EINDHOVEN UNIVERSITY OF TECHNOLOGY

MASTER THESIS

A Real Road Study of Automated Driving: The Influence of the Non-Critical Cue on Automation Surprise

Author:
Yuanzi Wang

Supervisor:
prof.dr. M.H.Martens

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science in Automotive Technology

in the

Future Everyday Group
Department of Industrial Design

February 2, 2023

Abstract

Human Machine Interface (HMI) is becoming increasingly important in Automated Vehicles (AVs) due to its key role as a link between vehicles and humans. We cannot comprehend it simply as visual or audio displays but also as channels or systems that interact with the users in the AVs. In level 3 and higher automated driving, the driver can perform Non-Driving Related Task (NDRT) while the Automated Driving System (ADS) operates the vehicle, as long as no critical events or no Take-over Requests (TORs) occur. An HMI should help avoid the Automation Surprise (AS) and minimise disturbances to NDRTs in this non-critical Highly Automated Driving (HAD) scenario.

This thesis develops a non-critical cue concept for the HMI in AVs. And we conducted a real-road experiment with the Wizard of Oz (WoZ) vehicle on a highway in Eindhoven. The within-subject experiment aims to test the influence of the non-critical cue on the driver's experience of the AS and distraction. Qualitative-oriented mixed methods are applied in the study with the data we collected, including questionnaires, video recordings, and interview recordings.

The experimental results suggest that the non-critical cue could potentially reduce the automation surprise in terms of observability, predictability and timeliness. Specifically, the observability could be improved by the non-critical cue directly or together with other factors, such as observation of the surroundings and the car dynamics. The car dynamics could function as a non-critical cue in terms of observability under certain condition. Additionally, the non-critical cue could distract the driver conducting NDRTs. However, the driver may feel distracted and unsafe if the non-critical cue is absent.

Contents

List of Figures

List of Tables

Acronyms

ACC Adaptive Cruise Control . . . 6
ADAS Advanced Driver Assistant System . . . 24
ADS Automated Driving System . . . ii, 1–4, 6, 8, 12
AS Automation Surprise . . . ii, 3–5, 9, 11, 12, 32–36, 43, 44, 47
AUTOAccD Automatic Acceleration and Data controller . . . 19, 50
AV Automated Vehicle . . . ii, iii, v, 1–7, 9–13, 15, 16, 32, 33, 35, 36, 39–42, 44–48
CQR Consensual Qualitative Research . . . 29
CUI Conversational User Interface . . . 12
DDT Dynamic driving task . . . 1, 5, 6, 8
GUI Graphical User Interface . . . 12
HAD Highly Automated Driving . . . ii, 4, 46–48
HMI Human Machine Interface . . . ii, iii, v, 1–3, 5, 7, 9–11, 32, 38, 42
IMU Inertial Measurement Unit . . . 22, 28, 47
IP Instrument Panel . . . 14–16
IRR Inter-rater Reliability . . . 25–28
LC Lane Centring . . . 6
LRT Light Rail Transit . . . 19
MSSQ-Short Motion sickness susceptibility questionnaire short-form . . . vi, 19, 52, 53
NDRT Non-Driving Related Task . . . ii, 2–4, 7–9, 12, 14, 16, 17, 32, 35, 37, 38, 42, 45, 47, 57
NHTSA National Highway Traffic Safety Administration . . . 5, 10
non-critical HAD Non-Critical Highly Automated Driving . . . 2–5
ODD Operational Design Domain . . . 1, 2, 4, 6, 36
RRADS Real Road Autonomous Driving Simulator . . . v, 12, 13
SA Situational Awareness . . . 8
SAE Society of Automotive Engineers . . . 1, 2, 4–6
TOR Take-over Request . . . ii, 2, 6, 10, 17, 43
TOT Take-over Time . . . 2, 43
UEQ User Experience Questionnaire . . . vi, 21, 24, 56
WoZ Wizard of Oz . . . ii, iii, v, 4, 5, 12–20, 40, 46

