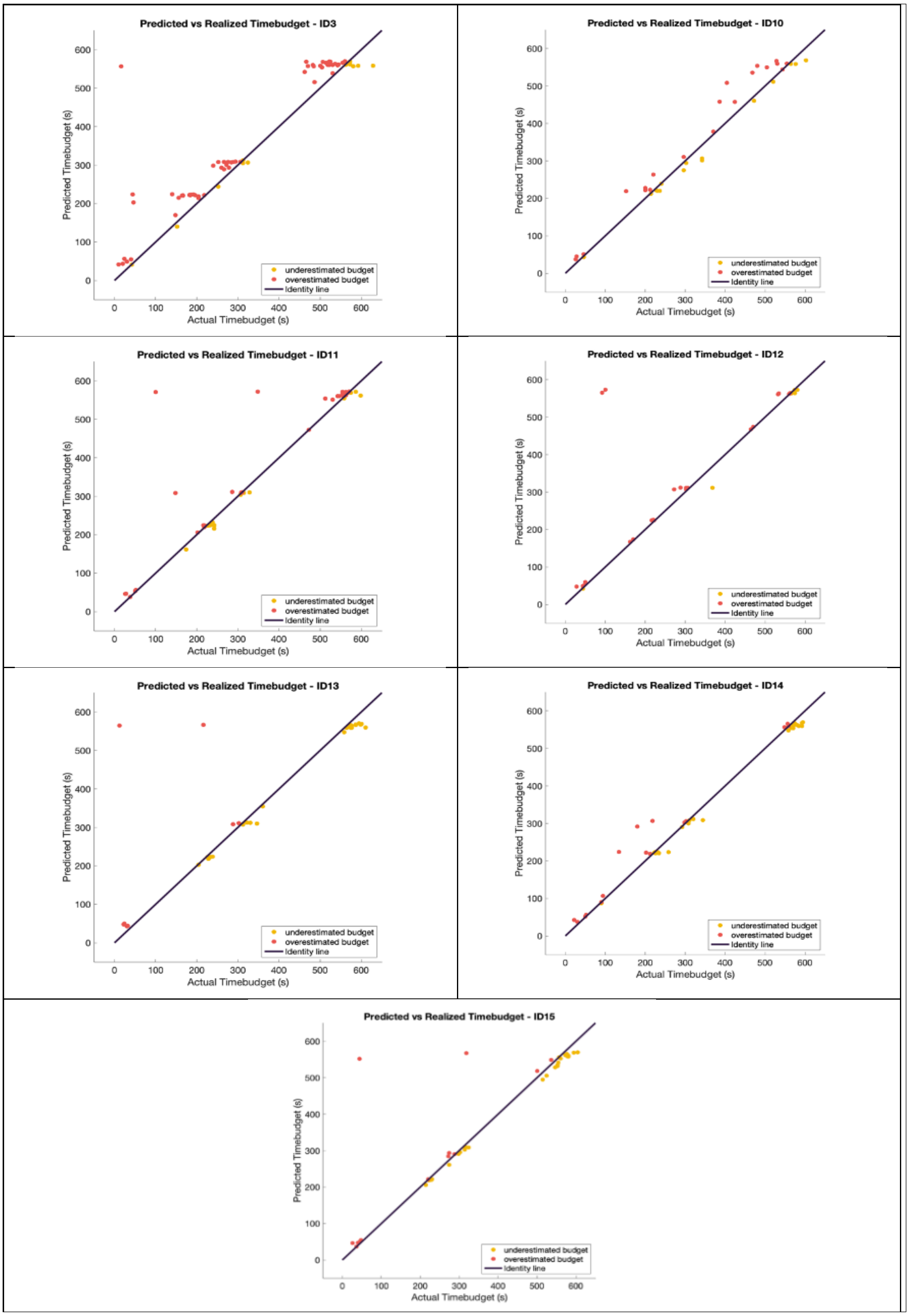


Figure 4.6 Predicted time budget versus actual time budget per driver for all drives[5].

As seen in the plots of Figure 4.6, the results mostly align around the identity line indicating a generally good match between the predicted and actual time budget. About the magnitude of the time-budget mismatch, Figure 4.7 illustrates the mismatch falls within the range of 60s, and most of them are between 20s.

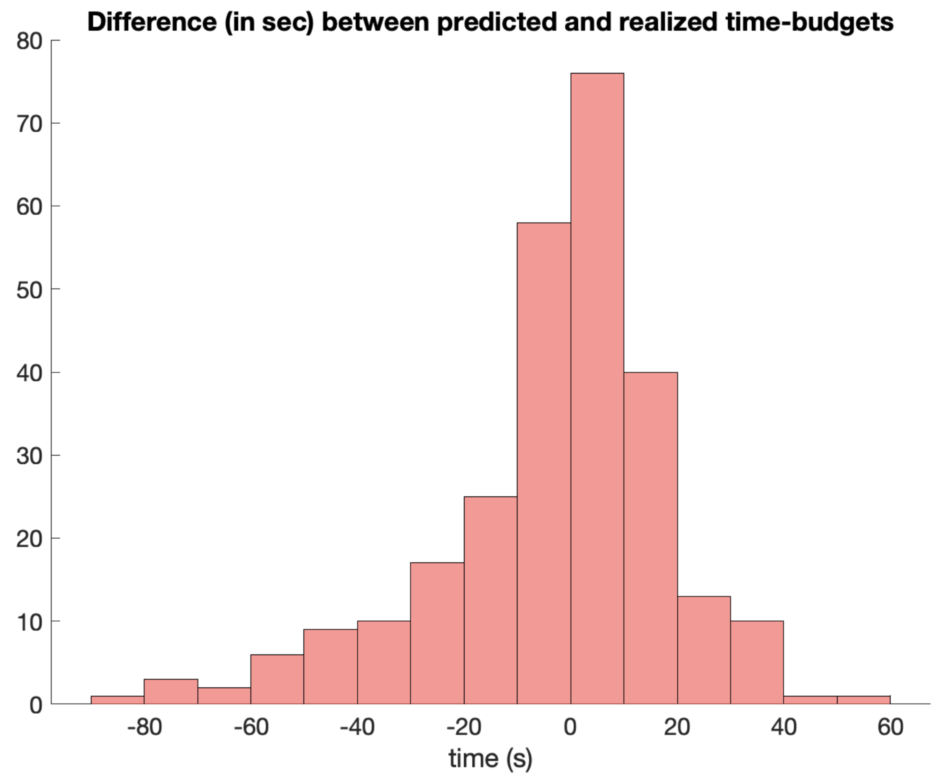


Figure 4.7 Histogram shows the magnitude of the time-budget mismatch.

Drivers' answers during the interviews support this observation. After Drive 2, two drivers (ID12, ID15) spontaneously mentioned the time budget when asked about Mediator's reliability in general: "*Yeah, maybe I can say the timing for the end of automation and start of manual and so on seemed reliable and useful* "(ID12)

After the last drive, one driver referred to the time budget as an example for Mediator's reliability: "*Yes. So, the timings were consistent*." (ID15)

When asked about the HMI, drivers also referred to the time budget as a good and helpful functionality (ID3, ID10, ID11, ID14, ID15) that can have added value to automated driving (ID3) especially for longer drives (ID10, ID11) and in higher automation modes (ID12, ID14).

5 As the test vehicle was used for other data collection effort outside MEDIATOR, ID number, in the following figure, doesn't refer to size of the dataset. Data set only contains 7 drivers.

“*Of course, it's good to know how long this will be available because it's very beneficial for us to plan. Okay, what we can do. Like, as I said, I know that I'll be in assisted driving for next 9 minutes. Then I can just quickly look at my notifications on my phone and so on*.” (ID10)

Only driver ID13 was a bit sceptical if the time budget information is really needed when having navigational information available.
“*I think it's kind of good prior to activation, like, you know there will be for so many minutes of good support. And it's... It's an indication that it's worth to activate and then prior to taking an exit, it's also…You know how long it's left. It could maybe help you, like if you should overtake someone or not. But then maybe it's different if you use it in combination with a navigation system, it would kind of give you the same amount of information. So, then I'm not sure if it’s needed*.” (ID13)

When directly asked if the information regarding the time budgets provided were reliable after the second drive, all seven drivers answered positively:

“*I wanted it to be activated more at some occasions. But at most occasions: Yes, it was. It was correct*.” (ID13)

One driver (ID15) expressed the wish to have the time budget shown even in situations where Mediator did not recommend a change in driving mode and, hence, did not show the time budget anymore after changing the mode against Mediator’s recommendation:

“*What I would have liked sometimes is when I was in assistance mode, but the system did not suggest that, then it would still be nice to see like the timings that the system would suggest on this road*.” (ID15)

One driver even stated that he/she used the time budget as indicator for driving directions after the second drive illustrating that he/she relied strongly on the provided time budgets:

“*I thought of using the map for the route, but then I decided, you know, like, after a while I know that I was looking at the centre stack because the centre stack sort of tells me where the manual driving is going to start, right. […]. I thought: Okay, now the system sort of will tell me because it's a clock ticking there. Like you can tell that in 3 seconds it will be in a junction. So, then I sort of relied on that system*.” (ID10)

4.4.2. RQ2: What is the driver’s perception of the reliability of the Mediator system after repeated exposures to the same static ODD change?

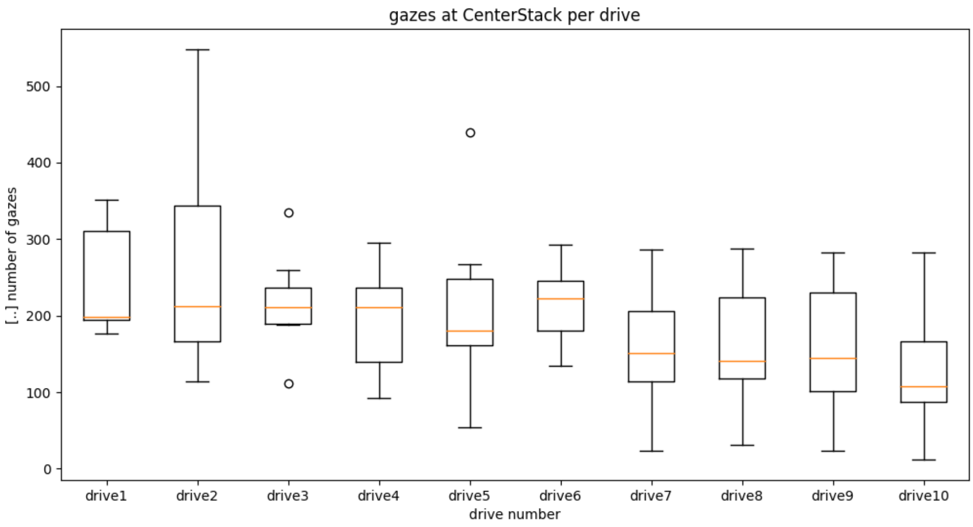
The results of the analysis were focused on comparing driver gaze behaviour over the duration of the drives and in between different configurations of the Mediator setup from the protocol.

4.4.2.1. Do drivers look at the Mediator display less over time?

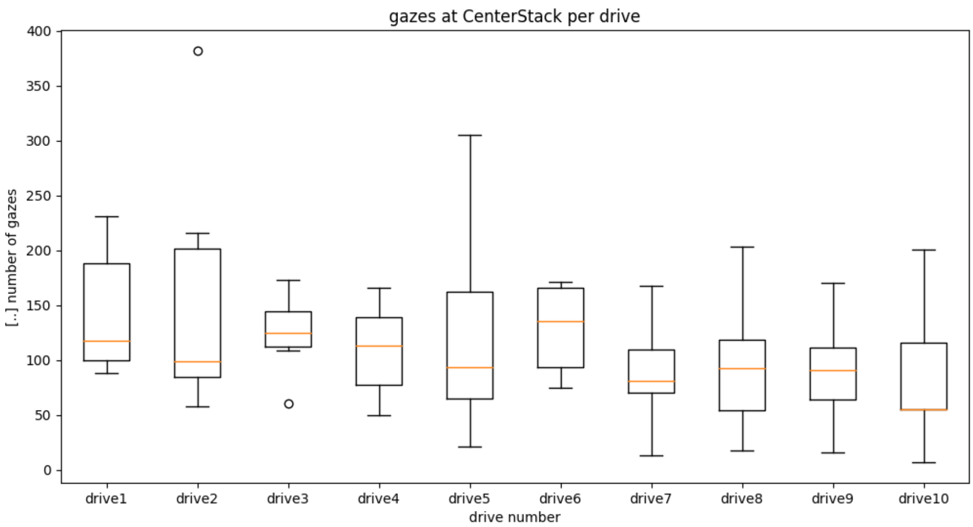
The gaze behaviour of drivers was analysed for the full drive and the highway section separately. Figure 4.8 illustrates the number of gazes at the Mediator display (gazes at Centre Stack) in the different drives (1 to 10) for the full drive (a) and the highway section only (b). It can be observed that in general there is a decreasing trend as the number of drives increases. It can also be observed that the number of gazes in drive 6 (short-time budget) is higher compared to the other drives which are according to expectations. The decreasing trend in the number of gazes might indicate that as drivers become more experienced with the system, they rely on the beeping sound instead of gazing at the Mediator display.

During the interviews, one driver (ID11) mentioned that they paid more attention to the audio cues compared to the visual cues on the display during the first drives.
"*I mostly hear the instruction instead of looking at the display. Especially on the second lap. When I get used to the Mediator system, for me, it's good enough to hear the warning sound*." (ID11)

Interestingly, the same driver stated after the last drive that he/she recognized the sounds but didn't process the different sounds anymore.
"*Actually, I did hear the audio, but it didn't go into my brain*" (ID11)



(a) Full drive



(b) Highway only

Figure 4.8 Number of gazes at the Mediator display (gazes at Center Stack) in the different drives with median values (yellow line). Top (a): for the full drive; Bottom (b): for the highway section only.

To test if these differences are statistically significant, repeated measures ANOVA was conducted to analyze if the differences among the four blocks of drives (blocks 1-4, 5, 6, 7-10) were statistically significant. No statistically significant difference was found at the 95% confidence level. However, for the full drive, a statistical difference was found at the 90% confidence interval.

Table 4.1 Repeated measures ANOVA analysis for the number of gazes among the different clusters of drives.

Source	Full drive	Highway only
ddof1	3	3
ddof2	15	15
F	3.1479	1.6728
p-unc	0.0562	0.2153
p-GG-corr	0.0701	0.2309
ng2	0.1680	0.0974
eps	0.8302	0.7509
sphericity	True	True
W-spher	0.6696	0.5212
p-spher	0.9167	0.7943

Table 4.2 Repeated measures ANOVA analysis for the number of gazes at the Mediator display for drives 1-6 (no distraction notification) and 7-10 (with distraction notification).

Source	Full drive (no distraction notification between drive 1-6 and 7-10)	Highway only (with distraction notification drive 1-6 and 7-10)
ddof1	1	1
ddof2	6	6
F	4.9363	5.4204
p-unc	0.0680	0.0587
ng2	0.2015	0.1651
eps	1.0	1.0

Despite the apparent downward trend in the number of gazes on the Mediator display, no statistical significance was found at the 95% confidence interval (but at the 90% confidence interval).

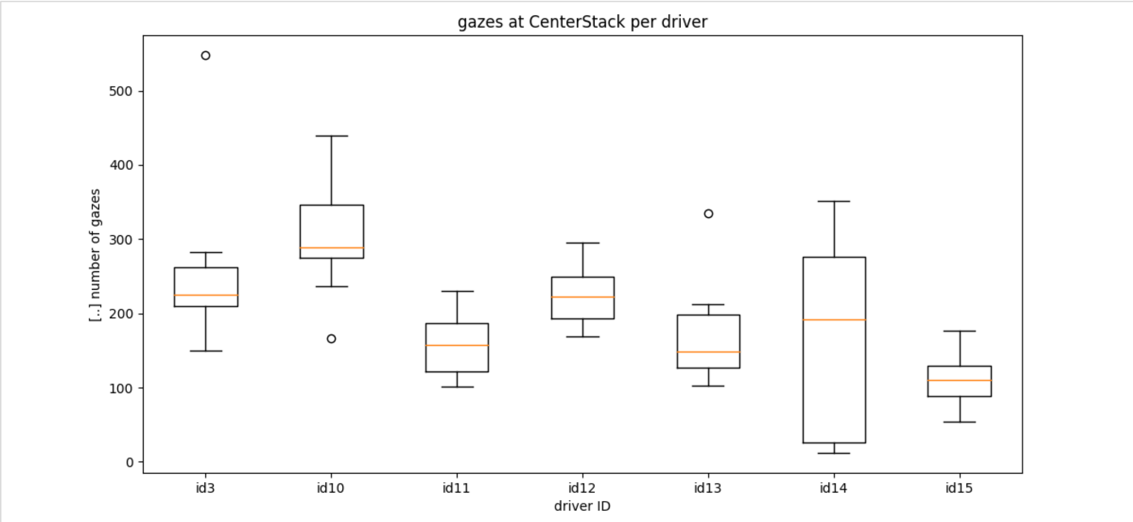


Figure 4.9 Individual results of the number of gazes on the Mediator display for each driver in each drive for the full drive with median value (yellow line).

From Figure 4.9 it can be observed that Driver ID10 looked at the Mediator display more often compared to the other drivers (switch-on time is longer, see the result of Automation Usage section 4.4.2.6), while Driver ID15 looked at the Mediator display less compared to the other drivers. Driver ID14 has a large variation in gaze frequency. When checking Driver ID14 gazes in the different drives (Figure 4.9) in the last four drives (i.e., drives 7-10 after the short time budget in drive 6), Driver ID14 barely looked at the Mediator display. It can be that the short time budgets in drive 6 have affected the Driver ID14 use of the Mediator system. It can be that Driver ID14 has experienced these short time budgets as annoying, and as a result, used less the Mediator system. Indeed, from the interviews it became clear, that driver ID14 was testing the limits of the distraction warning system (which was activated from drive 7 onwards) by, for instance, closing his/her eyes for short moments to trigger a distraction warning.

"*I think it was very I tried it a couple of times by picking up the phone, and then I think it was often quite correct. But I also tried to like, simulate nothing or like closing my eyes or crossing off, and then it did not activate at all*." (ID14)

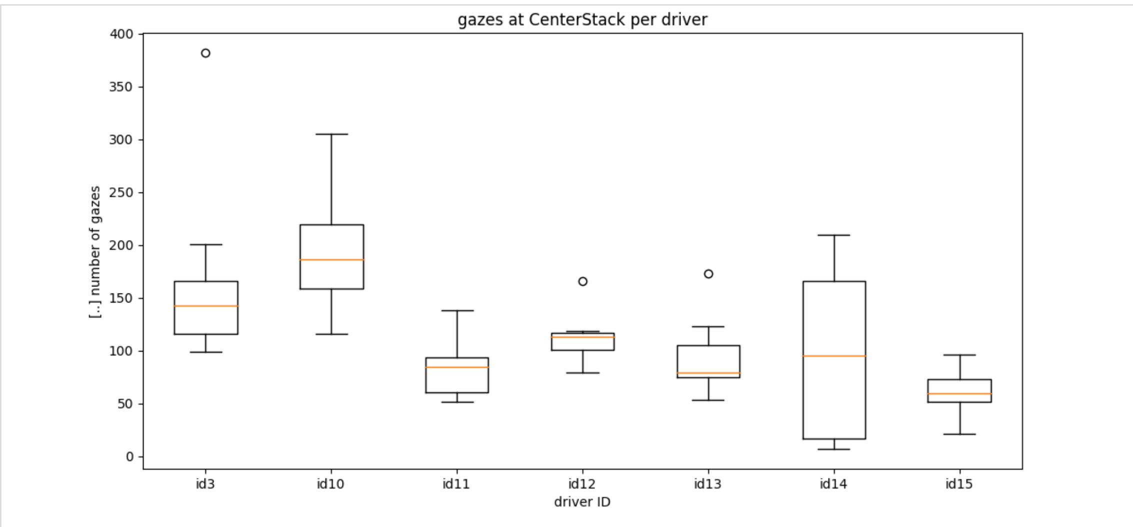
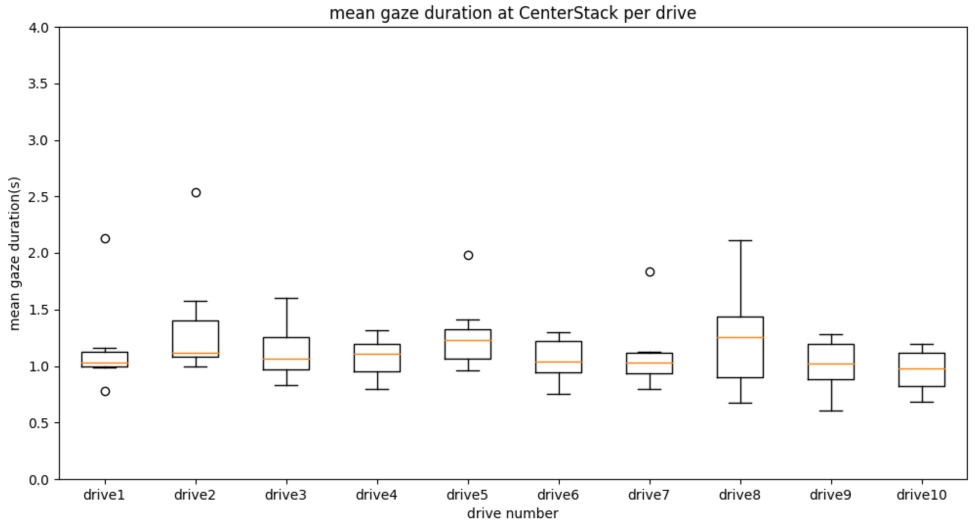
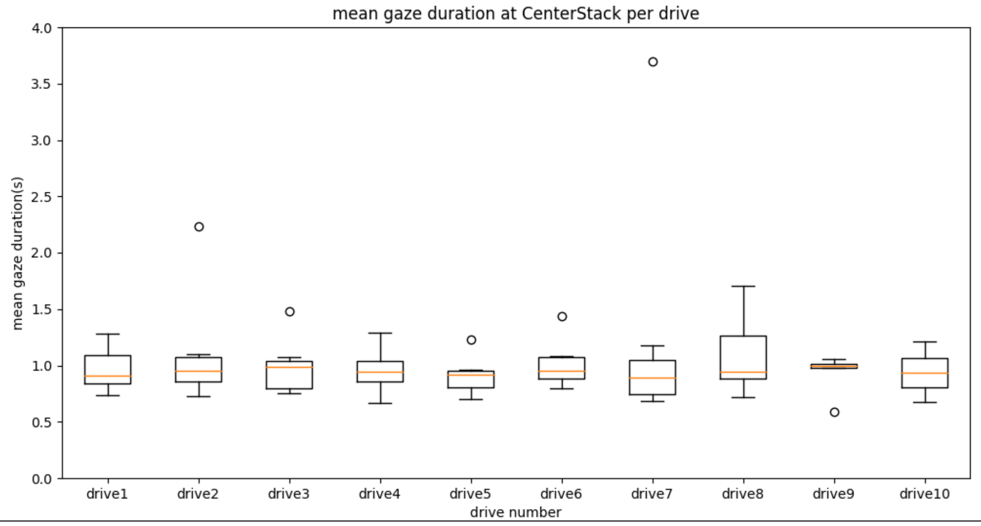


Figure 4.10 Individual results of the number of gazes on the Mediator display for each driver in each drive for the highway only, with median values (yellow line).