1 Introduction

Nowadays, cars are evolving from entirely manually operated machines to ones that include a variety of automated functions and are capable of making decisions on their own. When the vehicles are driving automatically, the vehicle may take actions not aligned with the driver's goal. Then the driver would perceive the vehicle's behaviour consciously or subconsciously, sometimes by combining the environmental information to collaborate with the vehicle. Therefore, the vehicle and its user should be seen as a joint cognitive system (Carsten and Martens, 2019). Then the Human Machine Interface (HMI) is becoming increasingly important as a bridge between automated vehicles and people in this system. It should help people complete the role transition from driver to passenger in certain scenarios conducting non-driving tasks. During this progress, the HMI should also communicate with the driver to achieve better interaction quality (Detjen et al., 2021). Here when we talk about HMI, we cannot comprehend it simply as visual or audio displays in the vehicles interacting with the users. It should be channels or systems that include various inputs from the driver to the vehicle and vice versa (Carsten and Martens, 2019). For example, the driver could instruct the car by touching the visual display as input. The car could provide haptic controls with somatosensory tactile feedback, such as in the driver's seat (Detjen et al., 2021). Meanwhile, the vehicle's dynamics and movements within a specific context, for example, resistance, pulses, vibrations, and physical guidance, are also the feedback from the vehicle considered to be part of the HMI used to guide and assist the human (Carsten and Martens, 2019).

According to the technological capabilities and human engagement, AVs can be classified from manual driving, where the human driver executes all of the driving tasks, to fully autonomous driving, where the human plays no role in driving (Kyriakidis et al., 2019). Among all these levels of automation, human-vehicle interactions exist. Different HMI would be adopted vary from different automation levels of vehicles. In this study, we adopted the standard Society of Automotive Engineers (SAE) International levels of automation (SAE International, n.d.), in which six levels are defined from level 0 (no automation) to level 5 (fully automation driving level). Starting from automation level 3, the Automated Driving System (ADS) itself begins monitoring the driving environment, and it performs all aspects of the Dynamic driving task (DDT) (in conditions for which it was designed). The level 3 automation is limited in its specific Operational Design Domain (ODD), which are the limited conditions that the automated system could work with. For example, the ODD of the level 3 traffic jam feature is an expressway in heavy traffic with access fully controlled (SAE International, n.d.) Within its limited ODD, the vehicle could be driven by ADS, and the driver could engage in the NDRTs. Still, the human driver is expected to respond appropriately to a request from HMI and to intervene. When it is outside the boundary of the ODD, such as leaving the highway or in terrible weather, it should be the driver to control the car. The HMI continue to play an essential role in the AVs. The response time of the driver evokes much research When the vehicle issues the TOR via HMI while the driver is conducting NDRTs(Zhang et al., 2019). The most often used measurement in the literature appears to be Take-over Time (TOT), which is defined as the time it takes for drivers to reclaim control from automated driving following the critical event in the environment or after receiving a TOR (Zhang et al., 2019). The takeover is a very popular topic of user interface research in automated driving (Ayoub et al., 2019). However, less HMI research focused on the general communication between the driver and the AV while the AV performs driving within its ODD.

Compared with the safety-critical situations, the period when the driver does not control the vehicle is the general situation when the ADS of AVs functioning normally in a SAE level 3 and higher level AVs. That is to say, this scenario would occur more frequently and last longer than the vehicle reaching its ODD limitation when travelling on the road qualified for automated driving function. Meanwhile, during this non-critical driving period, we want the driver to conduct the NDRTs comfortably without being surprised or disturbed by the unexpected manoeuvres of the vehicle. Thus, we want to focus on the driver's experience in the AVs of level 3 and higher, i.e., in highly automated driving. And to be more specific, we narrow down our attention to the normal conditions when the AVs of SAE level 3 and upper function automatically within its ODD where no emergency and TORs occur. We named this scenario as **Non-Critical Highly Automated Driving (non-critical HAD)**, and all our design and research are within the scope of this non-critical scenario shown in Figure 1.1.

Within the defined scope of this research, while the driver is focusing on the NDRTs, there is a possibility for the driver to get surprised because of the events that happened in the environment and the behaviour of the car. For example, the driver might be disturbed by the AV overtaking a huge truck when playing on the phone in the AV, and ADS is performing the driving tasks. The reason for the surprise could be the truck is quite huge and close to the AV, causing the sense of lack of safety. There is also no expectation from the driver that the AV would conduct the overtaking of the truck as the driver may be out of the monitor

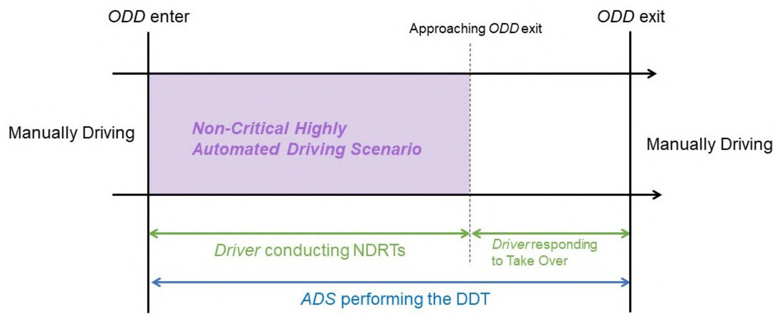

Figure 1.1: The non-critical highly automated driving scenario

loop and concentrating on the phone. A specific term "Automation Surprise (AS)" is defined in the aviation field, stemming from the pilots' inability to understand the detected conflict between automation and pilots (Dehais et al., 2015a). Victor et al. (2018) used automation expectation mismatch to describe the conditions when operators are surprised by automation. They do not understand why it behaves in a certain manner or what it will do next (Victor et al., 2018). In 2019, Carsten and Martens (2019) linked the AS to automated driving. They used the term primarily for situations where the driver actually notices the action taken by the vehicle is inconsistent with the driver's previous expectation.

With the development of automated vehicles, conflicts between the driver and the automated vehicle gradually appear. Here in this research, we re-defined the term "**Automation Surprise (AS)**" in the automotive field in the scenario of non-critical HAD scenario, that is, **in a conditionally Automated Vehicle (AV) where the driving tasks are performed by Automated Driving System (ADS), the driver is surprised by the unexpected behaviours of the vehicle and events in the environment while conducting Non-Driving Related Tasks (NDRTs)**.

Within the scope of this project, to be more specific, in the conditionally automated driving scenario while the driver is conducting NDRTs, we designed a cue as the HMI named ***non-critical cue*** that improves the user experience from two aspects. First, we want the non-critical cue to help the driver avoid the AS. It means the driver would not be negatively surprised by the behaviour of the AV and the events in the environment. Second, the driver should not be disturbed from conducting NDRTs by either the cue itself or other factors. Then the driver can concentrate on the NDRTs rather than keep looking up, monitoring the environment and the vehicle's movement.

To realize this, we would like the non-critical cue to be able to help the driver be aware of the situation to some extent but not drag the user back to the monitor loop suddenly and explicitly. Inspired by calm technology, we hope the non-critical cue could support human users to reconnect with their environment by providing information to the periphery of users' attention, i.e., cue the users implicitly (Weiser and Brown, 1996). Therefore, we did not employ explicit interaction interfaces, such as the visual display showing the status of the vehicle with text on the screen. Instead, we applied a vibrotactile cue and tested it in a real-road experiment with an experiment car modified based on Wizard of Oz (WoZ) method.

Research Goals

In the scenario of SAE level 3 conditionally automated driving, we envision that the AVs within this scope maintains the same interior as the manually driven vehicles. In an L3 AV, the cabin would still be equipped with the instrument panel with dashboard, steering wheel and brake. When the driving tasks are performed by ADS within the ODD, the signals from the manually driven vehicles are inherited. The speedometer on the dashboard would show the speed changing, and the steering wheel rotates itself. These signals should be able to be detected by the driver. Vehicle dynamics such as longitudinal and lateral acceleration also influence the driver's experience as feedback from the car (Carsten and Martens, 2019). The driver could sense when the vehicle was accelerating or decelerating because of the resistance and forces. We call all the mentioned signals and the car dynamics in a SAE level 3 vehicle as ***inherent cues***. In this project, we also propose the concept of the non-critical cue under the highly HAD scenario to help reduce the AS to the driver without interrupting the driver's NDRTs, thereby improving the quality of the journey with conditionally AVs. We would like to investigate the role of the non-critical cue and the inherent cues in conditionally AVs. The research questions are as below:

Research question 1: How does the non-critical cue influence the AS and distraction from conducting the NDRTs in the non-critical HAD scenario?

Research question 2: How does the inherent cues in the SAE level 3 AVs influence the AS?