Similar observations for Drivers ID10, ID14, and ID15 can be made from Figure 4.10 regarding the highway section only for the full drive. Driver ID ID10 looked at the Mediator display more compared to the other drivers, while Driver ID15 looked at the Mediator display less compared to the other drivers, and large variation for Driver ID14.



(a) Full drive



(b) Highway only

Figure 4.11 Gaze duration on the Mediator display for each driver in each drive with median values (yellow line). Top (a): for the full drive. Bottom (b): for the highway section only.

Figure 4.11 shows that no apparent differences can be observed between drives 1-6 (no distraction notification) and 7-10 (distraction notification) for both the full drive and the highway section only.

4.4.2.2. Do drivers look at the screen less in drive 5 (No Mediator system)?

To analyze whether drivers look at the screen less in drive 5 (baseline) when Mediator did not show the time budget compared to when Mediator displayed the time budget in drives 1-4 and 7-10, the mean number of gazes at the screen in drive 5 was compared to the mean number of gazes at the screen in drives 1-4 and 7-10. Drive 6 was excluded from this analysis as it included short-time budgets and frequent Transition of Control (ToC) suggestions. Figure 4.12 (left) shows that the mean number of gazes at the Mediator display in drives 7-10 is less compared to drives 1-4 and drive 5 (No Mediator). Figure 4.12 (right) shows the results of a boxplot of the same comparison. Statistical testing of the differences in the mean number of gazes at the screen did show a significant difference in the number of gazes at the Mediator display for the full drive but not for the highway (although the same trend as the full drive can be observed). The results are displayed in Table 4.3.

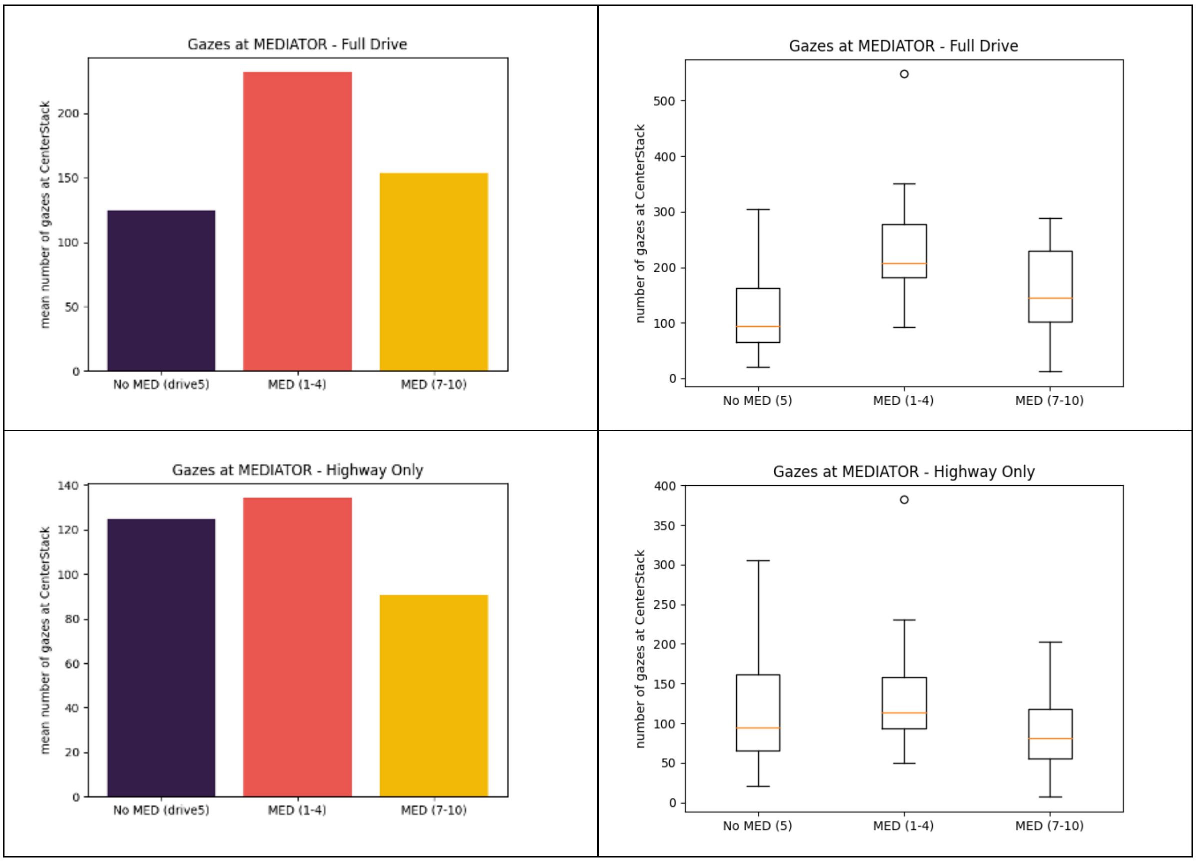


Figure 4.12 Mean number of gazes at the screen (left) and Boxplot (right) of the number of gazes at the screen, with median value (yellow line in the boxplot), without Mediator, with Mediator in drives 1-4, and with Mediator in drives 7-10.

Table 4.3 Repeated measures ANOVA of the mean number of gazes amongst the different clusters of drives (1-4, 5, 7-10).

Source	Full drive	Highway only
ddof1	2	2
ddof2	10	10
F	4.5746	2.4373

p-unc	0.0388	0.1373
p-GG-corr	0.0411	0.1439
ng2	0.1786	0.1004
eps	0.9636	0.9137
sphericity	True	True
W-spher	0.9622	0.9056
p-spher	0.9259	0.8201

Post hoc analysis was conducted to further investigate amongst which clusters of drives (1-4, 5, 7-10) the difference is significant. The results displayed in Table 4.4 show that the results are significantly different between clusters 1-4 and 7-10 for both the full drive and the highway section only.

Table 4.4 Post hoc analysis of the number of gazes amongst the different clusters of drives (1-4, 5, 7-10) for the full drive and the highway section only.

	Full drive			**Highway only**		
	0	**2**	**1**	**0**	**2**	**1**
A	med1-4	med7-10	med1-4	med1-4	med7-10	med1-4
B	med7-10	no med	no med	med7-10	no med	no med
Mean(A)	232.1111	153.8888	232.1111	134.3703	90.5185	134.3704
Mean(B)	153.8888	213.4285	213.4285	90.5185	124.8571	124.8571
diff	78.2222	-59.5396	18.6825	43.8518	-34.3386	9.5132
se	25.1133	48.2647	49.1038	16.8932	37.5867	38.4954
T	3.1147	-1.2336	0.3805	2.5958	-0.9136	0.2471
df	51.1416	7.6783	8.2033	49.1175	7.0224	7.7083
pval	0.0083	0.4693	0.9240	0.0327	0.6495	0.9670
hedges	0.8354	-0.5108	0.1575	0.6962	-0.3783	0.1023

To further investigate whether there is a difference between the beginning and end of drive 5 (No Mediator) in the number of gazes to the Mediator display, the number of gazes in both driving directions (i.e., way there and way back) was compared, for the full drive and the highway section only Figure 4.13 presents the results. No statistical difference was found in the median number of gazes at the Mediator display between one direction and the other direction when considering the full drive, however, a significant difference was found for the highway section only – drivers looked fewer times at the Mediator display on the way back compared to the way there.

This indicates perhaps that on the way there, drivers were missing or anticipating messages, but learned over time that the system is not active. When checking the total time looking at the Mediator display it can be noticed that the drivers looked less time on the way back compared to the way there. To check whether the total time looking at the Mediator display was affected by the driving speed or congestion level on the road (i.e., longer travel times), the mean, median and standard deviation of the driving speeds in drive 5 on the way there and the way back were calculated for the full drive and the highway section only and are presented in Table 4.5. No direct relation was found between the different driving conditions and the total time looking at the screen, as the driving speed on the way back was lower (i.e., the driving time was longer) than on the way there, but the total time looked at the screen on the way back was lower than on the way there.

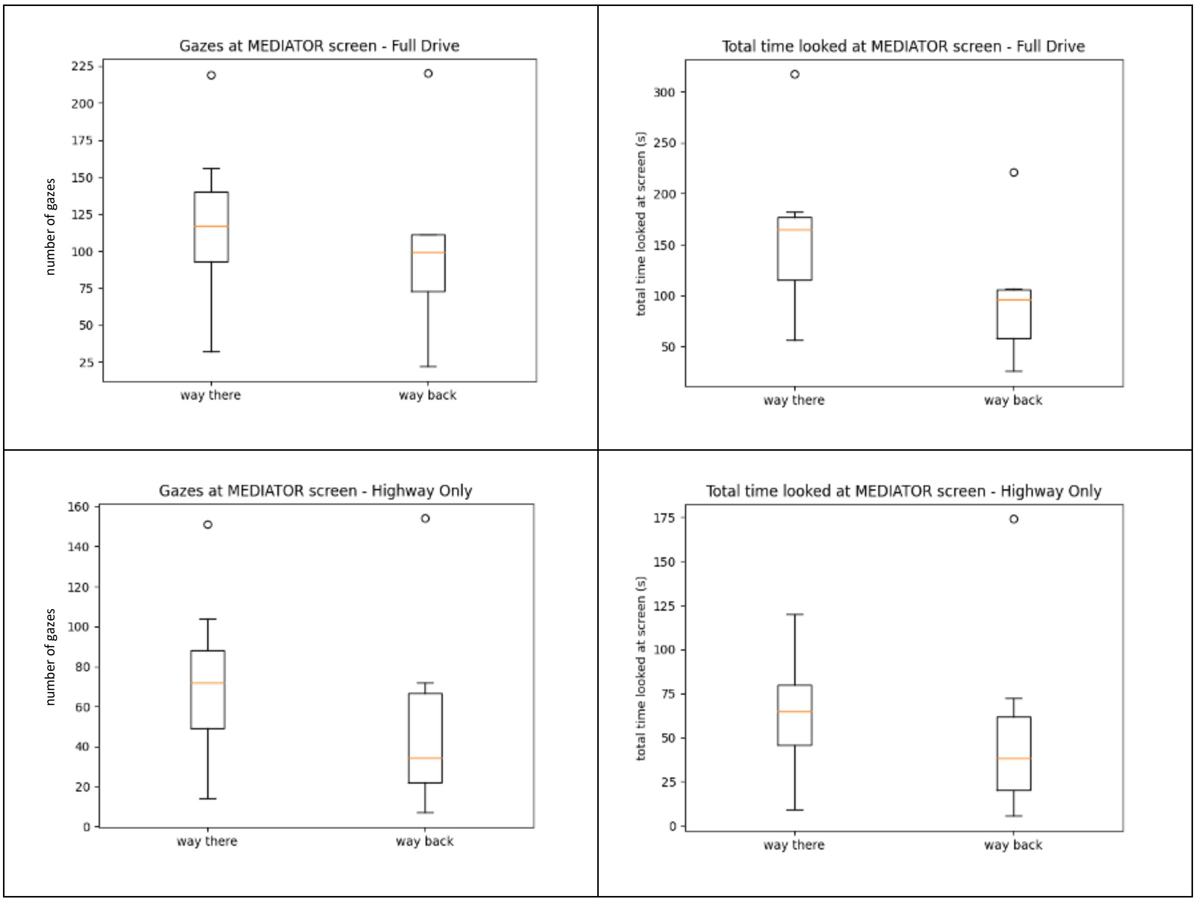


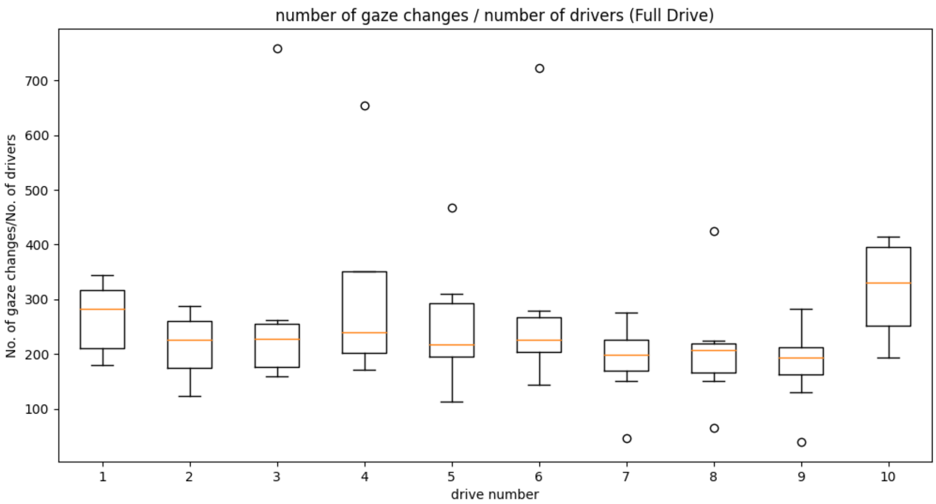
Figure 4.13 Number of gazes (left) and total time looked at the Mediator display (right), in drive 5 – no Mediator on the way there versus on the way back. Median values indicated by the yellow line.

Table 4.5 Mean, median and standard deviation of the driving speed in drive 5 on the way there and back for the full drive and the highway section only.

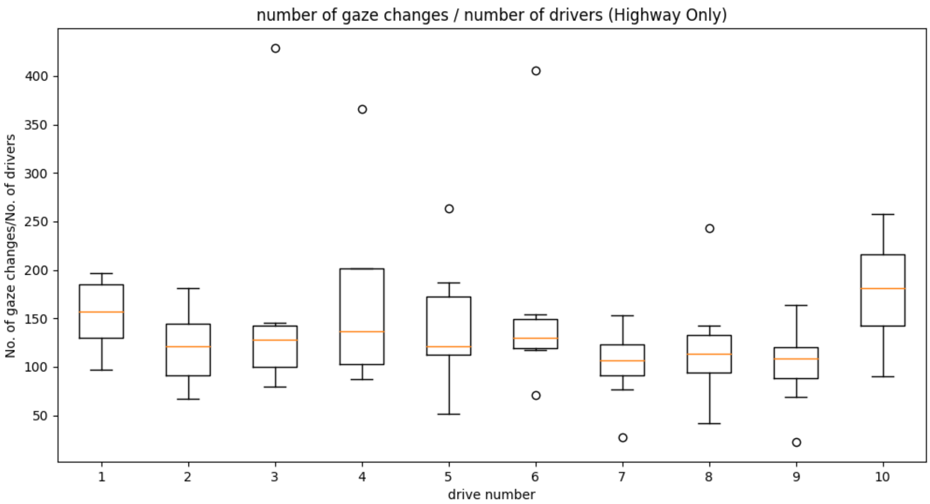
Driving speed	Full drive			Highway only		
	Mean (km/h)	Median (km/h)	Std. (km/h)	Mean (km/h)	Median (km/h)	Std. (km/h)
speed – way there	82.6	93.0	31.2	97.6	102.5	23.2
speed – way back	77.3	83.0	35.6	86.5	102.0	10.7

4.4.2.3. Does driver gaze behavior change over time (i.e., the number of gaze direction changes)?

The number of gaze direction changes in the full drive and highway section only were analyzed to investigate if driver gaze behavior changes over time. From the results presented in Figure 4.14, it can be noticed that there is an increase in the number of gaze changes in the last drive in both the full drive and the highway only. This could indicate that drivers started to rely more on the sound once experienced with the Mediator system instead of the visual message on the Mediator display.



(a) Full drive



(b) Highway only

Figure 4.14 Number of gaze direction changes normalized to the number of runs per drive and median values (yellow line).

The number of gaze changes per second (i.e., the number of gaze changes divided by the drive length) for the highway section was further analyzed by comparing the results when the autopilot was ON versus OFF. Figure 4.15 depicts the results of the number of gaze directions normalized to the drive length when the autopilot was OFF and when it was ON for the different drives. As expected, the number of gaze changes per second when the autopilot was OFF is lower than when the autopilot was ON, indicating that drivers scan their environment more with autopilot is ON.

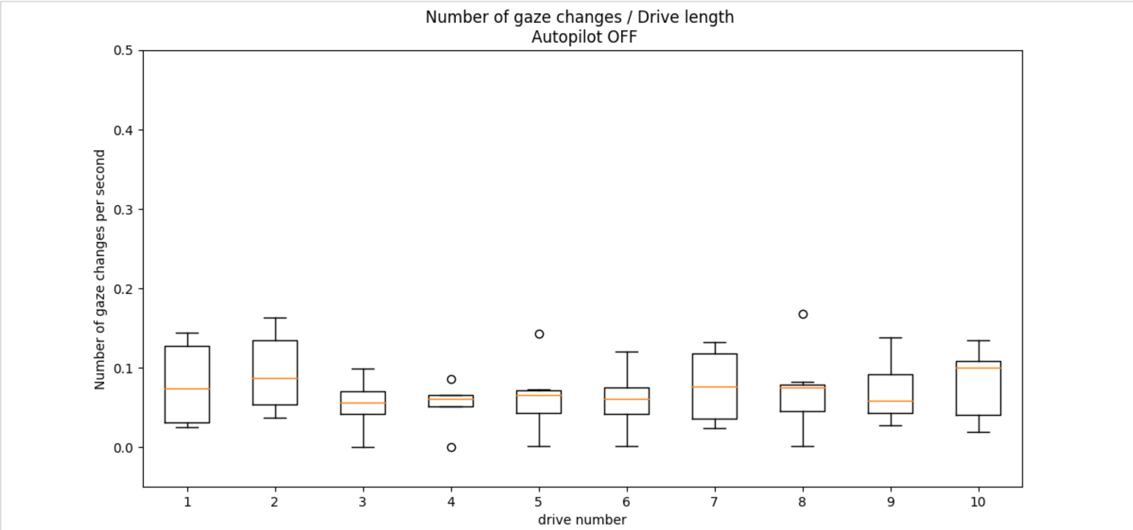


Figure 4.15 Median (yellow line) number of gaze changes per second and drive length for when the autopilot was OFF (top) and when it was ON (bottom) for the different drives on the highway section only. Left – Mean number of gazes at the screen; and Right: Boxplot of the number of gazes at the screen, without Mediator, with Mediator in drives 1-4, and with Mediator in drives 7-10.

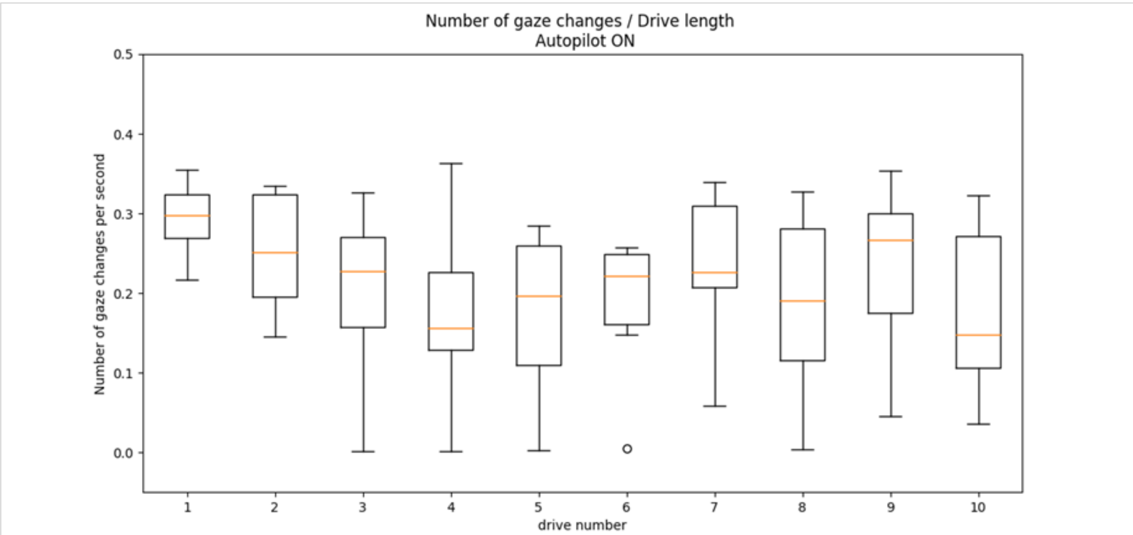


Figure 4.16 Median (yellow line) number of gaze changes per second from all drives when the autopilot was ON versus OFF.

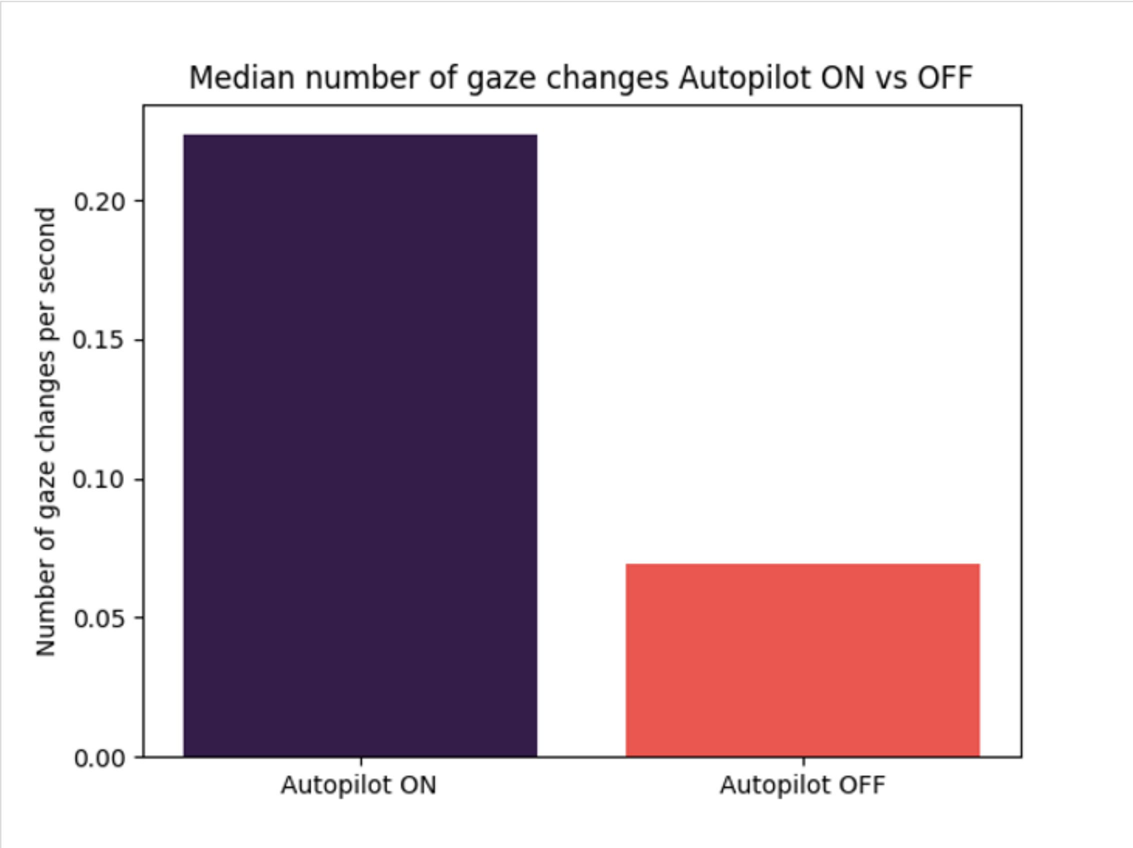


Figure 4.17 Median number of gaze changes per second for all drives when the autopilot was OFF and when it was ON.