Background 2

Within the scope of the non-critical HAD scenario in SAE level 3 AVs, we developed a non-critical cue and tested it in an experiment car to see if it can help reduce the AS by on-road experiment trials. In this chapter, literature and related work to this project will be reviewed and explained to help the reader have a better understanding of the topic.

There are two sections in this chapter, and each includes several subsections. In the first section, the general background information will be explained. The automation level will be described in detail to distinguish the difference between each level and among different standards. In this part, we further clarify the scope of this project. The second part of this section explains the calm technology and peripheral interface, which is the inspiration and foundation of the project's aim. Regarding the third part in this section, the HMIs in the vehicles are discussed. Then, we explain the AS in detail. The second section discusses the methodology we applied, including the Wizard of Oz (WoZ) method and the on-road experiments.

2.1 Background information

2.1.1 Automation Level

The automation level of AVs is measured by various standards. In the National Highway Traffic Safety Administration (NHTSA) classification, shown in Figure 2.1, the automation level is divided from level 0 (momentary driver assistance) to level 5 (full automation). According to NHTSA, from level 3 of automation, the system will operate the driving tasks. However, the driver should remain available to resume driving in NHTSA level 3. In level 4 or higher, the driver will perform more as a passenger and could manoeuvre the vehicle no more in a specific area or even on all conditions (National Highway Traffic Safety Administration, n.d.) However, in the PACT framework, the car is still in the assisting mode at level 4 (Richards and Stedmon, 2016).

In this research, we adopt the SAE (SAE International, n.d.) shown in Table 2.1. From level 0 (no automation) to level 5 (fully automation), six levels are defined in the standard. In addition, some terms were defined by SAE International (n.d.) to clarify the levels of automation of vehicles. Dynamic driving task (DDT) refers to all of the real-time operational and tactical functions required to operate a vehicle in on-road traffic. It excludes strategic functions such as

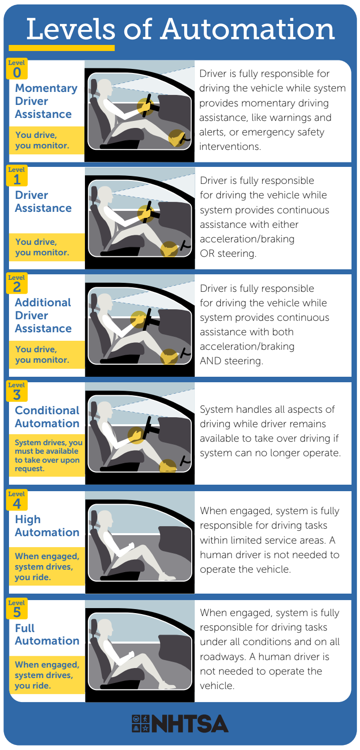

Figure 2.1: NHTSA Automation Levels (National Highway Traffic Safety Administration, n.d.)

trip scheduling and selection of destinations and waypoints. For example, operational tasks like vehicle motion control via steering, acceleration or deceleration, and tactical tasks such as manoeuvre planning are not DDTs. The DDTs could be performed by the human driver or the Automated Driving System (ADS). Here theADS means the hardware and software that can continuously execute the complete DDT. Particularly, ADS is a specific term for automation level 3 to level 5. Moreover, under level 5, the ADS could only perform the DDT under specific conditions, which are called Operational Design Domains (ODDs). For example, the implementation of LC could be specifically at low speed, high speed, or all speeds, which refers to different ODDs Especially, ODD limitations are only considered levels 1 through 4 rather than level 5 (SAE International, n.d.)

Monitoring of driving environment	Automation level	Definition
Human	Level 0	No Driving Automation
	Level 1	Driver Assistance
	Level 2	Partial Driving Automation
Automated Driving System	Level 3	Conditional Driving Automation
	Level 4	High Driving Automation
	Level 5	Full Driving Automation

Table 2.1: Levels of automation as defined by the SAE International (SAE International, n.d.)

With the concepts in mind, we can differentiate the automation levels in SAE easier. According to SAE International (n.d.), from level 0 to level 2, the driver performs part or all of the DDT. Starting from level 3, the conditional automation, the Automated Driving System (ADS) controls the entire DDT when engaged. Here we would like to specify the difference between level 3 and level 4 automation. In level 3, an ADS performs all aspects of the DDT within its limited specific ODD, and the human driver is expected to respond appropriately to a TOR. When in level 4, an ADS performs all aspects of the DDT the same as it in level 3. However, the user does not need to supervise or respond to a request to intervene. Because the level 4 feature AVs could automatically perform a minimal risk manoeuvre such as pulling over the vehicle to the side of the road within itsODD. Specifically, during a given vehicle trip, such as the robotaxis on the road with designed infrastructures, ADS with level 4 features could operate the vehicles with full automation (SAE International, n.d.) Currently, production vehicles support SAE Level 1 or Level 2 driving automation systems with ACC and LC (He et al., 2022). Automation level 3 will be the focal point of the popularization of automated vehicles in the near future. The experiment scope of this project is under automation level 3 defined by SAE.

2.1.2 Calm Technology and Attention Theories

Calm technology is the inspiration for how we would like to develop a non-critical cue in the AV rather than inform the user explicitly about the information. In this section, calm technology, peripheral attention, and attention levels according to situational awareness are introduced.

Calm Technology and Peripheral Interface

Weiser and Brown (1996) coined the term calm technology. They said, "*Calm technology engages both the centre and the periphery of our attention, and in fact moves back and forth between the two*.". The "periphery" here means the things that people are attuned to without attending to explicitly. However, we cannot imply that the periphery is trivial. It could move to the centre or the focal point of attention in the next moment and become crucial (Weiser and Brown, 1997). To be more specific, along with the ability to shift to the centre of our attention, the periphery increases our knowledge and ability to act without increasing information overload (Weiser and Brown, 1997). Calm technology should be able to help bring more details, including what just happened, what is happening, and what is going to happen, into the periphery rather than the centre of attention. Thus the user of the technology will not be surprised by the pre-attentive periphery filled with details engaged (Weiser and Brown, 1997).

A kind of calm technology known as peripheral interfaces enables a person to be aware of information from various aspects without feeling overburdened. (Matthews, Rattenbury, and Carter, 2007). When the user is conducting a primary task, peripheral displays allow the user to monitor a separate information source (De Guzman et al., 2004). For example, De Guzman et al. (2004) designed an expandable plastic ball that shows the status of the online cooperation partners. When the partner is offline, the ball would collapse and would expand to different sizes depending on the activities when the partner is online. A pulsing effect of the ball represents the message from the partner. The display allows the user to monitor pals' status without purposefully checking the computer, and reduces distraction from partner status changes so that they can concentrate on a different main work (De Guzman et al., 2004).

Under the context of our project, the primary task of the driver is conducting the NDRTs and it should be in the driver's centre of attention. Ideally, the events in the environment and the behaviour of the vehicles should be in the periphery of the driver with the help of the HMI in the vehicle.

Situational Awareness and Attention Levels

Situational Awareness (SA) is viewed as consisting of a person's state of knowledge about a dynamic environment. It incorporates the perception and comprehension of the elements in the environment and a projection of future states of the environment based on this understanding (Endsley, 1995). Matthews, Rattenbury, and Carter (2007) defined peripheral awareness as the amount of information displayed by the interface that people can use without focusing their attention. It can be considered as a subset of SA (Matthews, Rattenbury, and Carter, 2007). In the case of the driving scenario, the SA refers to the cognitive process that includes Scanning the visual field, Predicting what might happen, Identifying situations and objects, Deciding what to do next, and Executing various driving Responses, which is the SPIDER model. This attention-driven cycle is the basis for SA in driving and virtually all other complex activities that people perform in the vehicle (Sanquist, Brisbois, and Baucum, 2016). With ADS performing the DDT, there is no need for the driver to scan and identify the situation continuously. However, maintaining a certain level of anticipation could reduce the surprise by sudden events as the driver is not monitoring the environment consistently. Then the driver could decide to keep focusing on the NDRTs or prepare for resuming driving.

Especially for the tactile interface in the vehicles, Riener et al. (2017) differentiate between four classes of tactile stimuli, which are (a) Ignore, (b) Change blind, (c) Make aware, (d) Demand action, corresponding to releasing increased levels of attention (LOA).