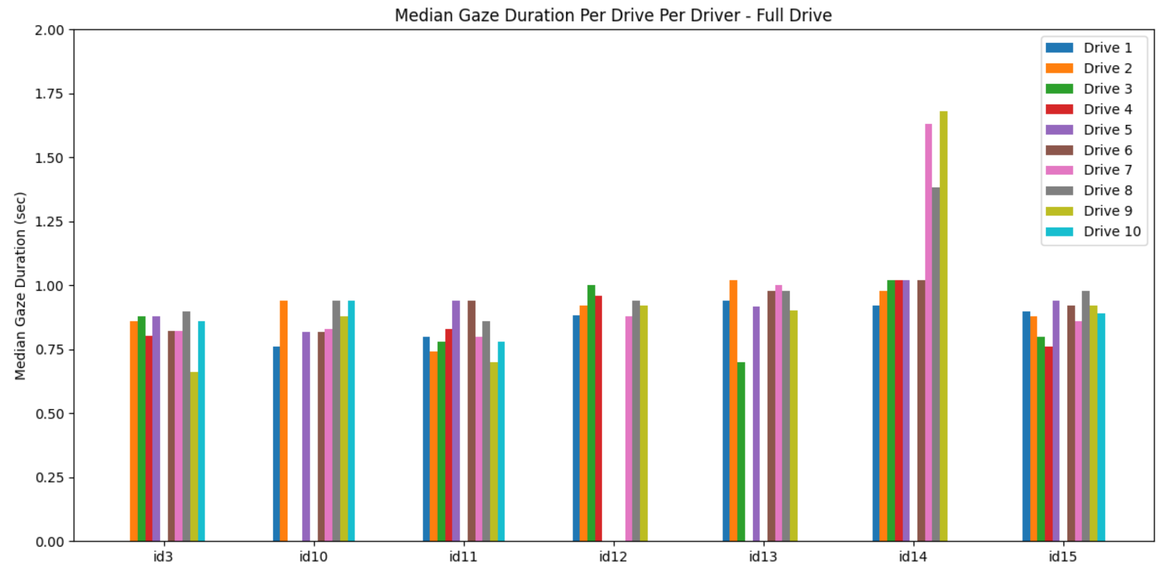
ANOVA analysis results in Table 4.6 show that this difference is statistically significant at the 95% confidence level.

Table 4.6 Repeated measures ANOVA analysis results of the comparison of the median number of gaze changes per second for all drives when the autopilot was OFF and when it was ON.

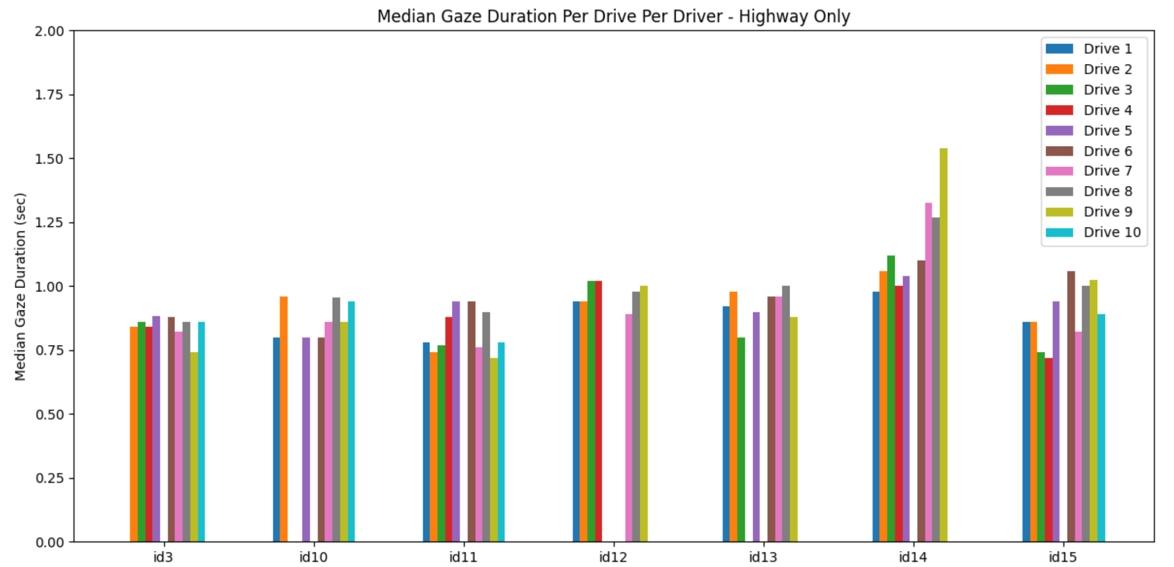
Source	Autopilot_status
ddof1	1
ddof2	6
F	23.5995
p-unc	0.0028
np2	0.6674
eps	1.0

4.4.2.4. Are there clusters of drivers in terms of gaze behavior?

Figure 4.18, Median gaze duration on the Mediator display for the full drive (a) and the highway section (b) in each of the 10 drives for each driver separately, presents the median gaze duration on the Mediator display (Center stack) in each of the 10 drives for each driver separately, for the full drive (top figure) and the highway section only (bottom figure). No major differences can be noticed between the different drivers and drives, except for driver ID14 who exhibited a relatively higher median gaze duration on the Mediator display in the last three drives. Notice that earlier (RQ2.1) it was found that Driver ID14 exhibited a large variation between the different drives with respect to the number of gazes at the Mediator display.



(a) Full drive



(b) Highway only

Figure 4.18 Median gaze duration on the Mediator display for the full drive (a) and the highway section (b) in each of the 10 drives for each driver separately.

4.4.2.5. Was the gaze behavior different (i.e., gaze changes, gaze length) in normal drives (1-4), special cases (5, 6), or distraction-system-active (7-10)?

The average number of gaze switches and the average gaze durations were calculated per drive cluster (i.e., drives 1-4, drive 5, drive 6, drives 7-10). Figure 4.19 presents the results for both the full drive and for the highway section. It can be observed that in drive 6 (short time budgets) drivers had higher number of gaze switches. No other apparent differences can be observed between the different clusters.

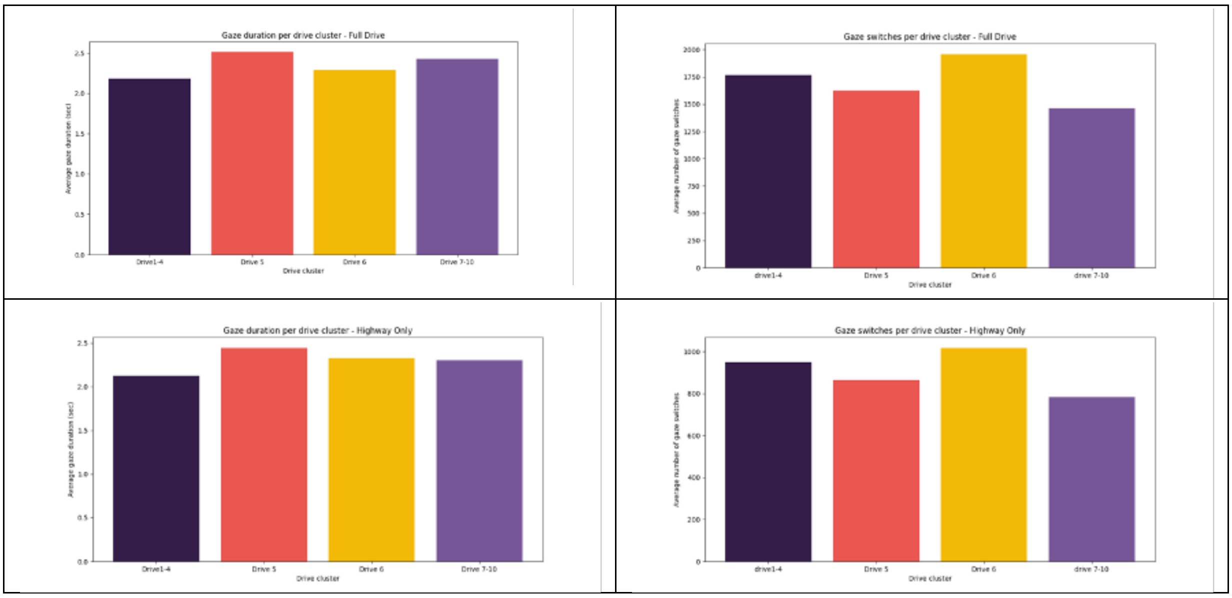
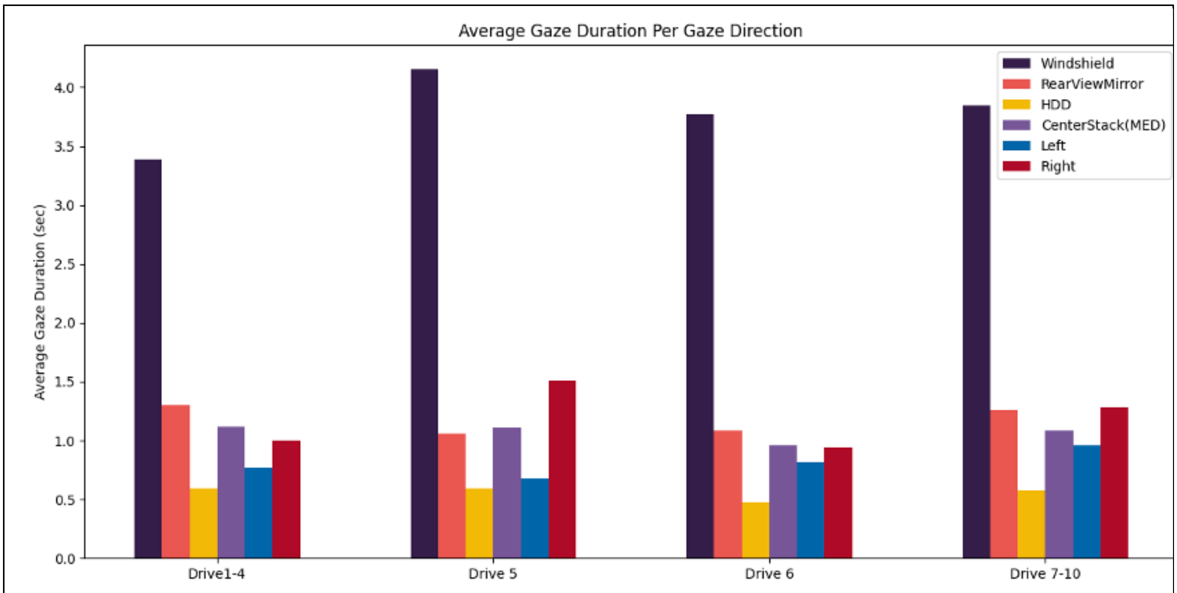
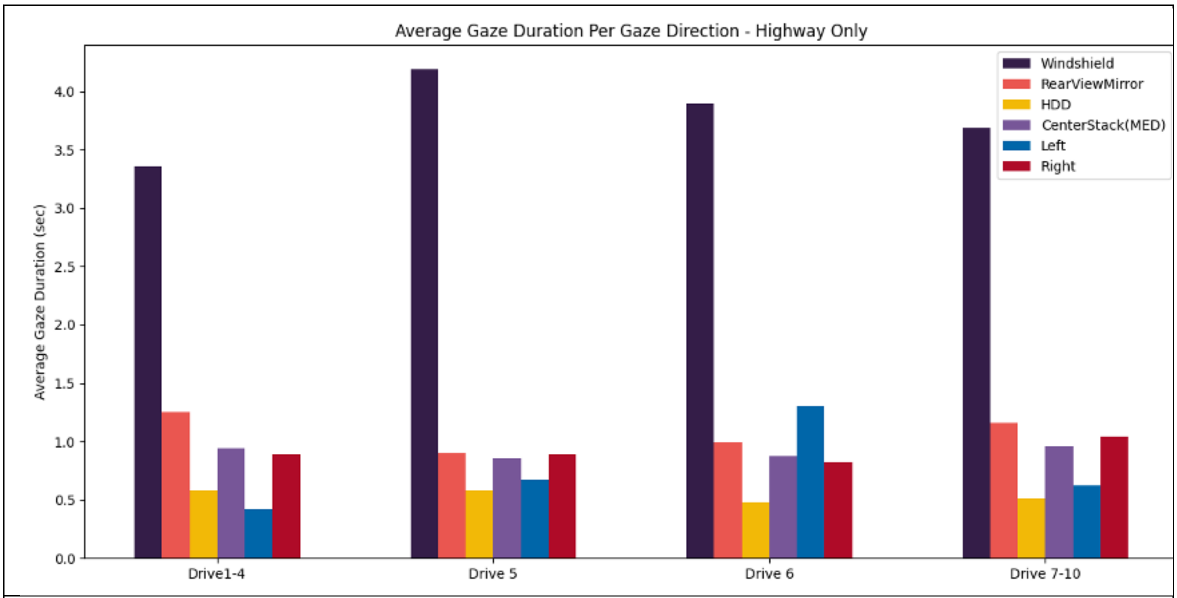


Figure 4.19 The average number of gaze switches and the average gaze durations per drive cluster for the full drive and the highway section.

The average gaze duration per gaze direction was also analyzed for the different clusters in the full drive and the highway section. The results are shown in Figure 4.20. The average gaze duration per gaze direction for the different clusters in the full drive (a) and the highway section (b). Again, no apparent differences can be noticed between the different clusters of drives.



(a) Full drive



(b) Highway only

Figure 4.20 The average gaze duration per gaze direction for the different clusters in the full drive (a) and the highway section (b).

4.4.2.6. Automation Usage

This section analyses the data regarding automation usage in the different drives.

4.4.2.7. Is there a difference in mean take-over time (i.e., switching to manual driving) between drivers and drives, after a suggestion from MEDIATOR?

Figure 4.21 presents the mean take-over time for each driver and drives. It is noticed that Driver ID3 seems to have a relatively shorter take-over time compared to the other drivers – this driver has driven for 16 runs – i.e., more than other participants – this could indicate a learning effect.

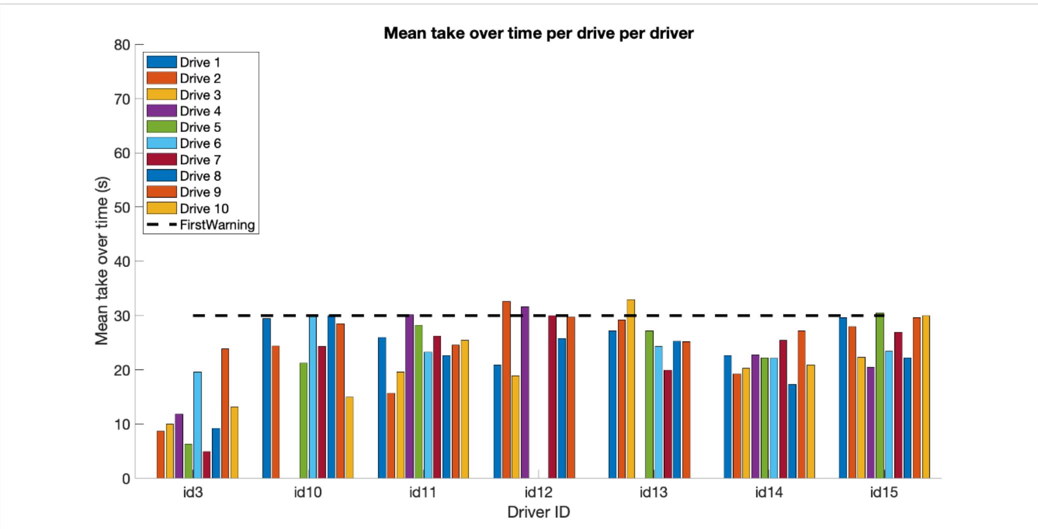


Figure 4.21 Mean take-over time for each driver and each drive.

4.4.2.8. If there's no time budget (Drive 5), is automation turned off later?

In drive 5 the Mediator system did not display the time budget for the drivers and therefore did not receive any information regarding the time budget availability. Figure 4.22 presents a box plot of the mean take-over time with (drives 1-4, 6, 7-10) and without Mediator (drive 5). As can be seen from the boxplot, the median values in the take-over times are very similar between the two conditions (with and without Mediator), however, there is higher variability in the drive without Mediator.

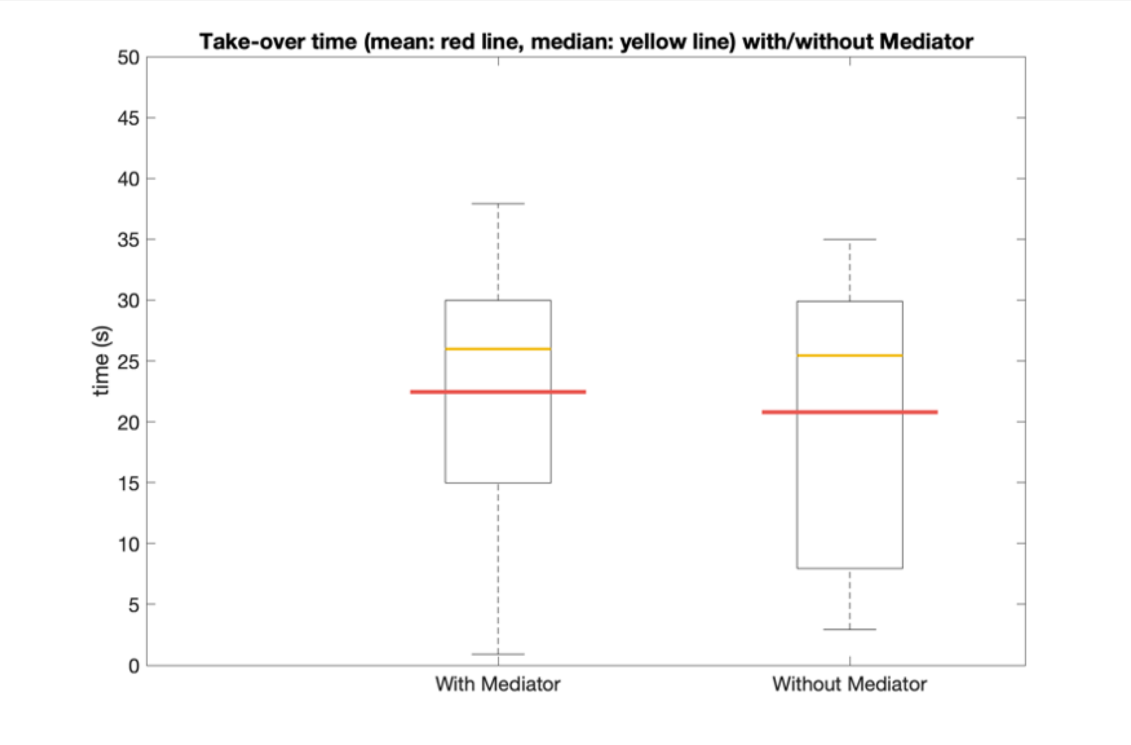


Figure 4.22 Box plot of the mean (red line) and median (yellow line) take-over time with and without Mediator.

4.4.2.9. Is there a difference in take-over time (i.e., switching to manual driving) between different conditions?

The take-over time of drivers was compared between the condition of 'end of ODD' and 'roadworks. Figure 4.23 shows that the mean take-over time in roadwork conditions, although lower compared to the end of ODD, is not significantly different (Table 4.7 presents the results of the statistical comparison). The lower value is probably because the roadworks are clearly visible to the drivers as well as possible road signage indicating upcoming roadworks which already could alert drivers.

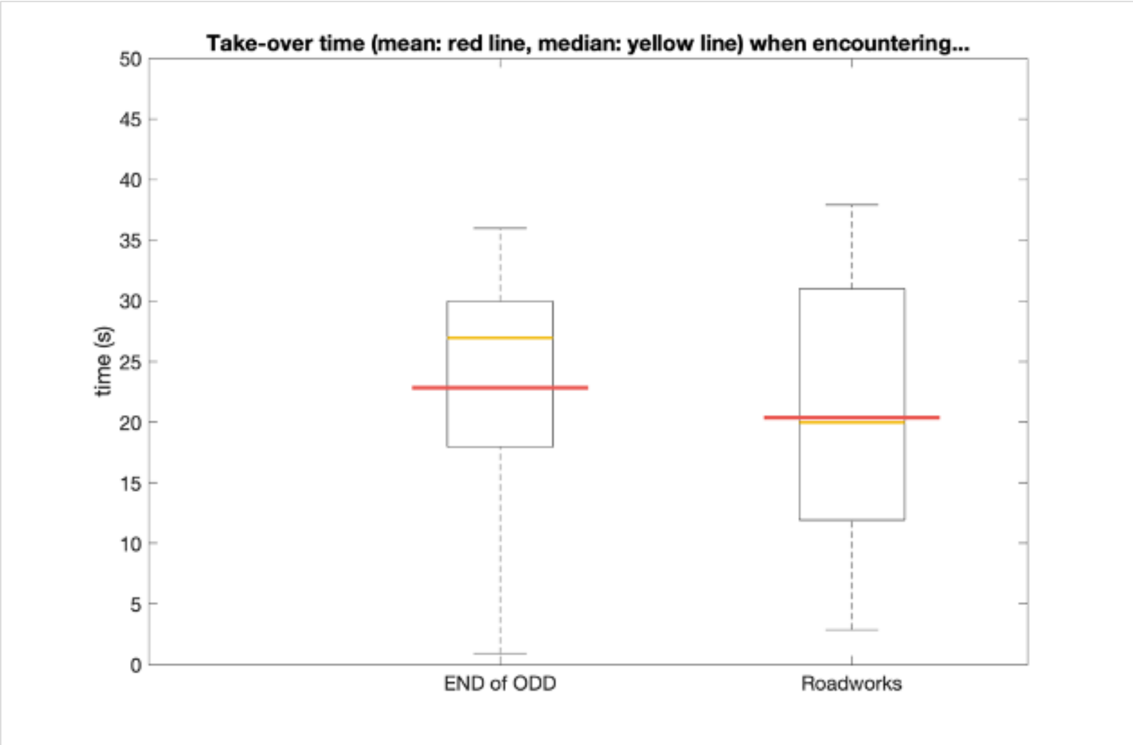


Figure 4.23 Mean (red line) and median (yellow line) take-over time (seconds) when encountering the end of the ODD and roadworks.

Table 4.7 Statistical comparison between the mean take-over time (seconds) when encountering the end of the ODD and roadworks.

Source	Groups	Error	Total
SS	277.3	26041	26318.3
df	1	270	271
MS	277.286	96.448	
F	2.87		
Prob>F	0.0911		

4.4.2.10. Distraction

In drives 7 to 10, in addition to the fact that the Mediator system was active, drivers also received warnings when they were distracted while driving, while in drives 1 to 4 drivers did not receive a warning when they were distracted, however, such distraction events were recorded.