The first attention level "Ignore", relates to the information from the environment. In this attention level, the driver is not required to pay attention to environmental information, such as the traffic condition at the intersection or whether the following car is keeping a safe distance when the driver interacts with fully automated vehicles. This attention level usually suits the user in the level 5 automation vehicle, which would be manoeuvred sustainably and unconditionally by ADS (SAE International, n.d.) The second level, "Change blind", is associated with the cues or notifications from the ADS. The driver will not manipulate immediate action to the cues when at this attention level. However, the notifications may evoke an action from the driver in a longer time (Riener et al., 2017). An example is the adaptive driver alert system designed byBeyer et al. (2010). The system could attain the level of driver's distraction and then give unobtrusive visual notification in the viewing direction to the driver to increase driver attention gradually (Beyer et al., 2010). There is no need for the driver to react to the cues from the system immediately. However, the distraction would be decreased gradually. The third and fourth levels are “Make aware” and “Demand action”, which are both related to the notification from the vehicle system to the driver. The "Make aware" level should be used when immediate action is required, but no danger would be caused if the action was delayed. The level "Demand action" requires the driver’s full attention and immediate action must be conducted to prevent hazardous situations(Riener et al., 2017). For example, the tactile low-fuel signal belongs to the "Make aware" level, because refuelling is the action required to be done in the short term and no emergency would be caused if the driver kept driving to the next gas station. The lane departure warning would be a notification in the “Demand action” level as delaying lane correction may cause a car crash.

In the level 3 automation, while the driver is conducting NDRTs, it would be ideal to keep the driver in the attention level "Change blind", where the driver could pay attention to their own tasks and not be scared or surprised by the events happening in the environment, such as braking and turning at the crossing or overtaking by a huge truck. Seeking a balance between the driver’s attention on the task and the environment is an important task of HMI in AVs.

2.1.3 HMI for AVs

Automotive HMI is a broad and complex concept, which includes a wide variety of designs and needs to adapt to different levels of autonomous driving. While within the same driving level, HMI can be designed for different purposes, such as assisting the driver in taking over or improving the driver’s trust and acceptance level to the AVs. There are many ways to classify HMI. Generally, HMI can be classified as visual, auditory, or tactile HMI. Boelhouwer (2021) created a framework with three categories, including suppleness, bodily experience, and situatedness to code the HMI, where design opportunities can also be discovered (Boelhouwer, Dijk, and Martens, 2019). Meanwhile, other standards have also been proposed to support HMI design and evaluation. For example, according to Carsten and Martens (2019), six goals are crucial for HMI design in AVs. They are 1. providing a required understanding of the AVs capabilities and status (minimise mode errors); 2. generating correct calibration of trust; 3. stimulating appropriate levels of attention and intervention; 4. minimise ASs; 5. provide comfort to the human user, i.e. reduce uncertainty and stress, and 6. be usable.

In the past decade, various human-machine interfaces (HMI) emerged with the continuous development of AVs. While we are advancing from the current partial automation (Level 2) to higher levels of automation, human taking over control when the system reaches its limit (e.g., automation failure, adverse weather) attract great attention, especially in conditional automation (Level 3) (Ayoub et al., 2019). 5.1% of total automotive interface design from 2009 to 2018 focused on taking over, which was the most popular user interface for AVs. Designing the interface providing trust and acceptance of AVs to the users (4.3%) is the second most popular domain(Ayoub et al., 2019). From the data, we can see the duration of vehicles conducting automated driving was not the major topic in the vehicle HMI design.

For the Take-over Request (TOR), two commercial models of Audi and Cadillac had already been equipped with the system formally. Besides, a variety of literature concepts were generated. Visual and audio messages were still the main output of the take-over request system, although some concepts combined vibration with physical shape-changing in the interfaces (Boelhouwer, Dijk, and Martens, 2019). For the HMI during automation, the visual display was used by all systems in commercial cars and most concepts in the literature regarding bodily experience. Driver's attention became one of the main focuses of the interfaces. For example, Karatas et al. (2017) designed mini robots on the dashboard, turning heads toward the road to redirect the driver's attention. Telpaz et al. (2015) used vibrations in different parts of the driver's seat to indicate approaching vehicles when the driver's attention was not on the road (Boelhouwer, Dijk, and Martens, 2019). The two examples were under the condition of level 3 automation defined by NHTSA, where the driver is supposed to monitor the system and expected to be available for both occasional control and full responsibility for driving within an acceptable amount of time(Telpaz et al., 2015).

As mentioned before in the introduction part, the HMI is not narrowly a display. It should be channels or systems that include various inputs from the driver to the vehicle and vice versa (Carsten and Martens, 2019). According to Carsten and Martens (2019), *the term "HMI" will here be used in that broad sense to encompass the full range of explicit as well as implicit communication between the human operator and the vehicle.* For instance, shown in Figure 2.2 Walker et al. (2021) developed and evaluated a novel in-vehicle interface with ambient light pulsing synchronously with the breathing pattern of the driver. The experiment shows there is no awareness of the ambient light manipulation, to be more specific, no physiological arousal was reported from the participants with this light-respiratory prototype. However, participants respond faster to the TORs (Liu et al., 2022). This is a perfect in-vehicle HMI example of engaging peripheral attention in the context of a level 3 automated vehicle and increasing the ability to act when attention moving to the centre (Weiser and Brown, 1997).

Figure 2.2: Light-respiratory HMI in AV (Walker et al., 2021)

2.1.4 Automation Surprise

As mentioned in the introduction, the research goal of this project is to explore whether the non-critical cue could reduce the AS in the conditionally automated driving scenario. The term Automation Surprise (AS) is firstly defined in the aviation field. It stems from the pilots' inability to understand the detected conflict between automation and pilots (Dehais et al., 2015b). Gaps and misconceptions in operators' mental models of automated systems are important aspects of human performance reduction, such as loss of situational awareness. Consequently, operators are often surprised by the automation; they do not understand why it behaves in a certain manner or what it will do next (Victor et al., 2018). Carsten and Martens (2019) linked the AS to automated driving. They used the term primarily for situations where the driver actually notices the action is inconsistent with the driver's expectation. Two types of AS were defined, which were the absence of expected action and the presence of unexpected action, respectively (Carsten and Martens, 2019). The first type of surprise often occurs when humans continuously monitor the automation system and the environment. Humans generate expectations of AVs' behaviour while monitoring, but vehicles do not behave as expected. The second type is more noteworthy for designing HMIs for level 3 and 4 automation driving systems. Because in level 3 and higher automated driving levels, people rarely or do not monitor the autonomous driving system, at this time, the conditions in the environment and the sudden reaction of AVs to it are more likely to alert or disturb the user.

Sarter and Woods (1995) defined the mode awareness for automation system as *"the ability of a supervisor to track and anticipate the behaviour of automated systems."* This concept was connected to AS by Monsaingeon et al. (2021) in automated driving. According to Monsaingeon et al. (2021), low mode awareness will cause mode confusion or mode error, which are typical examples of AS. Besides, mode awareness was used by them as the measurement in the human-automation interaction because high awareness could ensure the drivers' safety without provoking confusion(Monsaingeon et al., 2021). However, in our project, we would like to use AS to appraise the human-automation interface directly. Apart from helping the driver to take over the vehicle in time to ensure safety in an emergency circumstance, the interface in the AVs should also help create a comfortable travelling experience in the car by avoiding AS. The situation where the driver does not need to prepare for taking over in level 3 automation also deserves attention.

To minimise the AS timely warnings are needed to help drivers be aware of vehicle capabilities and monitor the road at proper times and be aware of the environment (Carsten and Martens, 2019). For example, the AS could be generated when a level 3 automated following car brake promptly as the front car decelerates in a sudden. If the driver is aware of the closely car-following situation with timely warnings or notifications, the AS could be decreased because the driver has the anticipation of the sudden brake managed by the ADS. In this project, we would like to explore a system fitted in level 3 AVs providing cues to the driver to help the driver minimize the AS and not interrupt manipulating NDRTs.

2.2 Methodology

In this section, the guiding methodology of the experiment, "Wizard of Oz (WoZ)", is introduced. The concept of WoZ and its applications in the automated driving research domain, especially in the on-road experiments, are mentioned.

2.2.1 Wizard of Oz (WoZ)

WoZ is a common method used to test or iterate design prototypes since 1983 in human factors. In this method, the "wizard" is the experimenter who will be hidden from the participant and simulate the behaviour of a theoretical intelligent computer application (Kelley, 2018). The user will not be aware of the presence of the wizard and is led to believe that the prototype or the system is fully operational (Salber and Coutaz, 1993). The WoZ could help eliminate the limitations of automated systems, and it is less expensive (Habibovic et al., 2016).