4.4.2.11. Is there a difference in Distraction when distraction warnings were active vs inactive?

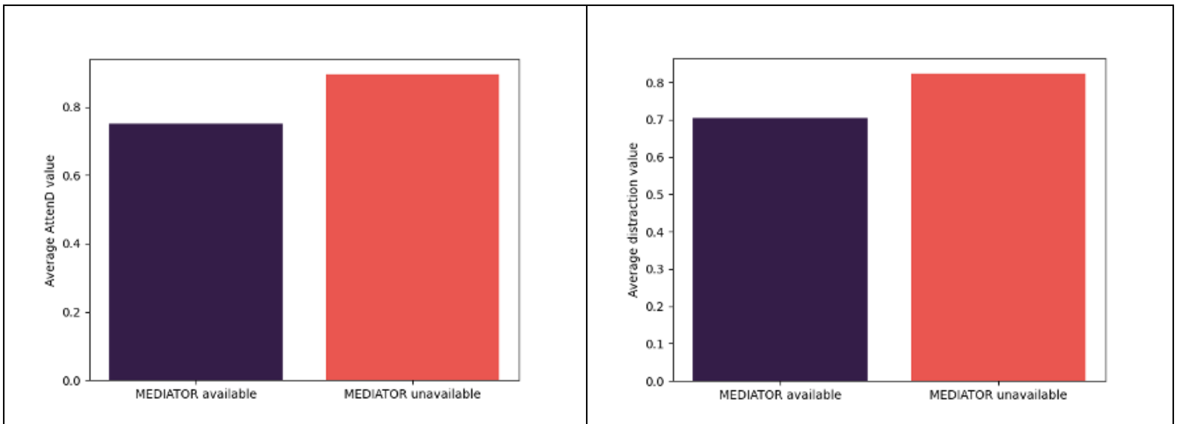
To test if the warning distraction affected the drivers, Distraction variable was compared between drives 1-4 and 7-10. Repeated measures ANOVA results presented in Table 4.8 indicate that there is no significant difference between when the distraction warnings were active vs inactive.

Table 4.8 Repeated measures ANOVA results when the Distraction warnings were active vs inactive.

Source	Distraction
ddof1	1
ddof2	6
F	0.337915
p-unc	0.582198
np2	0.029108
eps	1.0

4.4.2.12. Is there a difference in visual Distraction when the Mediator system is AVAILABLE vs UNAVAILABLE?

To test if there is a difference in the mean visual distraction value when Mediator was available versus unavailable, the *AttenD* and *Distraction* were compared between these two conditions. The *AttenD* value can range between 0-2. It is assumed that the driver has a buffer of 2s when looking away from the road. When looking away, the buffer is depleted and when looking back to the road the buffer fills up. Therefore, as the value is lower the driver is more likely to be distracted. If the buffer runs empty the driver is classified as distracted. The Distraction is defined as duration of eyes off road and threshold (>2s) was used to issue a warning.



a) AttenD	b) Distraction

Figure 4.24 Average value of AttenD (a) and Distraction (b) when Mediator was available versus unavailable.

It appears a difference between the two conditions. To test if this difference is significant, a repeated ANOVA test was conducted, and the results are presented in Table 4.9 and Table 4.10.

Table 4.9 A repeated measures ANOVA for the average AttenD value when Mediator was available versus unavailable.

Source	Mediator_available
ddof1	1
ddof2	6
F	7.040212

p-unc	0.037859
np2	0.311164
eps	1.0

Table 4.10 Repeated measures ANOVA for the average Distraction value when Mediator was available versus unavailable.

Source	Mediator_available
ddof1	1
ddof2	6
F	7.040212
p-unc	0.037859
np2	0.311164
eps	1.0

The results indicate that the differences in the average *AttenD* and *Distraction* values when Mediator was available versus unavailable are significant, with a higher likelihood that the driver is not looking to the forward view when driving with the Mediator system available. This result is expected, as when drivers inspect the Mediator system their eyes are away from the forward view. Although the difference is significant, it is not likely to impose higher risks compared to driving without Mediator, as the values are still close to each other.

4.4.2.13. Is there a difference in Distraction when pilot assist is ON vs OFF?

To test if there is a difference in the mean visual distraction value when pilot assist was ON versus OFF, the AttenD and Distraction were compared between these two conditions. Figure 4.25 presents the results.

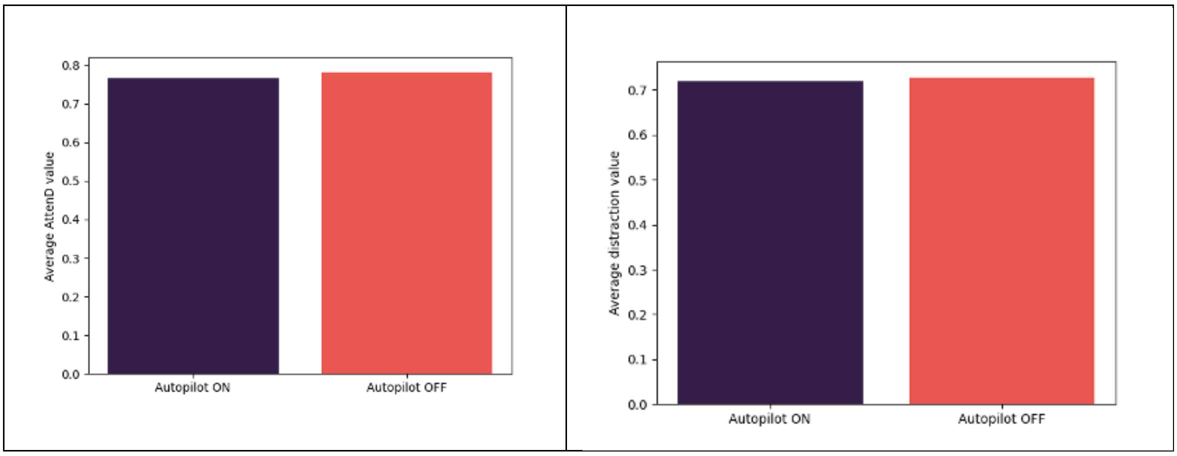


Figure 4.25 Average value of AttenD (a) and Distraction (b) when pilot assist was ON versus OFF.

No apparent difference can be seen between the two different conditions. Statistical tests presented in Table 4.11 and Table 4.12 confirm these observations.

Table 4.11 Repeated measures ANOVA for the average AttenD value when pilot assist was ON versus OFF.

Source	pilot_assist
ddof1	1
ddof2	6
F	0.015756
p-unc	0.90421
np2	0.000154
eps	1.0

Table 4.12 Repeated measures ANOVA for the average Distraction value when pilot assist was ON versus OFF.

Source	pilot_assist
ddof1	1
ddof2	6
F	0.084286
p-unc	0.781343
np2	0.000821
eps	1.0

A closer look into the single item measurements as well as the interview data revealed that drivers rated Mediator's reliability on average as good. Figure 4.26 shows that on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree drivers tended to agree to the statement "Mediator is reliable", especially after the first drives. In drive 5 the reliability was rated a bit lower which can be explained by missing recommendations and time budget information provided by Mediator. Drive 6 shows the highest variance in the rating indicating that recommending partly automated driving also for quite short sections is interpreted as an indicator for high reliability by some drivers and as indicator for a lower reliability by other drivers. During the last drives, the reliability was partly rated as quite low. One explanation might be the distraction warning which gave false alarms and, hence, reduced the perceived reliability of the system (e.g., during drive 7 and 9). Nevertheless, the results show that the overall reliability rating is not affected too strongly (e.g., drive 8 and 10), indicating that drivers can distinguish between the reliability of the distraction warning and reliability of the whole Mediator system (including, e.g., the time budget). The interview results show that the displayed time budget was rated as reliable information (see section regarding RQ1).

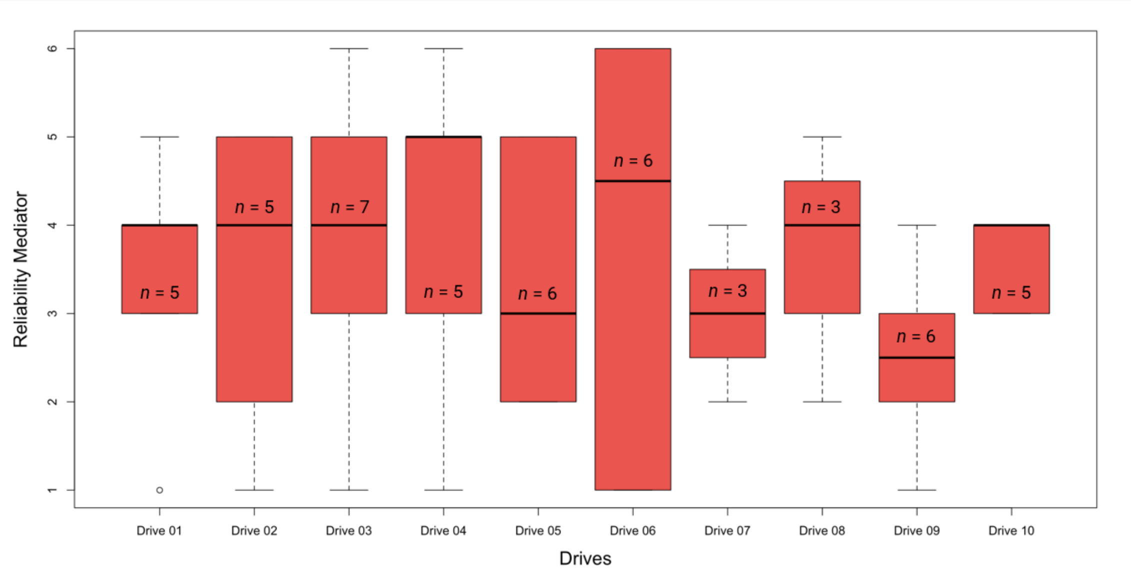


Figure 4.26 Drivers answers to the statement "Mediator is reliable." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree for all 10 drives. n shows the number of valid values per drive.

Some drivers are more critical compared to others (see Figure 5.1 in Appendix 0). Driver ID 12 and driver ID13 rated Mediator's reliability quite low in all drives. Also, driver ID3 rate Mediator's reliability on average slightly below the mean scale value (i.e., tendency to disagreement). Driver ID11 rated Mediator's reliability the highest.

Interview results indicate that after the first two drives all seven drivers agreed that Mediator is reliable. Drivers mentioned the consistency like giving the same recommendations in comparable situations (ID13, ID14, ID15), the correctly working functions (ID10) especially the time budget (ID12, ID15) and the timing of the audio warnings (ID11) as indicators for a good reliability.

"*It was consistent. I think it gave the same recommendation both times of two test drives.*" (ID13)

"*I think every time you go into the highway, exit the highway and engage, I mean, activate the ad mode, disengage the ad mode, I always get the warning. So I think the system is quite reliable.*" (ID11)

Driver ID12 stressed that in his/her opinion the Mediator system is not really necessary when using an L2 system, hence, he/she didn't elaborate deeper on the system's reliability except the time budget.

"*Maybe I can say the timing for the end of automation and start on manual and so on seemed reliable and useful. And if it was a full automation system, I would have found that information useful.*" (ID12)

Two drivers (ID3, ID10) mentioned some minor inconsistencies in the system's behaviour. Further, driver ID13 stated that, although the given recommendations were reliable, he/she did not fully agree with Mediator's decisions.

“*I didn't agree completely with the recommendation. Like when entering Alingsås, you enter a red light, but the road is still rather good. And I didn't agree with the recommendation to stop using the system, so then it becomes quite disturbing*.” (ID13)

After the last drive, drivers reported that they experienced consistent timings and messages as indicators for a good reliability (ID14, ID15), although they mentioned some inconstancies (ID10, ID15).

“*Yes. I experienced, and I don't really know if that is intentionally, but sometimes when I got the warning to take over again, sometimes it really disappeared after like three blinks, but sometimes it just kept going*.” (ID15)

Especially, the distraction warning was mentioned as less reliable (ID11, ID13)

“*Besides, some strange distraction warning, I think, is everything is what I expected*.” (ID11)

“*So and also the distraction warning was not very confident with both false positives and false negatives*.” (ID13)

Driver ID 3 and driver ID12 rated Mediator’s reliability quite low after the last drive because of several system shutdowns they experienced.

“*The system turns off a lot. I think in the beginning when I took my first drive, so it was no problem. I can have the system on all the time, but now it just turns off and I don’t know why*” (ID3)

“*There have been, I mean, unreliable situations of all sorts. The underlying system stopped working*.” (ID12)

4.4.3. RQ3: What timings are most appreciated by the drivers in repeated exposures to the same static ODD change?

4.4.3.1. Did some drivers switch back to manual driving later than others?

Figure 4.27 presents a boxplot of the time it took each driver to switch back to manual driving (analysis for the highway only). As can be clearly seen driver ID3 switched back to manual driving earlier than the other drivers.

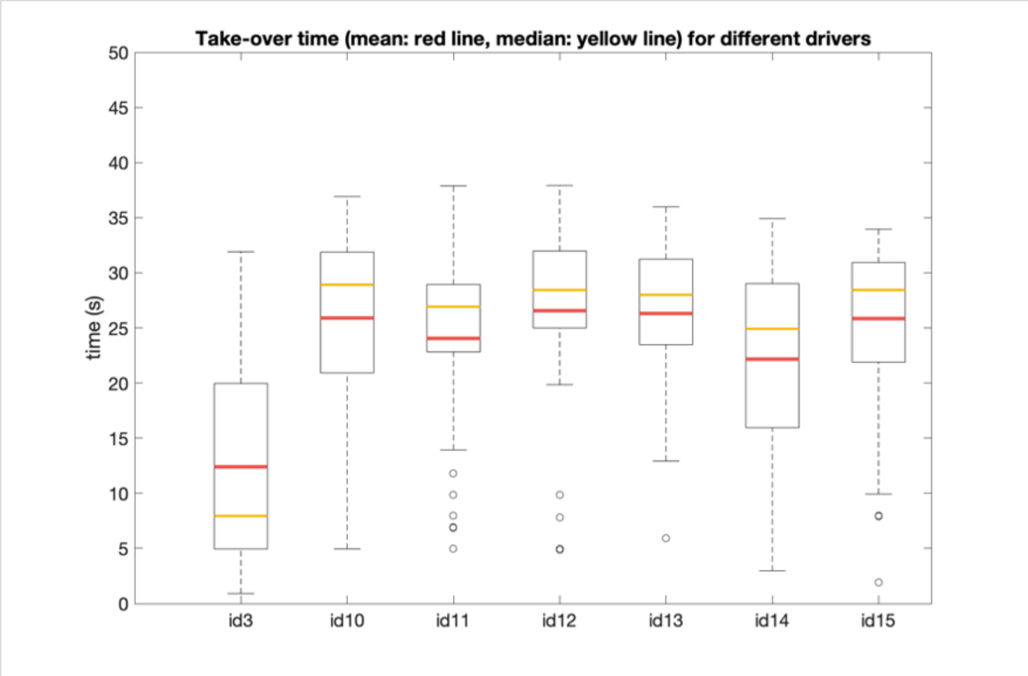


Figure 4.27 Boxplot of the time it took drivers to switch from automated to manual driving – Highway drive. Mean (red line) and median (yellow line) values are also reported.

The ANOVA analysis results (see Table 4.13) show significant differences between drivers.

Table 4.13 ANOVA analysis of the time to switch from automated to manual driving – full drive.

Source	Groups	Error	Total
SS	7765	18553.3	26318.3
df	6	265	271
MS	1294.17	70.01	
F	18.48		
Prob>F	6.16653e-18		

A post hoc analysis was conducted to analyse which drivers significantly differ from each other with respect to the time to switch from automated to manual driving.

4.4.3.2. Did some drivers switch on automation when automation becomes available much later than others do?

Figure 4.28 illustrates the boxplot results per driver for the full drive. As can be clearly seen driver ID10 has a distinct pattern from the other drivers. Table 4.14 of the ANOVA analysis also shows that there is a significant difference between the groups.

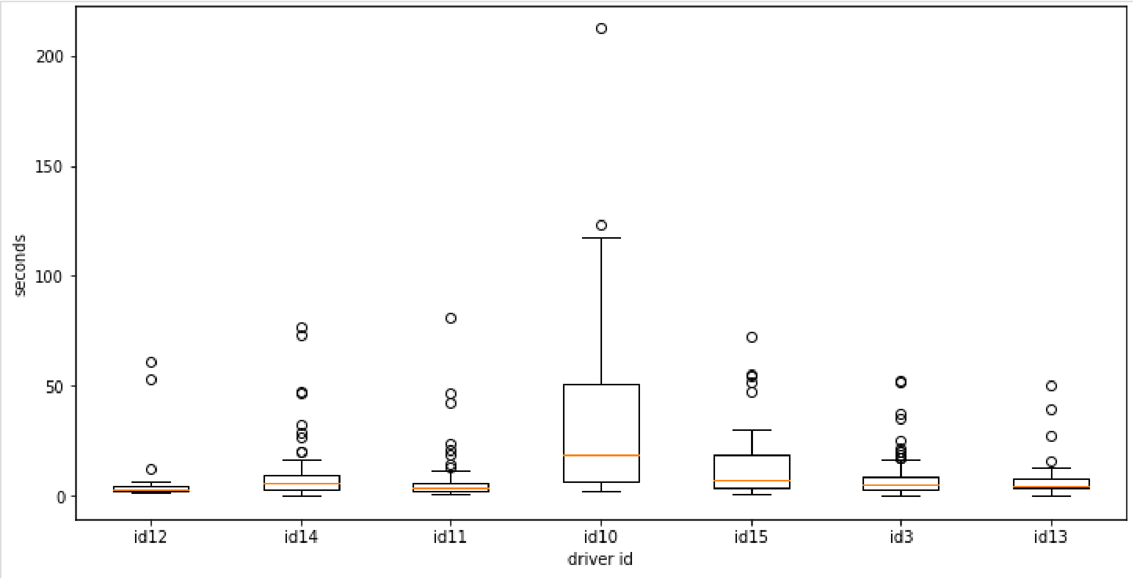


Figure 4.28 Boxplot of the time it took drivers to switch from manual to automated driving measured from the time of the first message sent that automation is available – full drive. Median value reported by yellow line.

Table 4.14 ANOVA analysis of the time to switch from manual to automated driving – full drive.

Source	Driver
ddof1	6
ddof2	314
F	12.266871
p-unc	2.093870e-12
np2	0.189889

Clearly, driver ID10 has a large variation in his/her time to switch to automation and took a significantly longer time to switch to automated driving from the moment automation became available compared to the other drivers.

During the interviews driver ID10 stated that he/she learned over time that he/she was not forced to respond to Mediator's recommendations to switch to automated driving mode and, hence, chose himself/herself when to change driving mode.

"*In the beginning, when the assisted driving was available, I was in the impression that I should really, you know, like respond to it, do the activated as possible. But then after a couple of drives, then I learned it. Okay, I can ignore this message whenever I want to activate it. So I did that so that there wasn't any problem with the timing because that relies on me. If I'm happy, then I'll do it. If I don't happy, I don't want to do it.*" (ID10)

Similar analysis was done for the highway part only. Figure 4.29 and Table 4.15 present the results which show the same pattern as for the full drive, i.e., driver ID10 differs from the other drivers.

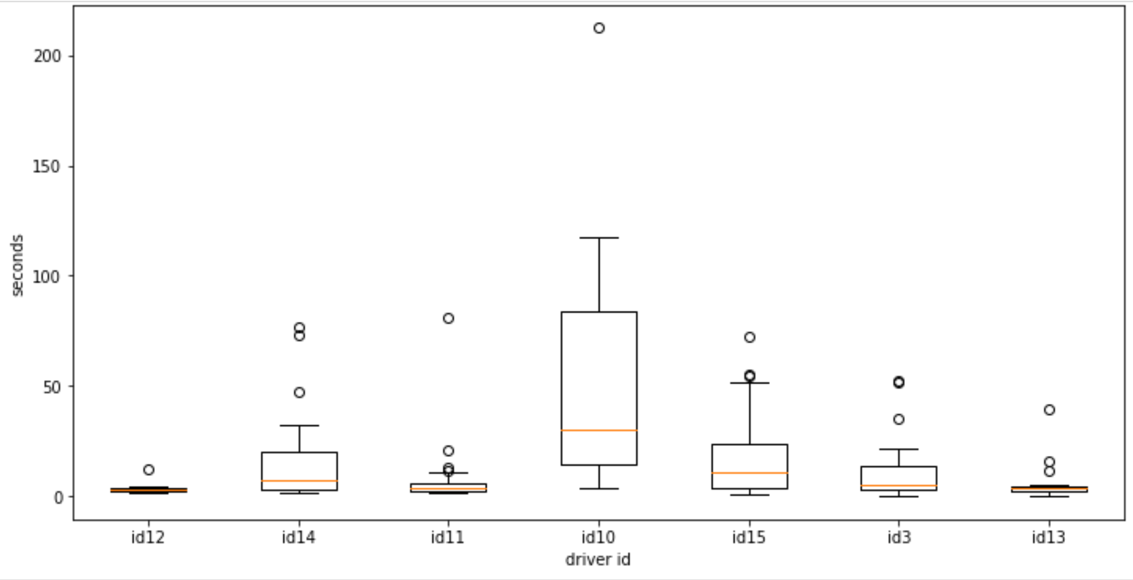


Figure 4.29 Boxplot of the time it took drivers to switch from manual to automation – highway only

Table 4.15 ANOVA analysis of the time to switch from manual to automated driving – highway only

Source	Driver
ddof1	6
ddof2	186
F	11.543339
p-unc	5.713705e-11
np2	0.271331

4.4.3.3. Did some drivers switch on automation much more than others (i.e., number of times)?

Figure 4.30 presents the results of the number of times drivers switched automation on during the full drive. Two observations can be made, the first is with respect to the general downward trend in number of switches as the number of drives increase; the second is the different patterns that drivers ID14 and ID15 have from the others – these two drivers had a lower number of times switching automation on.

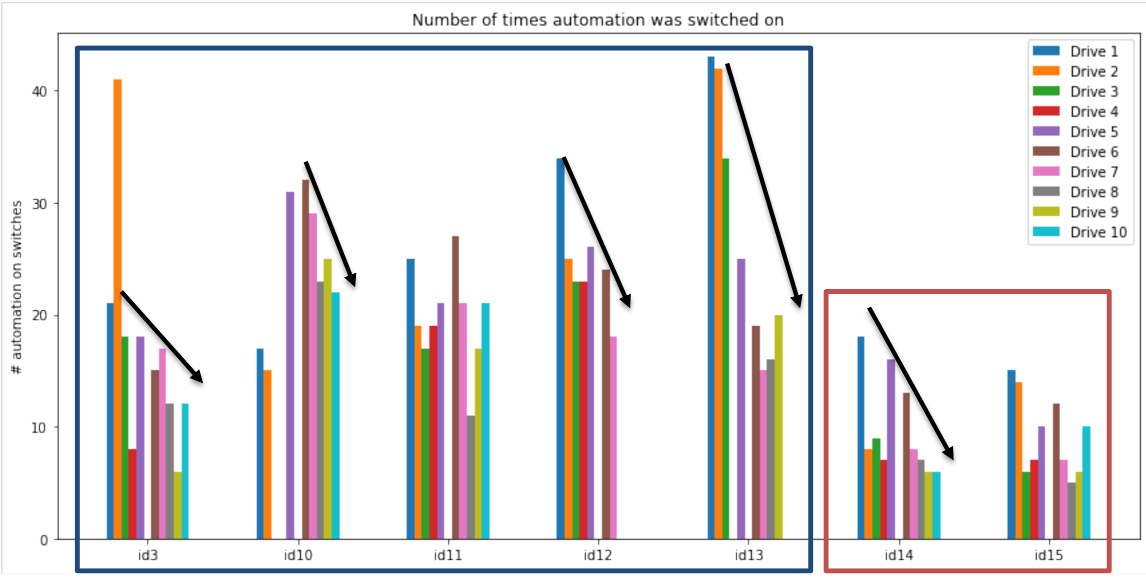


Figure 4.30 The number of times drivers switched automation on during the full drive.

This raises the question of whether drivers keep the system on more (i.e., longer) and thus switch less or they use the system less with time. To investigate this, the median automation duration per drive and driver was analysed and the results are presented in Figure 4.31.

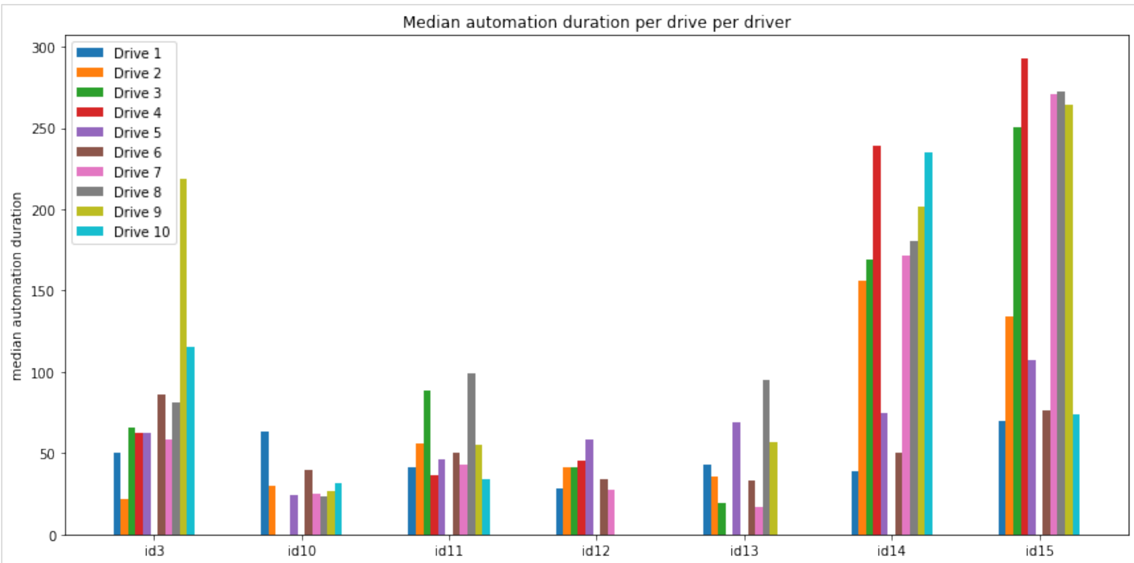


Figure 4.31 Median automation duration per driver per drive.

There is no upward trend visible in terms of the automation duration (median), so, it seems to be that drivers use the automated system less with time. Nevertheless, the figure reveals another interesting trend. Driver ID14 and ID15 who switched on automated driving less often than the other drivers (see figure x) drove longer durations in an assisted driving mode in most of the drives compared to the other drivers.

During the interviews, both drivers reported that they drove in the assisted driving mode most of the time and that they tried to push Mediator's limits (e.g., driving longer in assisted driving mode than recommended).

"*I like it to be active. It's a nice feature to have when driving on more highway sort of roads. […] I try to push it a bit. And, but yeah, once almost every time where the ODD said they would end, it was pretty difficult to continue driving with the system on. However, there were sometimes that I activated the system when we were outside of like, what was the original plan.*" (ID14)

"*I think it was always I. I drove most of the time assisted.*" (ID15)

4.4.3.4. Is there a difference in gaze behavior before and after take-over request?

Figure 4.32 presents a boxplot and median of the gaze duration 30s before vs 30s after the Mediator message is sent to the driver. As can be seen from these figures, no apparent differences in the gaze behaviour can be noticed.

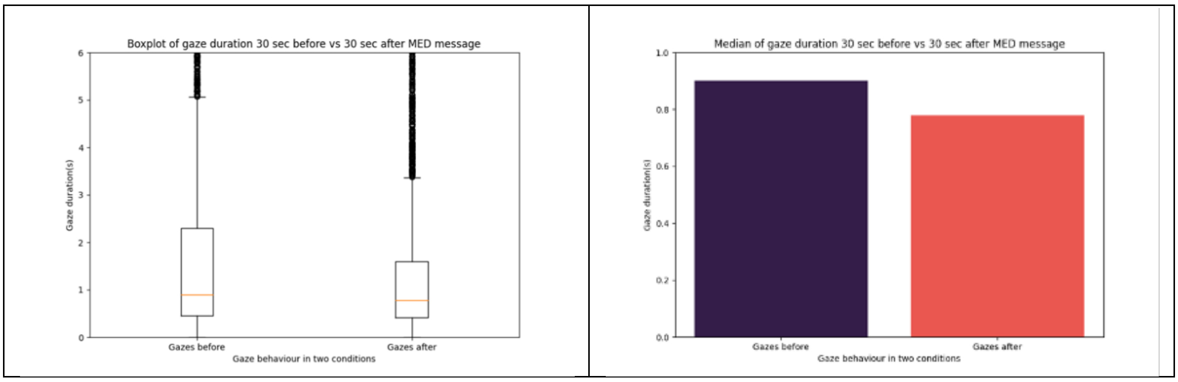
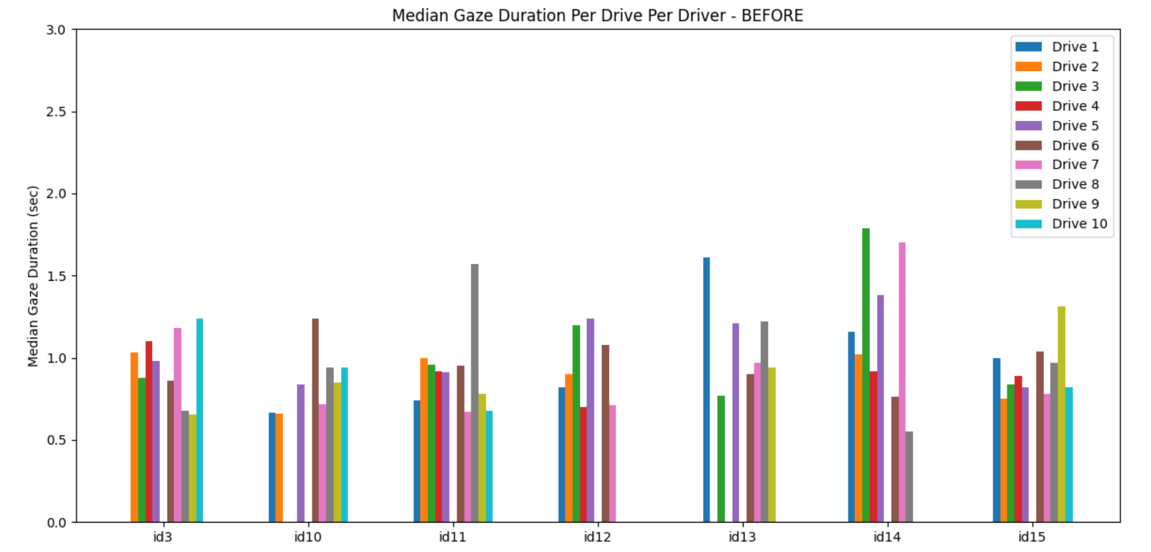
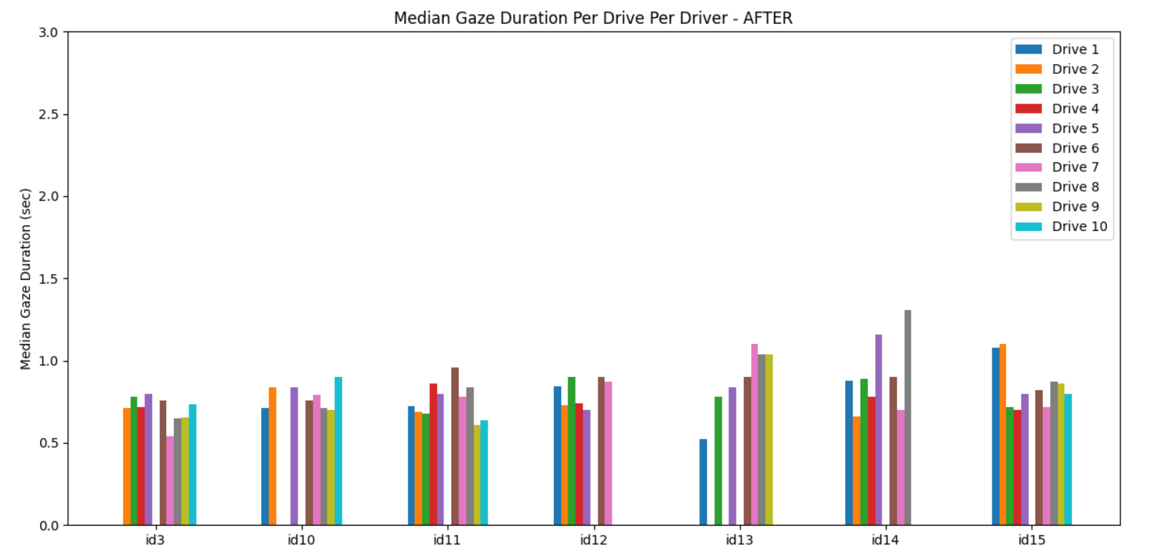


Figure 4.32 Boxplot (left) with median value (yellow line), and median (right) of gaze duration 30s before vs 30s after the Mediator message is sent to the driver.

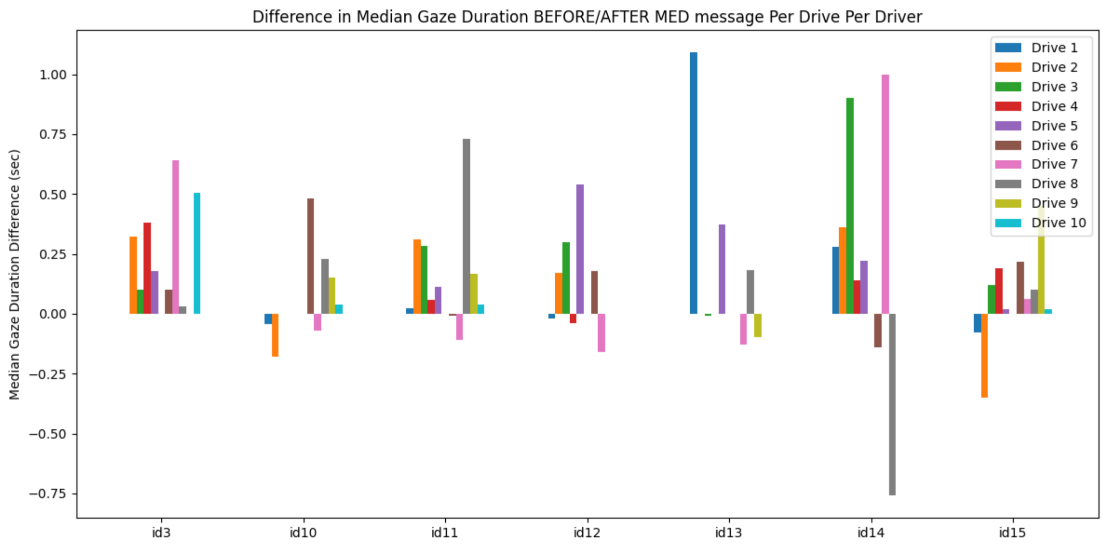
To further investigate this, the median gaze duration per driver and drive before and after take-over requests on the highway section by the Mediator system and the difference between them were plotted as shown in Figure 4.33. No clear trend can be identified when looking at these figures.



(a) before



(b) after



(c) difference

Figure 4.33 Median gaze duration per driver and drive before (a) and after (b) take-over requests by the Mediator system, and the difference between them (c).

Table 4.16 Repeated measures ANOVA of the median gaze duration before and after take-over requests by the Mediator system

Source	Driver
ddof1	6
ddof2	186
F	11.543339
p-unc	5.713705e-11
np2	0.271331

A closer look into the single item measurements revealed that drivers rated Mediator's warnings when a takeover was necessary as appropriate and the warning times on average as good Figure 4.34 shows that for the first drives, drivers somewhat agreed to the statement "*When Mediator asked me to retake control, I was warned in an appropriate way*." After drive 5 drivers disagreed to the statement because no messages (i.e., recommendations or warnings) were sent by Mediator. After drive 6 the answers showed the highest variance indicating that some drivers preferred Mediator's recommendation to use assisted driving also for shorter sections meaning that the warnings to retake control appeared more often (and sometimes shortly after switching to assisted driving). Other drivers rated this behaviour of Mediator as less appropriate. After drive 9 and 10 the appropriateness of the warnings was rated a bit lower on average compared to the other drives. One possible explanation might be that drivers were influenced in their answers by the oversensitive distraction warning.

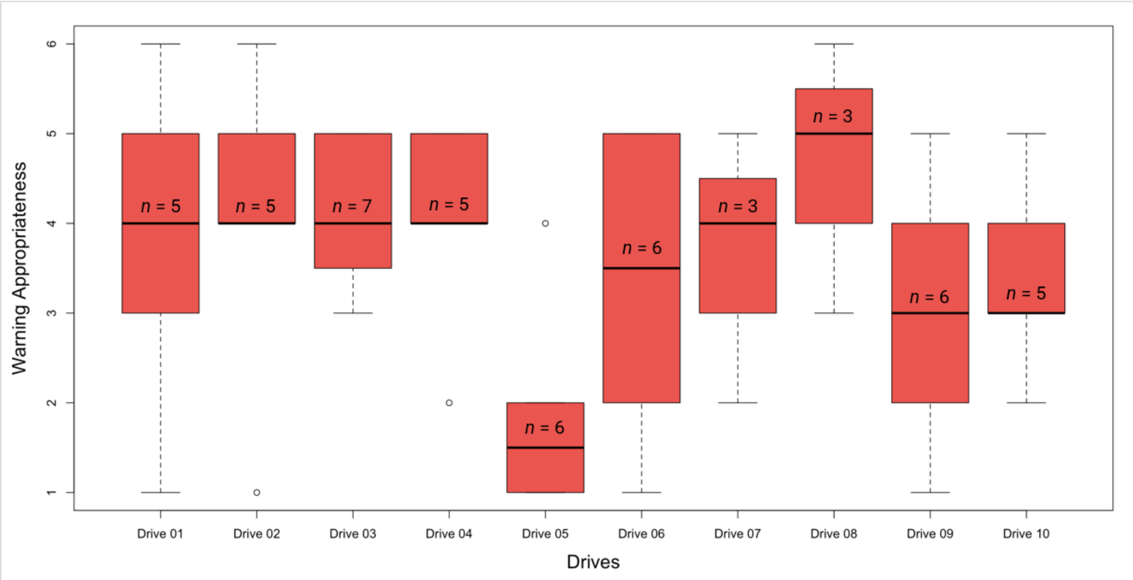


Figure 4.34 Drivers answers to the statement "When Mediator asked me to retake control, I was warned in an appropriate way." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree for all 10 drives. n shows the number of valid values per drive.

Driver ID12 and ID13 are the most critical ones rating the appropriateness of the warnings as quite low (see Figure 5.2 in Appendix 0. Main reason for the rating might be that both drivers are not agreeing with Mediator's recommendations and are not convinced of Mediator's usefulness for level 2 automation at all.

"*And I didn't agree with the recommendation to stop using the system, so then it becomes quite disturbing*." (ID13)

"*Again, I don't think it makes sense with an L2 system*." (ID12)

Figure 4.35 reflects drivers' agreement to the statement "*When Mediator asked me to retake control, I was warned with sufficient time to do so safely*". Results show that drivers were mostly satisfied with the timings of the warnings. During the interviews, driver ID3, ID10 and ID15 stated that they were satisfied with the timing of the warnings.

"*That felt good. That was it was on the right time to still to still react. Not too early*." (ID15)

Hence, the used configuration (first warning to retake control 30 seconds before end of ODD) seems to be appropriate in most situations. Only after drive 5 drivers disagreed with the statement because Mediator didn't send any messages (e.g., warnings or recommendations) at all.

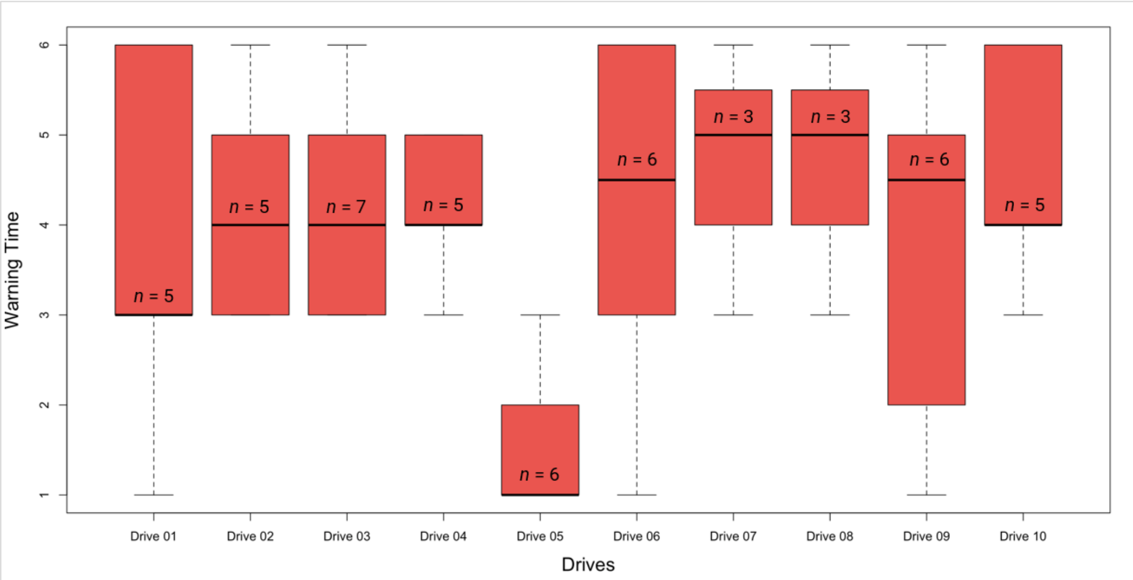


Figure 4.35 Drivers answers to the statement "When Mediator asked me to retake control, I was warned with sufficient time to do so safely." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree for all 10 drives. n shows the number of valid values per drive.

Drivers ID12 and ID13 were less satisfied with the timings which is in line with their quite low rating of the appropriateness of the warnings at all, most likely due to disagreement with Mediator's recommendations or low perceived usefulness for level 2 automation (see Figure 5.3).

"*You know, given my overall impression, I think that the best timing would have been no timing at all*." (ID12)

Further, driver ID 13 stated during the interviews that he/she would prefer later warnings (e.g., 10 seconds before end of ODD). "*When it ends, I thought the countdown started a bit too early, Like 30 seconds. It became quite a lot of beeps then, like the last 30 seconds, I rather like five or 10 seconds at most, I Think since you're driving any control, you should be sufficient*." (ID13)

On the other hand, driver ID10, ID11 and ID14 would have preferred earlier warnings.

"*It might be good to have at least for me, it might be good to have like early then I can prepare myself if I'm engaged in some of the tasks*" (ID10)

"*So when I reach the end of the highway, there's a speed change sign from the 100 to 80. I'm aware I should decelerate and try to adapt to the new speed. And then at that moment I expect Mediator to warn me in advance because I saw the plate and after a couple of seconds, the Mediator starts warning me and that period one, two seconds I feel stressed because I'm thinking there should be one warning or indication, but there's not*." (ID11)

"*I wanted to have like end of the handover done or said that it would be done 15 seconds earlier*." (ID14)

Hence, some drivers seem to prefer earlier warnings, other drivers seem to prefer later warning. Additionally, drivers' preference might vary according to the situation (e.g., fast approach to a highway exit vs. slower approach to an intersection in a city).
"*And in two locations, as I mentioned, I think I should take back control earlier. Yeah, so I both did. But in general, like the exit, the highway, I do it when they ask me to take control, it's naturally*." (ID11)

Moreover, drivers' answers indicate that their preference might change with more experience (e.g., repeated exposure to the Mediator system), for instance driver ID11 who expressed the wish for earlier warnings after the first two drives but was satisfied with the timings after the last drive.
"*I think the timing is okay*". (ID11)

Results indicate that the timing of warnings, especially for retaking control over the driving task, should be adaptable to drivers' preferences (that might change over time) and situations' requirements. Drivers' answers point in the same direction.

"*Like being able to individualize those. Um. And, yeah, maybe like the time. The warning or time before the warning comes or when the warning is*." (ID14)

"*Maybe like the environment in the next 10 minutes. Like I can imagine if, for example, if I use ad drive over 5 hours, I might go to sleep and quite sleepy. So if the Mediator or some other system tells me in front of me next 20 minutes there was a road construction you have to care for*." (ID11).

4.4.4. RQ4: How do drivers evaluate the Mediator system with its functions and the HMI after repeated use?

4.4.4.1. Trust

Figure 4.36 Drivers answers to the statement "*I can trust Mediator*." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree for all 10 drives. n shows the number of valid values per drive Figure 4.36 shows drivers' reported trust in Mediator after each drive on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely disagree. Results show that drivers' trust increased over the first four drives. After drive 5, the trust was quite low which is in line with the expectations as Mediator, for instance, did not give any recommendations or show any time budget information. After drive 6, the reported trust is quite high, although the variance is highest reflecting drivers' different levels of trust in Mediator when recommendations to use assisted driving appeared for shorter sections. For the last four drives, trust was rated very differently for the separate drives but with a lower variance between the drivers. After drive 7 and 8, only three drivers were able to evaluate the Mediator system due to technical reasons. Hence, the rating need to be interpreted carefully as they represent the opinion of only three drivers. In the last two drives, trust was rated quite low. This can partly be explained by the above-mentioned oversensitivity of the distraction warning function as well as experienced problems with the whole Mediator system (e.g., several system shut downs) during the last drives.

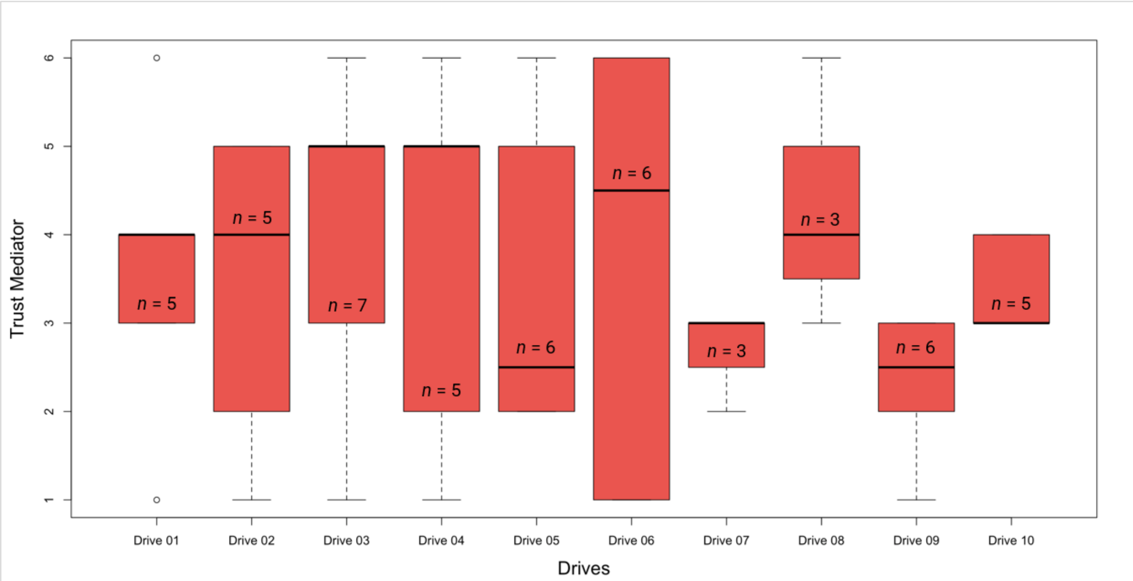


Figure 4.36 Drivers answers to the statement "I can trust Mediator." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree for all 10 drives. n shows the number of valid values per drive.

Driver ID3, ID12 and ID13 rated their trust in Mediator quite low over all drives (see Figure 5.4 in Appendix 0). Driver ID12 justified his/her low rating by stating that, in his/her opinion, Mediator was not useful for L2 automation. "*No, not at all. As I said, I didn't think it was useful at any time*." (ID12)

During the interviews after the second drive, six drivers stated that they trusted Mediator (ID3, ID10, ID11, ID13, ID14, ID15). "*I think, yeah. I mean, when it said it was available, it sure performed well. So in that sense, I guess it was trustworthy*." (ID13)

Driver ID10 reported even high enough trust in the system to get engaged in secondary tasks. "*The trust already started to build in. So I started using my phone sometimes because I know that when the vehicle, you know, when there's going to be a disengage message, it will definitely tell me when it's going to do that*." (ID10)

Further, results show that drivers' trust ratings a quite similar to their reliability ratings (see Figure 4.36 and Figure 5.4) indicating that trust in Mediator is closely related to perceived reliability of Mediator. Drivers' interview answers support this assumption. "*Yes. Since it was reliable. Yes, I trusted. I sometimes chose to not follow it, but it was predictable and yeah, I could trust it*." (ID15)

"*And the other thing is that I think it's quite strange that the system allows you to drive the system […] But then it shouldn't allow the assistance instead to reactivate it. Okay. So I'm not sure if you actually need the "Don't drive mode". It's either, I could decide that myself or use the recommendation when it is suitable*" (ID13)

"*If they can make like, a marker, virtual marker on the road, I mean, trust them more than I know: Okay, Mediator knows more than me. Yeah. Okay. I would trust it more*" (ID11)

"*that the LED bar starts to shrink one by one, I think. But then this happens on the driver unit, especially mainly on the dashboard, but it doesn't reflect on the ambient light or the LED bar at the bottom. So I was so I was just trying to, you know, look at it when there is a shrinking happening* *there, but there is no shrinking happening there. Maybe that's something like from the design requirement. So that was that's something which I also like notice. So basically if he can a little bit work on the luminosity and also the you know where how the lights are because we're using lights and then the text as a mode of communication. So probably making it much more details would help I think.*" (ID10)

"*The LED bars that we have now, maybe that the luminosity or maybe, you know, like the light intensity and these things could be like improved*." (ID10)

"*Well, I suppose the system would have learned when I don't have my hands on the wheel. If then Mediator had told me. But I need to have my hands on the wheel. It would be like a better experience I suppose*." (ID3)

"*I wanted it to be activated more at some occasions"* (ID13)

"*that I didn't find the system available what I wanted*." (ID13)

After drive six,
"*Maybe it's just I got I got familiar with the system, or I started to like, trust the system more*." (ID10)

"*When I took that drive today, it just turned up, like, ten times. And I don't know why, And I think that. The car was Tourney with no reason. So it was… today I didn't trust it at all*." (ID3)

After the last drives,
"*So every time when the things when I switch between assisted and manual mode or manual assisted mode, it's also like I probably keep looking at the green Icon most of the time. Okay. Compared to the, you know, the LED bar. […] Indicating maybe then my attention might be more on that. Yeah. Let's say. This is something I just realised it like every time when I activate it to ensure that if it is activated. But I basically. Yeah. Was more looking at the cluster icon. […] I would say, like trusting the instrument cluster information.*" (ID10)

"*That I think in other situations it wasn't that difficult. It was quite straightforward. So, you know, when you're activate it, when you're assisted and someone. So yeah, now another situations, yeah, probably. I would trust the system, I would say*." (ID10)

"*In general, Yes. Yeah. So it was as I said. It was consistent when the, the takeover requests would come. And it was also I mean there was always something on the, on the road that I could relate to why it was, would have been nice to, to know in at once why actually the request would come. But when it comes then it's yeah, you could always relate to something like a traffic light or end of highway or something that was going on*." (ID15)

"*Yeah, I think I do. I do trust that*." (ID11)

"*The like the driving like the driving recommendations on manual or automated that I still think is reliable and trustworthy and so on*." (ID14)

"*Also, the time left, I think, is some trusting, like when you're when it says like 10 minutes, then you know that it will be available at that time*." (ID13)

“*But then I didn't particularly trust everything then, like since I wanted to use it more. But when it's said it's that I should not use it, I use it anyway. So in that sense I didn't trust it. But when you're in highway and have it come through a lot then yeah.”* (ID13)

“*I think it's the same. So I don't do that. No, I did before, but not right now*.” (ID3)

4.4.4.2. Usage intention

Figure 4.37 shows drivers' usage intention after each drive on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely disagree. Results show that after the first three drives, drivers would tend to use Mediator in future cars. After drive 5, usage intentions decreased a bit, most likely due to missing support from Mediator in this drive. After drive six, variance in answers is highest reflecting drivers' different preferences for Mediator offering automated driving also for shorter sections. During the last drives, use intentions are lowest, probably because of the oversensitive distraction warning.

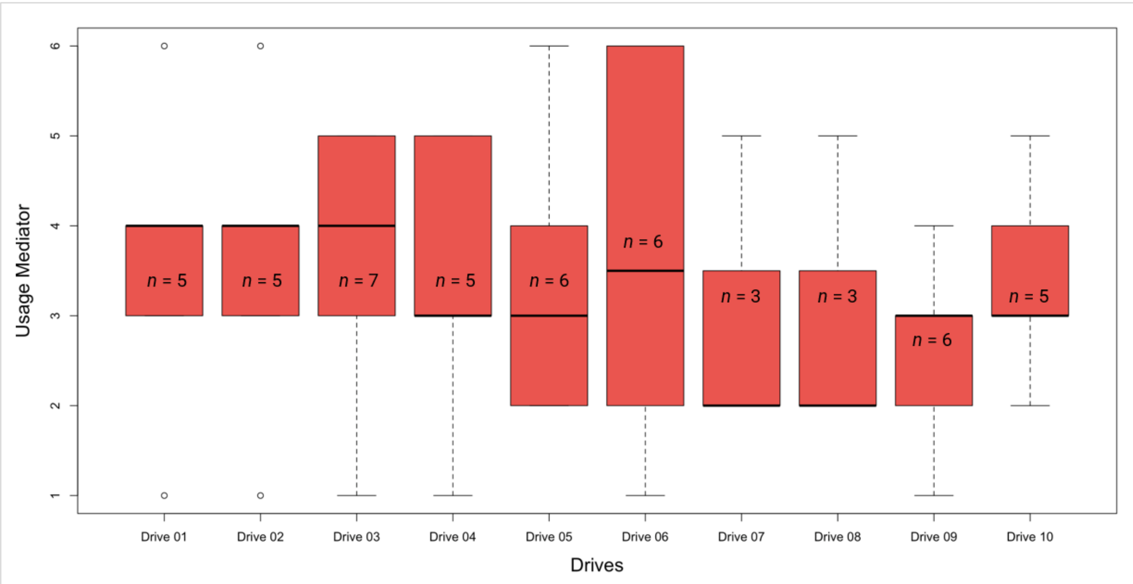


Figure 4.37 Drivers answers to the statement “I would use Mediator if it was in my car.” on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree for all 10 drives. n shows the number of valid values per drive.

Driver ID3, ID12 and ID13 and ID14 have only low to medium intentions to use Mediator in their future cars (see Figure 5.5 in Appendix 0). Driver ID10 and ID11 show higher intentions to use Mediator in the future.

During the interviews, three drivers (ID3, ID13, ID14) agreed to use Mediator in their future cars but with several prerequisites like the chance for personalization (e.g., shut off distraction warning or beeping sounds) and improvement of the system (e.g., distraction warning). “*Yeah. If I had it in my car, I would use it. But the. I suppose, but. But without annoying beeping*.” (ID3)

Driver ID12 stated that he/she would not use Mediator for L2 driving but for higher automation levels. “*No, not as it is here. But again, then, combined with unsupervised driving, maybe. I think it can be useful*.” (ID12)

4.4.4.3. Comfort

Figure 4.38 shows drivers' experienced comfort after each drive on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely disagree. Results show that drivers' experienced medium comfort after the first drives with no change after drive 5 (no recommendations and time budget offered by Mediator). After drive six, driving comfort was rated highest but also with the highest variance between drivers. The high variance might reflect drivers' differing preference for Mediator offering automated driving also for short sections and, hence, more often. During the last four drives, experienced comfort was reduced, probably due to distraction warning and repeated problems with the system.

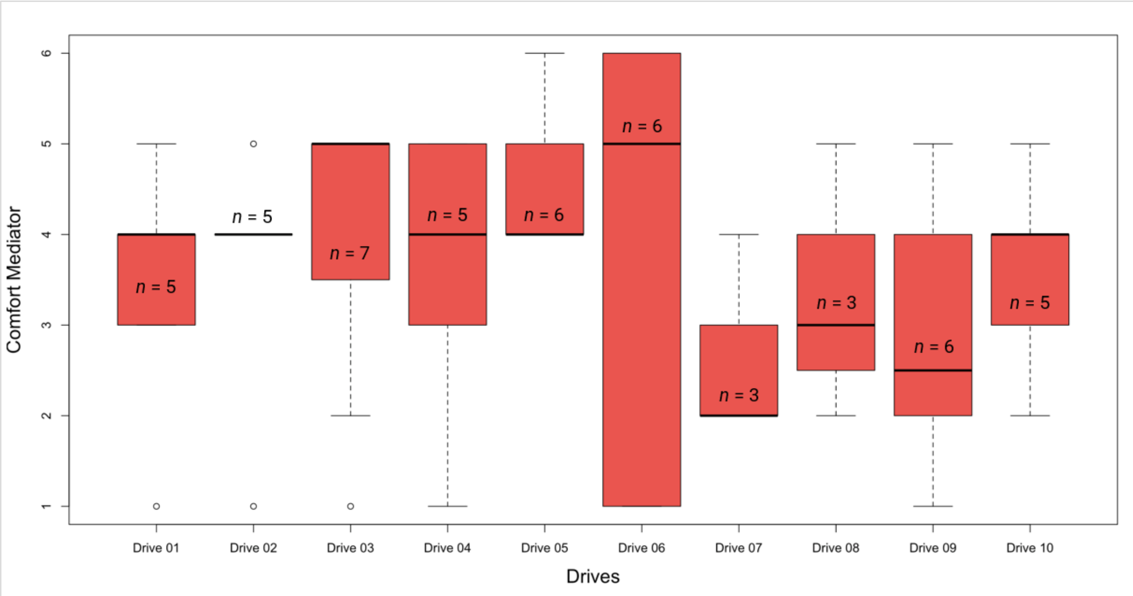


Figure 4.38 Drivers answers to the statement "Driving with Mediator was comfortable." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree for all 10 drives. n shows the number of valid values per drive.

Driver ID12 and ID13 felt least comfortable with Mediator (see Figure 5.6 in Appendix A.4). Driver ID10 and ID11 expressed highest comfort values.

4.4.4.4. Confidence in using Mediator

Figure 4.39 shows drivers' confidence in using Mediator after each drive on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely disagree. Results show that drivers' confidence in using Mediator increased after the first drives, and decreased a bit after drive five (no recommendations or time budget) and drive six (Mediator offers automated driving also for short sections). Again, the variance between drivers were highest after drive six, reflecting different preferences of the drivers. After the last four drives, drivers rating of their confidence varied between the drives with low ratings after drive nine and medium ratings after the last drive (same level as after the first two drives). Results indicate that the system needs some improvement (e.g., regarding the sensitivity of the distraction warning) to ensure high confidence in system usage.

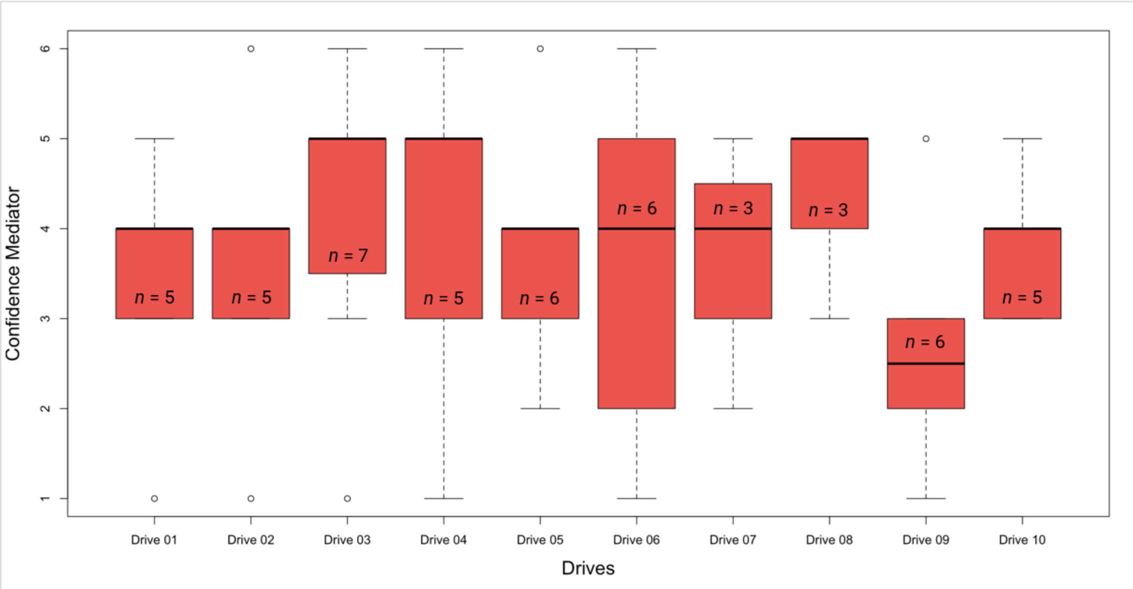


Figure 4.39 Drivers answers to the statement "I felt very confident using the system." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree for all 10 drives. n shows the number of valid values per drive.

Driver ID3, ID12 and ID13 felt least confident in using Mediator (see Figure 5.7 in Appendix A.4). Driver ID10, ID11 and ID14 expressed highest confidence values.

4.4.4.5. Perceived Safety

Figure 4.40 shows drivers' perceived safety after each drive on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely disagree. Results show that drivers' perceived safety was on a medium level after the first three drives with a slight increase after drive four. Although, variance between drivers was also very high after drive four indicating that drivers experienced drive four quite differently. After drive five (no recommendations and time budget provided), perceived safety was rated lowest. After drive six (automated driving was offered also for short sections), perceived safety was again rated on a medium level but with a high variance between drives, most probably due to different preferences for the experienced configuration. After the last drives, perceived safety was again on a medium level.

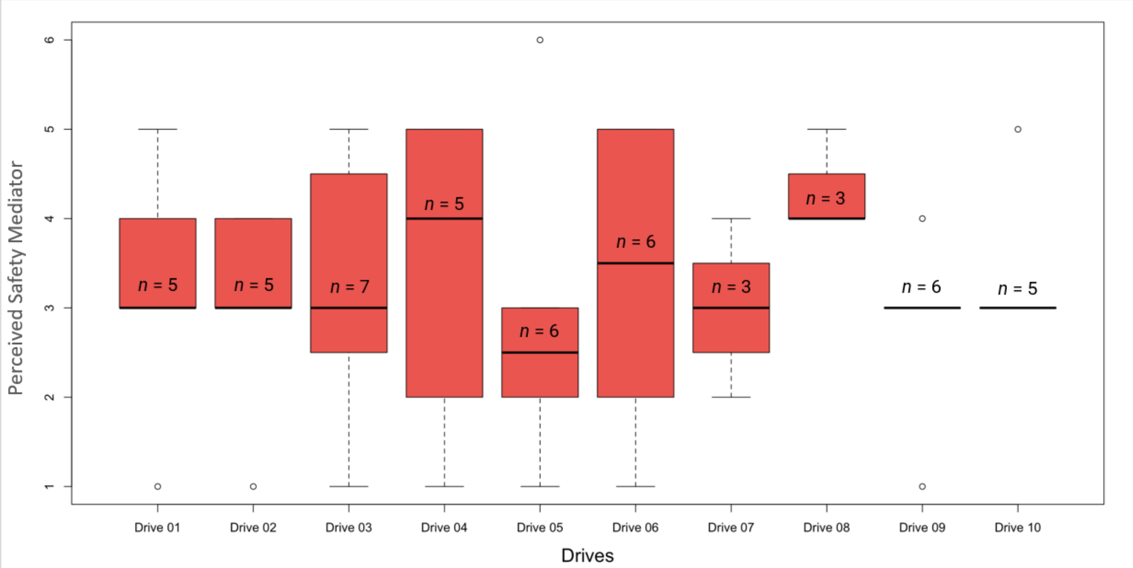


Figure 4.40 Drivers answers to the statement "Mediator provides security." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree for all 10 drives. n shows the number of valid values per drive.

Driver ID12 and ID13 felt least safe when using Mediator (see Figure 5.8 in Appendix A.4). Driver ID10 and ID11 expressed highest safety values.

4.4.4.6. Awareness

Figure 4.41 shows drivers' agreement to the statement "*Mediator improved my awareness of the driving automation*." after each drive on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely disagree. Results show drivers tended to agree to the statement, that Mediator improved awareness of the automation system. After drive five, the rating was a bit lower with a high variance in drivers' answers. This might be due to the fact that Mediator was not giving any recommendations or provided time budget information during the drive, which might have caused a reduced awareness of the driving automation for some drivers. On the other hand, Mediator was still indicating the current driving mode by different colours of the LED stripes, ambient lightning and the displayed route in the centre stack. Hence, for some drivers, the reduced functionality of Mediator still helped to improve awareness. Again, variance was highest after drive six, indicating that for some drivers the suggestion to activate automated driving even for short sections helped to increase awareness of the driving automation, and for others it was less helpful.

Again, driver ID12 and ID13 reported lowest improvement of awareness of the driving automation by Mediator (see Figure 5.9 in Appendix A.4).

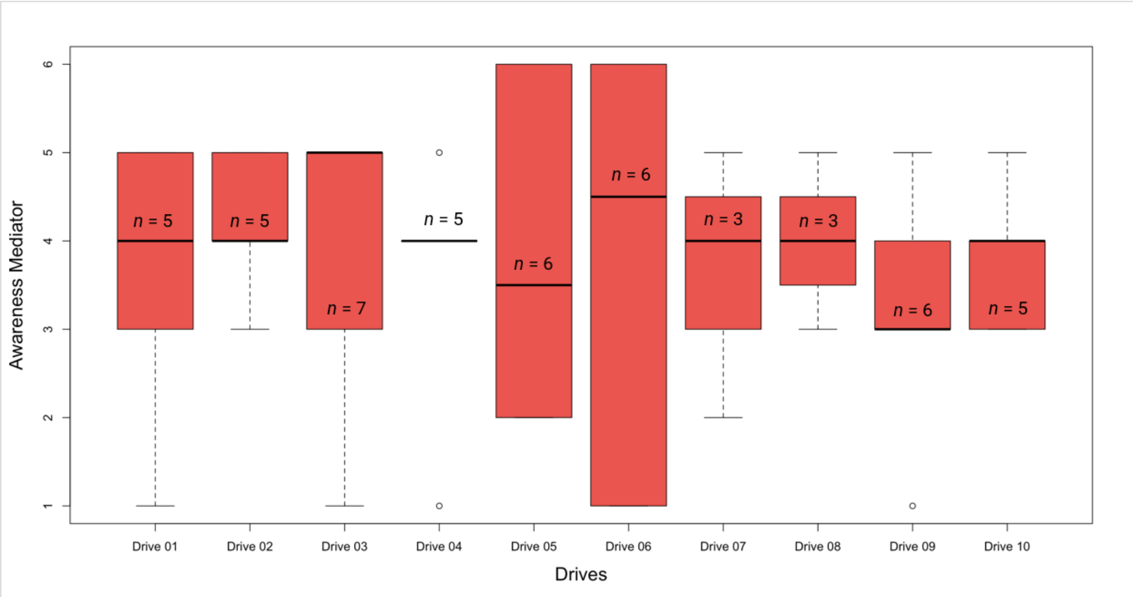


Figure 4.41 Drivers answers to the statement "Mediator improved my awareness of the driving automation." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree for all 10 drives. n shows the number of valid values per drive.

4.4.4.7. Benefit

Figure 4.42 shows drivers' agreement to the statement "*I consider Mediator as a beneficial system for (partly) automated driving*." after each drive on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely disagree. Results show that drivers tend to agree to the statement. Again, variance was highest after drive 6 reflecting drivers' different preferences for Mediator offering automated driving even for short sections.

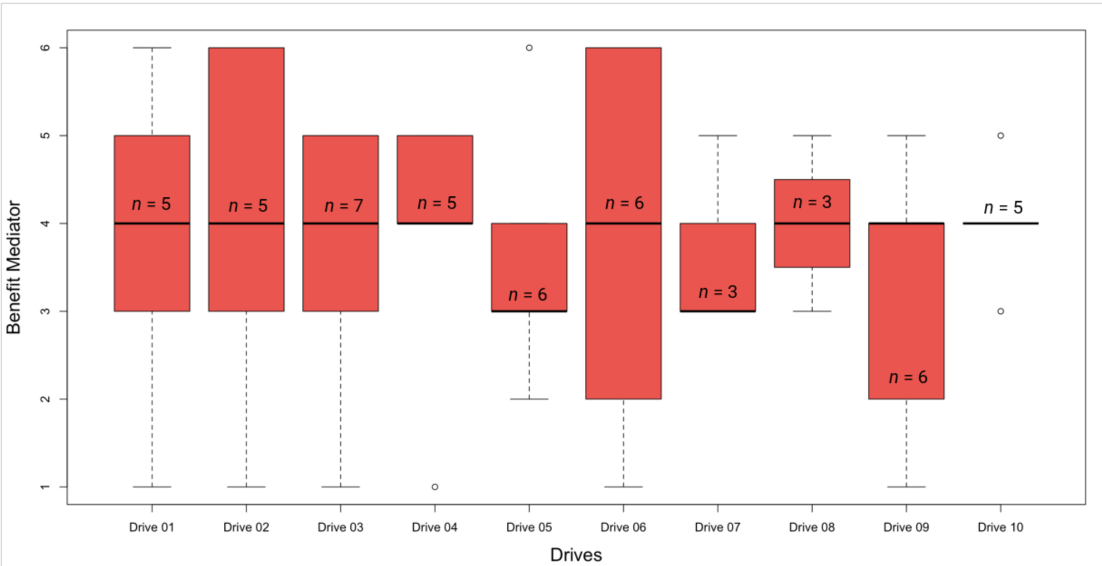


Figure 4.42 Drivers answers to the statement "I consider Mediator as a beneficial system for (partly) automated driving." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree for all 10 drives. n shows the number of valid values per drive.

Driver ID3, ID12 and ID13 rated Mediator's benefits for (partly) automated driving the lowest (see Figure 5.10 in Appendix A.4). As reported by ID12, this is partly because no added value for L2 driving was experienced.

4.5. Discussion

The main objective of the study was to evaluate the technical integration of the full Mediator system in which all components in actual working form came together and interacted in real-time.

4.5.1.1. Functionality

The main objective of the in-vehicle prototype was to provide accurate "time budget" estimation to predict when the drives need to switch from automated driving to manual and vice versa. The results show that prototype had a good matching when it comes to calculating the automation fitness and unfitness. When it comes to automation usage, take-over times are very similar between the two conditions (with and without Mediator). However, there is higher variability in the drive without Mediator.

During interviews with the drivers, the time budget is evaluated as a good and helpful functionality that can have added value to higher automated driving modes. When it comes to trusting the functionality of the prototype, overall interview results show that, with repeated exposure, drivers trust in Mediator increased compared to the baseline system (Drive 5). However, due to the small sample size of drivers, there is a larger variance between the drivers' responses. The driver state estimation (distraction alerts) was reported to be oversensitive.

4.5.1.2. Validity

The data collection was carried out using professional test drivers. In total, 70 hours of driving data and interview data were collected during the study. Due to repeated exposure to the Mediator system, there was some learning effect observed when analysing the gaze duration between different drives over the course of the study. The decreasing trend in the number of gazes toward the Mediator HMI indicated that drivers become more experienced with the system. The study carried out has its limitations due to the small sample size of the data collected. Thus, TI in-vehicle prototype is an exploratory study with limited validity when it comes to representing the results on a population level. However, the integrated system performed as designed and valuable feedback were obtained from the professional drivers.

4.5.1.3. Reliability

The reliability of the system was carried out by comparing driver usage of the prototype system across different configuration of Mediator system based on the experimental design. The main analysis was carried out on the interview data and questionnaire data recorded. Drivers' responses were overall good when it comes to reliability of the Mediator system. From the interview data it was observed that audio HMI was more reliable compared to visual HMI displayed on the Mediator display. However, due to experimental design reliability was rated low for baseline trial (drive 5) compared to others. Nevertheless, the results show that the overall reliability rating is not affected too strongly indicating that drivers can distinguish between the reliability of the distraction warning and reliability of the whole Mediator system. Based on the interviews, drivers recommend improvements in terms of reliability of the Mediator system e.g., adding navigational information, preventing automated driving when it is not recommended, more information why a take-over is requested and reduced transparency if the system can handle specific unexpected situations.

5. Conclusion

The deliverable summarizes the on-road evaluations of the vehicle-integrated Mediator system. Three real-life on-road studies were conducted to test the overall performance of the Mediator system and its effects on safety-relevant behaviours, driver reactions and driver opinions.

First Study in Italy

The Italy study aimed at the evaluation in real-life conditions the usability, the acceptance, and the perceived trust of the designed MEDIATOR HMI solutions. To reach this purpose, the HF in-vehicle prototype was used. This demo vehicle allowed 16 naïve participants to have an "automated" vehicle experience, without being in a real automated vehicle, because the prototype was always driven, using the standard primary controls of the vehicle, by the driving wizard seated on the right seat. Different MEDIATOR HMI solutions (visual, acoustic, vocal, luminous, …) located in different parts of the vehicle (e.g., participant frontal display, shifter, steering wheel, seat and belt and centre dashboard display) were evaluated during the on-road trials in their different states, which changed according to the tested use cases. The participants experienced, on a mixed scenario of 46 km, seven use cases in which handover and takeover were requested by the user or initiated by the system for some different reasons such as drowsiness, distraction, desire to take back the control from the user...; in addition to handover and takeover also an improvement of driver fitness during an automated driving mode was tested.

Participants trusted the MEDIATOR HMIs, and the usability was scored as "good". The acceptance of the "automated" vehicle experience was positive as well as that of the MEDIATOR HMI.

- Participants at the end of the trial said they appreciated the experience they had with the "automated vehicle". This, plus the behaviour they had during the trials (e.g., not considering at all the driving wizard) indicates that, even if the WoOz vehicle prototype was not functional, they "forgot" there was the driving wizard and they interacted with the examiner as they were in an automated vehicle.
- This user testing allowed to identify advantages and disadvantages of the MEDIATOR HMI solutions in an ecological context.
- For example, participants appreciated that:
- in dangerous situation like distraction or drowsiness, there were more alerting warnings (e.g., acoustic feedback and seat belt haptic feedback) respect to safer uses cases
- The vehicle mode was shown
- The Head Unit HMI was very visible and some of its icons were well comprehensible like that in the use case in which drowsiness was considered.
- Some disadvantages were on:
- handover and takeover latency time because it was considered too long
- handover escalation in dangerous situation (e.g., drowsiness and distraction) was too long
- the usage of LEDs on steering wheel, at least in the position they tested them, because they were not well visible during the daylight and probably annoying during night
- visibility and legibility of HMI on the Cluster
- some icons (e.g., steering wheel with only one hand and strikethrough steering wheel) not very understandable.

- Considering these results, it was possible to advise some redesign guidelines to enhance even more the MEDIATOR HMI solutions usability for the next trials. Then, this Italian user testing study allowed to select the most appropriate HMIs and refine them, following the User-Centred Design process, so to be tested during the following MEDIATOR project on-road study in Sweden.

Second Study in Sweden

The Swedish study with the HF in-vehicle prototype focussed on the evaluation of acceptance and safety of the Mediator system. In general, the Mediator system was preferred over the Baseline system, but this preference was strongly influenced by age. It is advised to increase the cohesion in the HMI and improve clarity of icons and warnings through, for example, making more use of learned affordances as well as testing the system on specific participant samples (such as elderly vs young driver). Several novel features in the Mediator system showed great comfort and safety potential. The time budget information was both appreciated and used by participants to be continuously aware of when a change in responsibility would take place. The combination of warnings and continuous mode and time budget information decreased the total duration in which participants were distracted and the maximum uninterrupted period of distraction. Especially the latter can have a significant effect on road safety, as long continuous periods of being distracted severely reduces situation awareness. Also, the elaborate Mediator takeover ritual, which included reasons for why a takeover was happening, was more appreciated and understood than the takeover ritual of the Baseline system. Finally, the Mediator system includes an active proposal to increase the level of automation which was much appreciated by participants. They also strongly agreed that they would increase automation use if such a feature was available in their car. If automation is indeed safer than manual driving, this feature could therefore potentially improve road safety.

Overall, the Mediator system showed great potential for improving both driver comfort and safety in future vehicles and provides future designers inspiration for new ways of interacting with the driver in the upcoming generation of automated vehicles.

Third Study in Sweden

The Swedish study with the TI in-vehicle prototype aimed to evaluate the reliability and validity of automation state detection and decision logic using a test vehicle with a simplified MEDIATOR HMI solution. The study carried out was an exploratory study to understand the technical integration concept. The study involved data collection efforts with professional drivers to participate and experience different configurations of the MEDIATOR HMI, automation state, decision logic, and driver state detection. Participants drove ten times in a specific route designed to fulfil the use cases and scenarios described within the MEDIATOR project. Both Quantitative data (from the test vehicle) and qualitative data (from structured interviews) were collected and analysed to answer the research questions selected. Analysis of quantitative data was done by categorising the drives into highway only and full drive.

Results

The results show that, implemented Mediator in-vehicle prototypes have good overall reliability when comparing the time budget predicted and observed. From the interviews and analysis, we also see that the distraction warning implemented in the vehicle was too sensitive. It did not have a significant effect on test driver usage of automation functionality and overall safety of the drives. The simplified HMI implemented in the prototype did not affect their gaze behaviour, in fact they looked at the Mediator display less overtime. The interview analysis showed that Mediator system improved metrics such as trust, comfort, perceived safety, usage and intention of automated driving.

The results of the three on-road studies show that novel features in the Mediator system, such as active proposals to increase the level of automation and the combination of distraction warnings and continuous mode awareness support, can improve comfort and safety. Furthermore, the technical implementation and validation of features such as time budget predictions of upcoming automation availability and real-time decision making proved successful. The study results also provided insight on possible improvements for future development of the system, such as improved fatigue mitigation strategies. In general, the Mediator system showed great potential to improve both safety and comfort in future automated vehicles.

References

Ahlström, C., Kircher, K., Kircher, A. (2013). "A Gaze-Based Driver Distraction Warning System and Its Effect on Visual Behavior". In IEEE Transactions on Intelligent Transportation Systems, 14 (2), 965-973, doi: 10.1109/TITS.2013.2247759

Åkerstedt, T., Gillberg, M., Subjective and objective sleepiness in the active individual.,Int. J. Neurosci., 1990, 52:29–37.

Åkerstedt, T., Anund, A., Axelsson, J., & Kecklund, G. (2014). Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *Journal of sleep research*, 23(3), 242-254.

Athmer, C., Spaan, M.T.J., Li, Y., Liu, Y., Bakker, B. (2022). Simulation of Decision Making in the Mediator System, Deliverable D3.2 of the H2020 project MEDIATOR

Bengler K., Omozik, K., Müller, A. I. (2020) The Renaissance of Wizard of Oz (WoOz) – Using the WoOz methodology to prototype automated vehicles. In Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2019 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

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

Brooke, J.B. (1996). SUS: A 'Quick and Dirty' Usability Scale.

Callaghan, E., Holland, C., & Kessler, K. (2017). Age-related changes in the ability to switch between temporal and spatial attention. Frontiers in Aging Neuroscience, 9(FEB). https://doi.org/10.3389/fnagi.2017.00028

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

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

Ferris, T. K., Suh, Y., & Miles, J. D. (2016). Investigating the Roles of Touchscreen and Physical Control Interface Characteristics on Driver Distraction and Multitasking Performance. www.ATLAS-Center.org

Fiorentino A., Bakker B., Brase D., Karlsson J., Knauss A., Santhià T., Thalya P., Toffetti A., van Grondelle E., Wang, D., (2022). Final in-vehicle prototypes, Deliverable D2.10 of the H2020 project MEDIATOR.

FOT-Net and CARTRE. (2018). FESTA Handbook Version 7. Retrieved from https://www.connectedautomateddriving.eu/wp-content/uploads/2019/01/FESTA-Handbook-Version-7.pdf

Franke, T., Attig, C., & Wessel, D. (2019). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. International Journal of Human–Computer Interaction, 35(6), 456-467, DOI: 10.1080/10447318.2018.1456150. https://ati-scale.org/

Grondelle, E.D. van, Zeumeren, I. van, Bjorneseth F., Borowsky, A., Chandran, T., Cleij, D., … Christoph, M. (2021). HMI Functional Requirements, Deliverable D1.5 of the H2020 project MEDIATOR

Horrey, W. J. & Wickens, C. D. (2018). In-vehicle glance duration: Distributions, tails, and model of crash risk. *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2018, no. 1, pp. 22–28, Jan. 2007.

ISO 9241-210:2010 Ergonomics of human-system interaction - Part 210: Human-centered design for interactive systems. International Organization for Standardization, Geneva.

Jansen, R. J., Tinga, A. M., de Zwart, R., & van der Kint, S. T. (2022). Devil in the details: Systematic review of TOR signals in automated driving with a generic classification framework. Transportation Research Part F: Traffic Psychology and Behaviour, 91, 274-328. doi:https://doi.org/10.1016/j.trf.2022.10.009

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. International Journal of Cognitive Ergonomics, 4(1), 53–71. https://doi.org/10.1207/s15327566ijce0401_04

Johns, M. W. (1991). A new method for measuring daytime sleepiness: The Epworth sleepiness scale. SLEEP, 14(6), 540-545.

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. ACM Transactions on Office Information Systems, March 1984, 2:1, pp. 26–41.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D. & Ramsey, D. J. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. In Proceedings of National Highway Traffic Safety Administration, Department of Transportation (DOT), Washington, DC, USA, vol. 594, 2006, pp. 1–5.

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

Kyriakidis, M., de Winter, J. C., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., ... & Happee, R. (2019). A human factors perspective on automated driving. Theoretical issues in ergonomics science, 20(3), 223-249.

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

Mano, D., Berger, N., Larsson, A., Brännström, M., Knauss, A., Toffetti, A., … Christoph, M. (2021), Quantified Markers for Degraded Automation Performance, Deliverable D1.3 of the H2020 project

May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. Transportation research part F: traffic psychology and behaviour, 12(3), 218-224.

Metz, B., Rösener, C., Louw, T., Aittoniemi, E., Bjorvatn, A., Wörle, J., … (2019). Evaluation methods. Deliverable 3.3 of the H2020 project L3Pilot. Retrieved from https://l3pilot.eu/download/

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

Müller, A. I., Weinbeer, V., & Bengler, K. (2019, September). Using the wizard of Oz paradigm to prototype automated vehicles: methodological challenges. In Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications: Adjunct proceedings (pp. 181-186).

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

Östlund, J., Nilsson, L., Carsten, O., Merat, N., Jamson, H., Jamson, S., . . . Anttila, V. (2004). HASTE deliverable 2: HMI and safety-related driver performance (GRD1/2000/25361 S12.319626).

Pereira Cocron, M., Vallejo, A., Delgado, M., Wilbrink, M., Anund, A., Krupenia, S., … (2019). Evaluation Framework. Deliverable 7.3 of the H2020 Project ADAS&ME. Retrieved from https://www.adasandme.com/dissemination/public-deliverables/

Post, J.M.M., Ünal, A.B., & Veldstra, J.L. (2020). Deliverable 1.2. Model and guidelines depicting key psychological factors that explain and promote public acceptability of CAV among different user groups. SUaaVE. http://www.suaave.eu/wp-content/uploads/sites/17/2021/02/SUaaVE_WP1_D1.2_20200930_V100.pdf

Rodarius, C., Duflis, J., Fahrenkrog, F., Rösener, C., Várhelyi, A., Fernandez, R., … (2015). Test and Evaluation Plan. Deliverable 7.1 from the FP7 Project AdaptIVe. Retrieved from https://www.adaptive-ip.eu/index.php/deliverables_papers.html

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

Said, S., Gozdzik, M., Roche, T. R., Braun, J., Rössler, J., Kaserer, A., Spahn, D. R., Nöthiger, C. B., & Tscholl, D. W. (2020). Validation of the Raw National Aeronautics and Space Administration Task Load Index (NASA-TLX) Questionnaire to Assess Perceived Workload in Patient Monitoring Tasks: Pooled Analysis Study Using Mixed Models. Journal of Medical Internet Research, 22(9), e19472. https://doi.org/10.2196/19472

Simons-Morton, B. G., Guo, F., Klauer, S. G., Ehsani, J. P., & Pradhan, A. K. (2014). Keep your eyes on the road: Young driver crash risk increases according to duration of distraction. Journal of Adolescent Health, 54(5 SUPPL.), S61–S67. https://doi.org/10.1016/j.jadohealth.2013.11.021

Van der Laan, J.D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research - Part C: Emerging Technologies, 5, 1-10. http://www.hfes-europe.org/accept/accept.htm.

Appendix A Forms used in on-road study with the Technical Integration in-vehicle prototype

A.1 Consent form

Date: ______________

Consent to participate

Name of Research Project: Veoneer on-road study

Principal investigator(s): Prateek Thalya and Stefan Andersson

Dear participant,

Please read carefully the explanation of the study and the instructions and make sure you understand all parts of the research. In case you have questions please ask the experimenter. Then, please complete the following details and sign at the end of the document.

After reading the research description, I, undersigned

First Name: **Last Name:**
ID:

1. Hereby declare my willingness to participate in the experiment as detailed in the document describing the study
2. Hereby declare that the PI/Experimenter: ______________________________
Explained to me,
    - The nature of the study and I understand the purpose of the study and the instructions
    - That I am free to decide not to participate in the study and I am free to stop my participation at any time during the experiment. Decision against participation or termination of the study will have no negative consequences
    - That my personal identity will be kept confidential by any of those who are involved in the study and that my personal details will not be published in any publication including scientific publications.
    - That both video and audio will be recorded as part of the data collection.
3. I hereby consent whether the collected video and audio will be used for further analysis by the research team only.
    - ☐ Yes
    - ☐ No

Health declaration

I, undersigned

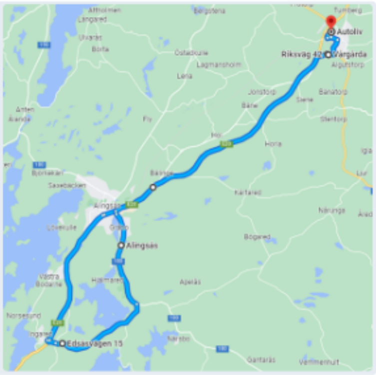
- Declare that I have no health issues and that I am not under any medical treatment.
- Declare that I am not suffering from nausea, head hakes, dizziness or any other issue that can disturb my driving or my ability to operate a computer.
- Declare that all parts of the experiment were explained in details and that all my questions regarding the experiments and its parts were answered.

A.2 Post-drive questionnaire form

Information about the drive

2

1.**Was there bad weather on any of the road segments?**



- ◯ No
- ◯ Fog
- ◯ Light Rain
- ◯ Heavy rain
- ◯ Snow falling
- ◯ Snow on the ground
- ◯ Other

3

How did you feel during the drive? *

	1	2	3	4	5	6	7
Drowsy (1) - Alert (7)	◯	◯	◯	◯	◯	◯	◯
Bored (1) – Excited (7)	◯	◯	◯	◯	◯	◯	◯
Calm (1) - Stressed (7)	◯	◯	◯	◯	◯	◯	◯

4

What best describe the traffic conditions during the drive? *

◯ Easy

◯ Medium

◯ Hard

5

How did you experience Mediator? Please indicate your level of agreement with the following statements. *

	1 Completely disagree	2	3	4	5	6 Completely agree
I would use Mediator if it was in my car	○	○	○	○	○	○
I consider Mediator as a beneficial system for (partly) automated driving	○	○	○	○	○	○
Driving with Mediator was comfortable	○	○	○	○	○	○
I can trust Mediator	○	○	○	○	○	○
Mediator is reliable	○	○	○	○	○	○
Mediator provides security	○	○	○	○	○	○
I felt very confident using the system	○	○	○	○	○	○
Mediator improved my awareness of the driving automation	○	○	○	○	○	○
When Mediator asked me to retake control, I was warned in an appropriate way	○	○	○	○	○	○
When Mediator asked me to retake control, I was warned with sufficient time to do so safely	○	○	○	○	○	○

A.3 Interview questionnaire

5.1.1.1. A.3.1 After Drive 2

Main questions	Sub questions	Check
<<Mediator in general>>		
How was your experience with Mediator during the first two trips?		O
Was there anything unexpected occurring during the drives? (i.e., Did you use the event button?)		O
What is your opinion about Mediator?	Was Mediator **reliable**?	O
	Did you **trust** Mediator? What would help you to have **more** trust?	O
	Was Mediator **easy** to **understand**?	O
<<Experience with automated driving >>		
Did Mediator help you improving your experience with assisted/automated driving?	Was Mediator created in a way that increased your trust in assisted/automated driving? How?	O
	Was Mediator increasing your **awareness** regarding the status of the assisted/automated driving system?	O
	Did Mediator increase your **perceived safety** while using assisted/automated driving?	O
<<HMI & Functionality of the system>>		
How would you rate the various elements of Mediator?	+++ screen with picture of whole dashboard (including LEDs, ambient lighting, display) were shown +++ Was the **information** provided by Mediator **clear**? (e.g., displayed images, sound concept, voice message, light concept)	O
	How was the **amount of information** provided? Was any important information **missing**? (Content or other modes like vibrations)	O
	Were you always **clear** about **why** you received the various **messages from Mediator**? (e.g., offer that Mediator can drive in assistant driving mode, takeover request to manual driving)	O
	How was the **timing** of the messages, i.e., the amount of time the messages were sent before an actual event (e.g., takeover) occurred?	O
	How clear was it to you who was **responsible for driving** and what you were allowed to do? i.e., **light concept** communicating the driving mode and an upcoming mode change	O
	Was it always clear to you, **when** assisted driving is available and for **how long** it is available?	O
	Did you activate assisted driving when it was **offered** by Mediator? Why? Why not?	O

	Did you follow Mediator’s requirement to drive **manually** immediately? Why? Why not?	O
	Was it clear to you when the actual **mode transfer** was happening and when it was finalized? Which information helped you in this regard?	O
	+++ screen with HMI image was shown +++ Was the **color scheme** of the street helpful? How?	O
	What do you think about the different **time budget** information (remaining time in a specific driving mode)? Do you think this has an added value to automated driving? Is there anything that should be changed?	O
	Was the information regarding the time budgets provided **reliable**?	O
	Did you experience any **driver state detection** events (e.g., distraction detection)? If yes: Was it helpful and why? If no: Do you think this would be a benefit to Mediator?	O
What is your overall opinion of Mediator?	What **benefits** did you encounter when using the system compared to driving without the system?	O
	What **drawbacks** (disadvantages / challenges) did you encounter when using the system compared to driving without the system?	O

5.1.1.2. A.3.2 After Drive 6

Main questions	Sub questions	Check
<<Mediator in general>>		
How was your experience with Mediator during the last trips?	Has your opinion regarding Mediator **changed** over the six drives? How? Why? e.g., Did you recognize new / other aspects, functionalities, icons, sounds… compared to the very first drives? Did you pay more attention to aspects that were not in your focus during the first drives?	O
Was there anything unexpected occurring during the drives? (i.e., Did you use the event button?)		
<<Comparison between the different drives>>		
You experienced different configurations of Mediator. Can you describe which differences you recognized?	+++ screen with pictures showing baseline vs. Mediator from the first drive were shown +++ In the first drive of this session, Mediator was not providing you the time budgets, the color scheme was not animated and you did not receive suggestions / warnings for take-over. How did you experience this configuration compared to the one last time you were driving? Did you miss the information? Which one did you miss the most? Do you have an idea, why you did not recognize the change?	O
	+++ screen with pictures showing short time budgets and repeated messages were shown +++ In the second drive of this session, Mediator was offering assisted driving also for very short durations and frequently offered you to change driving mode to assisted driving. How did you experience this configuration compared to the one last time you were driving? Do you have an idea, why you did not recognize the change?	O

5.1.1.3. A.3.3 After Drive 10

Main questions	Sub questions	Check
<<Mediator in general>>		
How was your experience with Mediator during the last trips?	Has your opinion regarding Mediator **changed** over the last drives? How? Why?	
Was there anything unexpected occurring during the drives? (i.e., Did you use the event button?)		
What is your opinion about Mediator?	Was Mediator **reliable**?	O
	Did you **trust** Mediator? What would help you to have **more** trust?	O
	Was Mediator **easy** to **understand**?	O
<<Trust>>		
Did Mediator help you improving your experience with automated driving?	Was Mediator created in a way that **increased your trust in automated driving**? How?	O
	Was Mediator increasing your **awareness** regarding the status of the automated driving system?	O
	Did Mediator increase your **perceived safety** while using automated driving?	O
<<Comparison between the different drives>>		
You experienced different configurations of Mediator. Can you describe which differences you recognized?	Last time we already talked about the drive where Mediator was not providing you the time budgets. We also talked about the drive where Mediator was offering assistant driving also for very short durations and frequently offered you to change driving mode to assistant driving. During your last visit, the **distraction warning was activated**. Did you experience any distraction detection warnings? When and why? What do you think about this functionality? +++ screen with slide showing the different configurations was shown +++ Please **order** the different configurations (Mediator standard, Mediator standard with distraction warning, Mediator without time budgets and recommendations, Mediator with short durations of assisted driving) regarding your preferences.	O
	Which functionalities / parts of Mediator did you **value the most**? / Did you miss the most?	O
	Which functionalities / parts of Mediator are **not so important / helpful** in your opinion? / Did you miss the least?	O
<<HMI>>		
How would you rate the various elements of Mediator?	+++ screen with picture of whole dashboard (including LEDs, ambient lighting, display) & HMI image were shown +++	O

	For the test drives, the Mediator information were shown on one screen in the center stack. Which **positions** would you prefer for the information to be provided to you?	
	Do you want information **why** a certain driving mode is available or no longer available? Do you **always** want this information? On what does it depend on (e.g., experience or type of reason)?	O
	How was the **timing** of the messages (i.e., the amount of time the messages were sent before an actual event (e.g. takeover) occurred?	O
	Would you like the opportunity to get some functionalities (e.g., the timing for the messages) **individualized**? What information / functionalities should be adaptable due to individual preferences?	O
	How **far in advance** should the next section of the route showing the availability of a different driving mode be displayed? (e.g., 10 minutes, 1 h, right at the beginning of the drive / after changing the driving mode no matter how long it takes to reach the section)	O
	How long should the **availability** of a particular mode be indicated (e.g., via the image on the upper right corner) if the driver does not activate it? The color coding of the displayed street will always be available. Should the availability of the mode be indicated **acoustically**? If yes, **how often** (number of alerts, frequency) should the driver be reminded?	O
	How long needs the automation have to be available for it to be displayed at all?	O
<<Potential for improvement>>		
Within our project, we are also investigating use cases in which Mediator can offer an autopilot mode. What situations can you think of where you would like to have this kind of support?	Imagine Mediator **detecting a traffic jam** ahead and offering to drive for you through the traffic jam. Would you like to have such a functionality?	O
	Imagine Mediator **detecting an incoming email** and offering to drive for you so you can read and answer the email. Would you like to have such a functionality?	O
	Imagine Mediator **recognizing that you are getting tired** and offering to drive for you for a while so you can rest. Would you like to have such a functionality?	O
	For which **levels of automation or use cases** would you consider such as system as most useful? (e.g., fatigue, distraction, comfort, SAE level 2 supervision, SAE level 3 Takeover request…)	O
What is your overall opinion of Mediator?	What **benefits** did you encounter when using the system compared to driving without the system?	O
	What **drawbacks** (disadvantages) did you encounter when using the system compared to driving without the system?	O

Would you use Mediator in your future car? Please explain why yes or no or under what conditions?	What would you change to **improve** Mediator?	O
	What do you see as **main challenges** of such a system to be introduced to the market? (e.g., technical challenges, implementation, user interaction, privacy, acceptance…)	O

A.4 Average ratings for each driver

5.1.1.4. A.4.1 Reliability

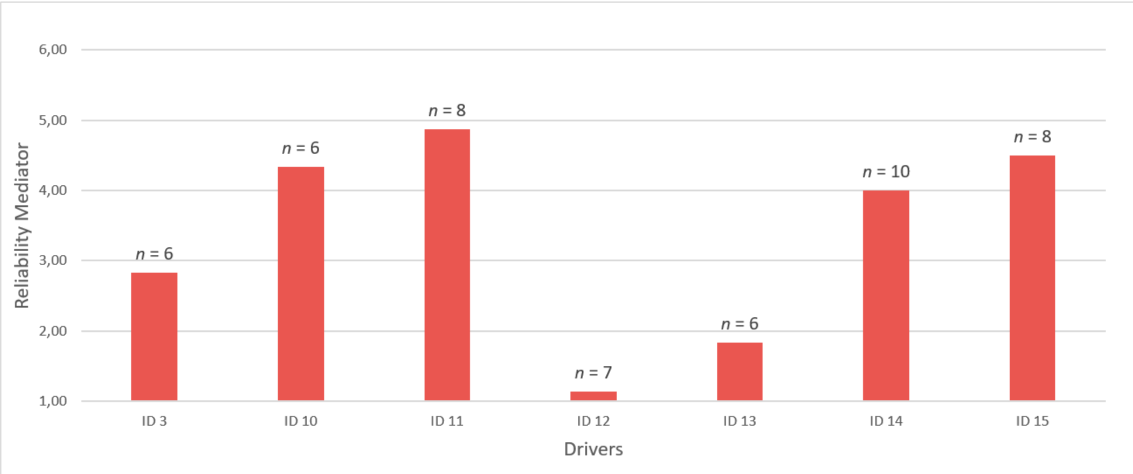


Figure 5.1 Average answer to the statement "Mediator is reliable." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree over all drives for each driver ID. n shows the number of valid values per driver.

5.1.1.5. A.4.2 Appropriateness of Warnings

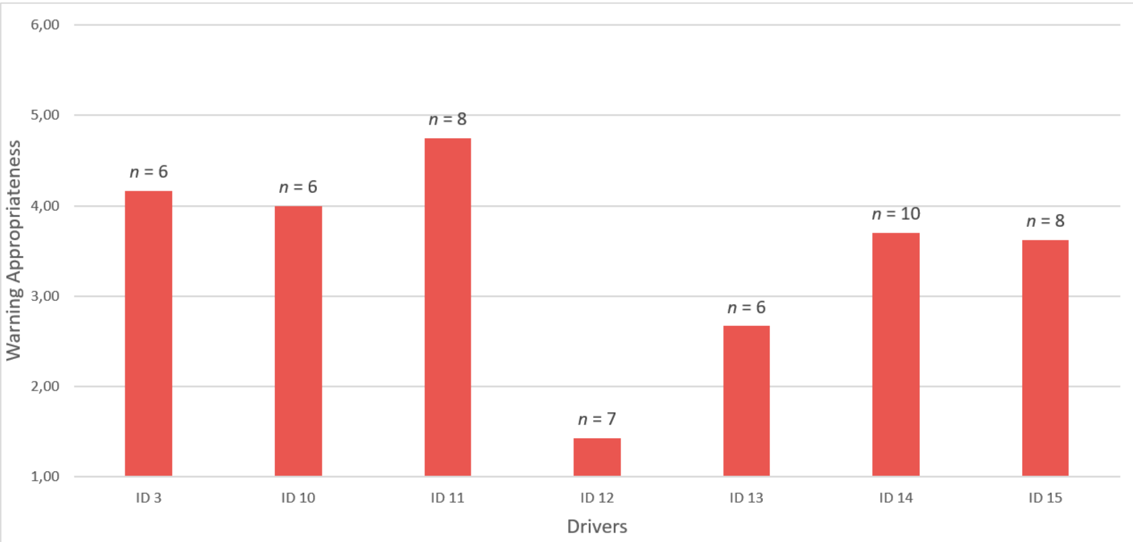


Figure 5.2 Average answer to the statement "When Mediator asked me to retake control, I was warned in an appropriate way." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree over all drives for each driver ID. n shows the number of valid values per driver.

5.1.1.6. A.4.3 Timing of Warnings

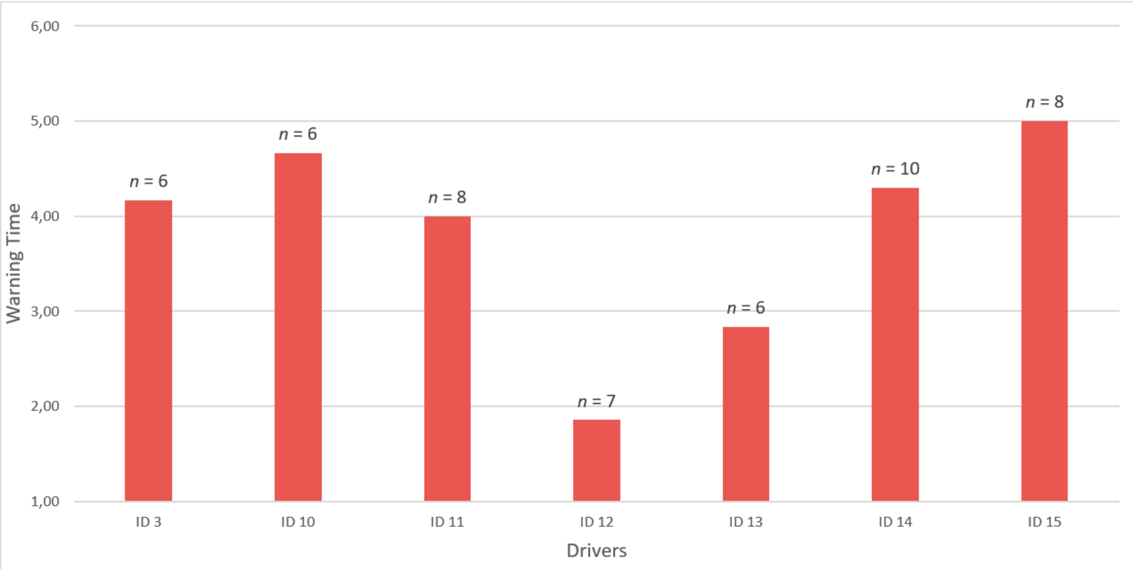


Figure 5.3 Average answer to the statement "When Mediator asked me to retake control, I was warned with sufficient time to do so safely." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree over all drives for each driver ID. n shows the number of valid values per driver.

5.1.1.7. A.4.4 Trust

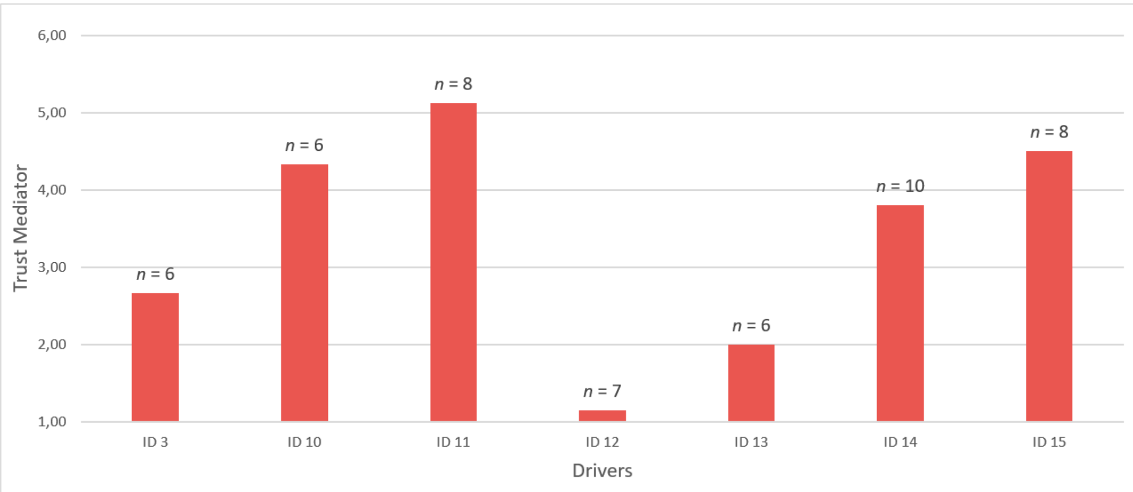


Figure 5.4 Average answer to the statement "I can trust Mediator." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree over all drives for each driver ID. n shows the number of valid values per driver.

5.1.1.8. A.4.5 Usage Intention

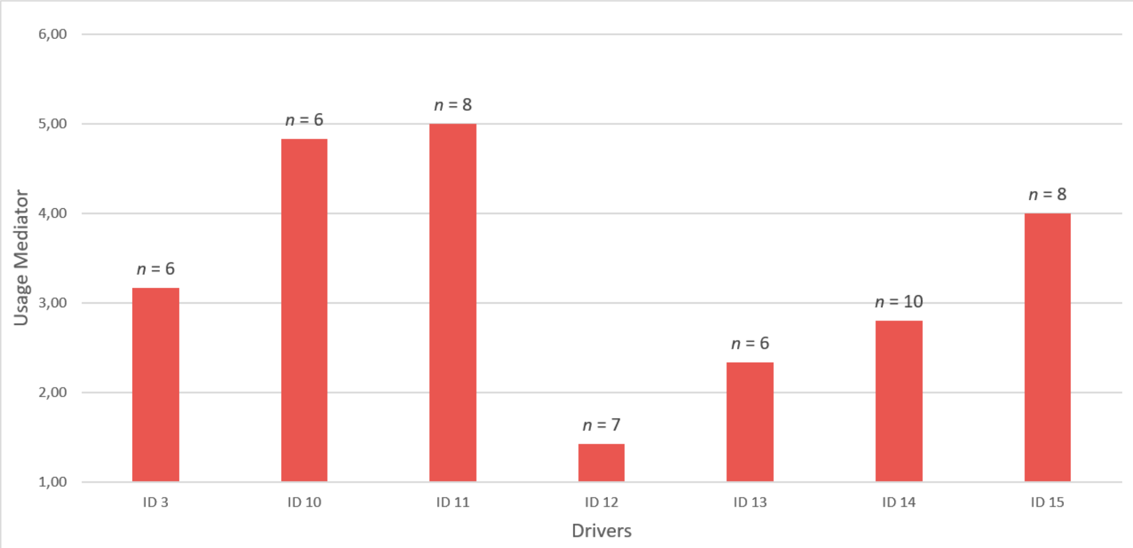


Figure 5.5 Average answer to the statement "I would use Mediator if it was in my car." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree over all drives for each driver ID. n shows the number of valid values per driver.

5.1.1.9. A.4.6 Comfort

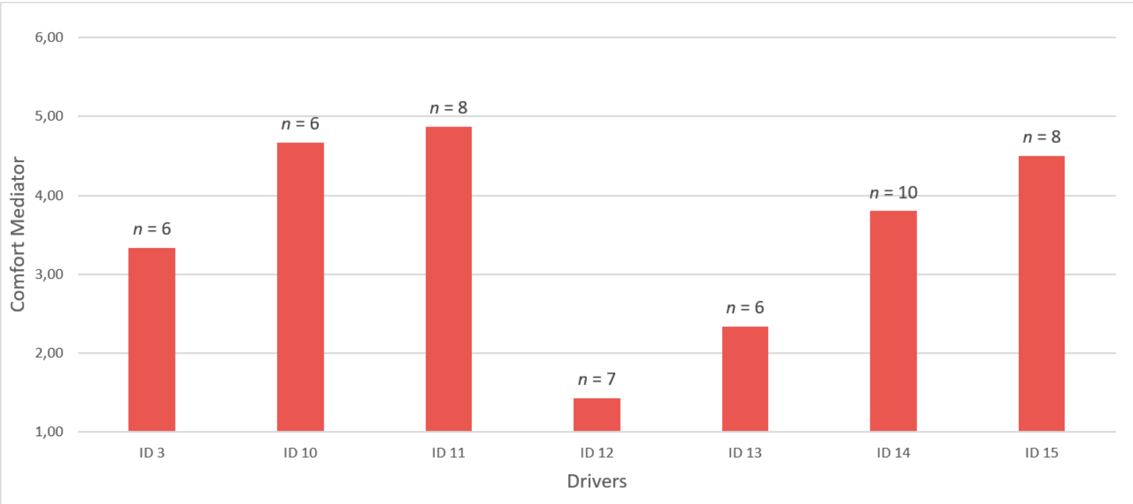


Figure 5.6 Average answer to the statement "Driving with Mediator was comfortable." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree over all drives for each driver ID. n shows the number of valid values per driver.

5.1.1.10. A.4.7 Confidence in Using Mediator

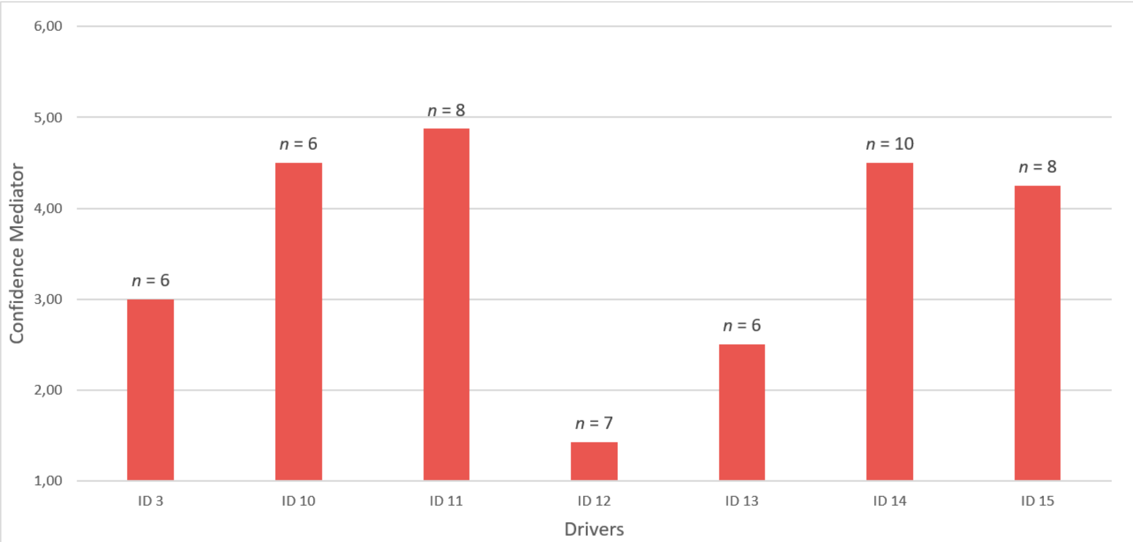


Figure 5.7 Average answer to the statement "I felt very confident using the system." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree over all drives for each driver ID. n shows the number of valid values per driver.

5.1.1.11. A.4.8 Perceived Safety

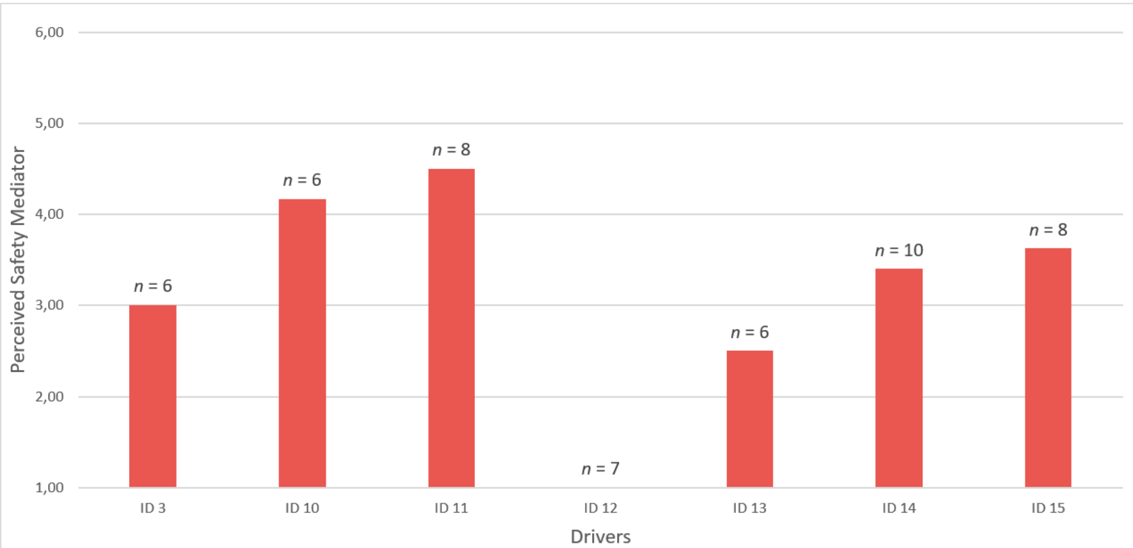


Figure 5.8 Average answer to the statement "Mediator provides security." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree over all drives for each driver ID. n shows the number of valid values per driver.

5.1.1.12. A.4.9 Awareness

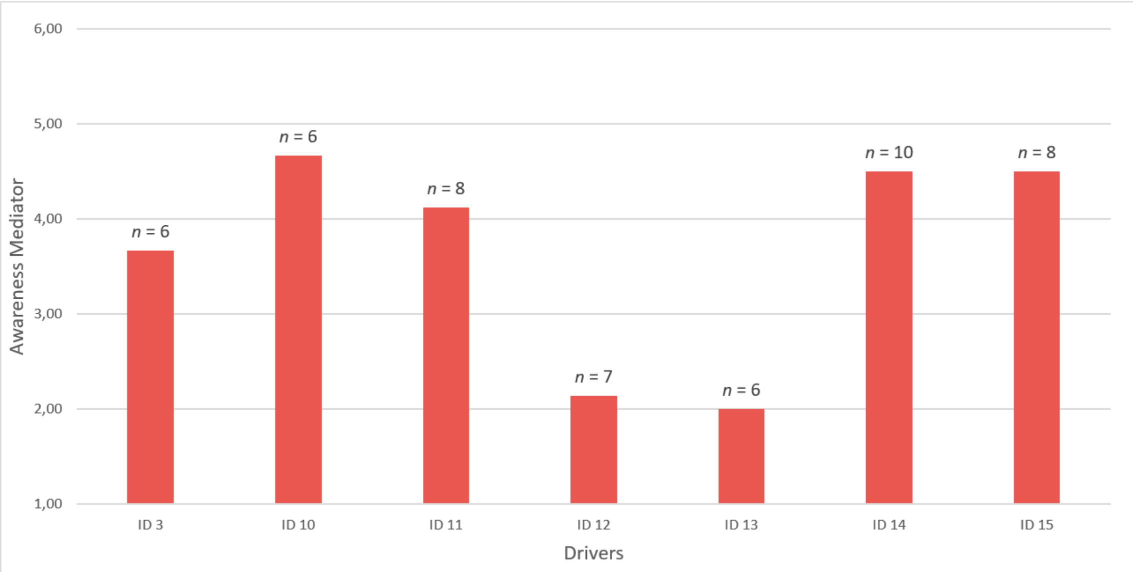


Figure 5.9 Average answer to the statement "Mediator improved my awareness of the driving automation." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree over all drives for each driver ID. n shows the number of valid values per driver.

5.1.1.13. A.4.10 Benefit

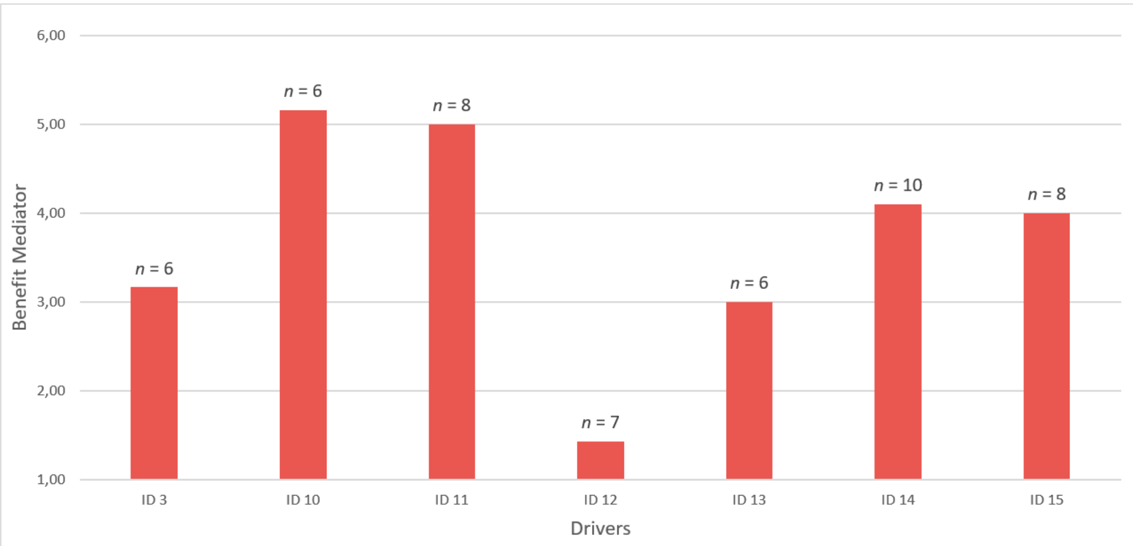


Figure 5.10 Average answer to the statement "I consider Mediator as a beneficial system for (partly) automated driving." on a 6-point Likert scale ranging from 1 – completely disagree to 6 – completely agree over all drives for each driver ID. n shows the number of valid values per driver.