2.2.2 WoZ Car

The WoZ is widely used in the automotive domain to evaluate various interfaces of vehicles (e.g., gesture-based, speech-based), with both driving simulators and real traffic environments (Habibovic et al., 2016). For example, Ruijten, Terken, and Chandramouli (2018) compared the user's trust of AVs with Conversational User Interface (CUI) and Graphical User Interface (GUI) with the driver simulator by experimenter playing the speech of the CUI.

Baltodano et al. (2015) developed a Real Road Autonomous Driving Simulator (RRADS) platform and first introduced the WoZ method to evaluate and test prototypes or iterate designing with on-road cars. The RRADS platform proposes an on-road, WoZ autonomous car simulation environment by modifying a commercial car. Two

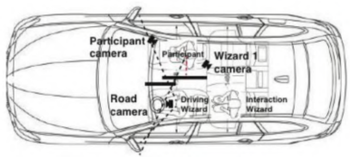

Figure 2.3: The layout of RRADS platform (Baltodano et al., 2015)

"wizards" are required in the experiment. One is the wizard driver, who will manoeuvre the car on the road. Another is the interaction "wizard", who is the experimenter activating the prototype and monitoring the experiment during the journey. The participant would sit in the front passenger's seat. As shown in Figure 2.3, a partition was designed and set between the driver's seat and the passenger's seat to prevent the participant from seeing the driver and ensure the participant could have front, side, and rear views of the environment. A steering wheel was also implemented in the participant area to give the participant the mindset of behaving as a driver rather than a passenger (Baltodano et al., 2015). The results of this experiment showed that the RRADS may be an effective way to evaluate prototypes and scenarios specific to open road human-autonomous vehicle interactions (Baltodano et al., 2015).

Another example of the real traffic experiment with the WoZ method is an on-road automated vehicle simulator developed by Karjanto et al. (2018). The partition in this on-road simulator is to separate the front and back cabin of the vehicle (shown in Figure 2.4). The participant would be in the back of the cabin, and a television connecting with a GoPro will show the front view. In the context of creating an autonomous vehicle, the seats in the back cabin could be in several variations mimicking different autonomous driving scenarios (Karjanto et al., 2018). Detjen, Pfleging, and Schneegass (2020) used a similar setup of the WoZ vehicle to explore the trust and acceptance of the AVs in the highly automated driving scenario.

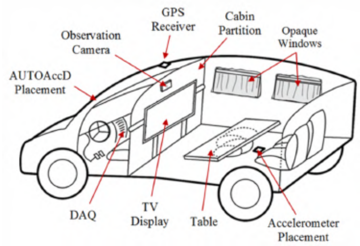

Figure 2.4: The layout of on-road automated vehicle simulator developed by Karjanto et al. (2018)

Methods | 3

In this chapter, the experiment methods are described. It includes the experiment design, the participants, the materials prepared and used in the experiment, the procedure and the measurements.

3.1 Experiment Design

The **within-subjects method** was used in the experiment. All participants experienced the scenario twice, i.e., travelling on the pre-selected route for two rounds. In one of the two rounds, the participant experienced the no-critical cue. To minimise the adaptation of the experiment set-up and the scenario, it is randomised for each participant in which round the prototype is engaged. The inherent cues were exposed to the participants, who are required to conduct the NDRT, in both rounds. Qualitative-oriented mixed methods were applied to the study. Both quantitative and qualitative data were collected through questionnaires, video recordings and semi-structured interviews.

3.2 Materials

In this section, all the materials prepared for the experiment are listed and explained, including the WoZ vehicle, which is the moving Mobility Lab of the Industrial Design Department, Technology University of Eindhoven. The experiment trials are on the real road. The pre-selected route is also introduced in this section.

3.2.1 WoZ vehicle

The experiment was conducted with the WoZ car (shown in Figure 3.1), adapted from a Renault Espace modified by Karjanto et al. (2018) in their previous study.

There were originally three rows in the cabin of the WoZ car. A partition separated the cabin apart, with one row in the front, where the wizard driver and the experimenter sat (same as the layout in Figure 2.4). The second row was removed, so the participant sat in the third row in the modified back cabin, which simulated the driving area, as shown in Figure 3.4. The Instrument Panel (IP), steering wheel and etc, were implemented in it. The interior of the

Figure 3.1: The WoZ vehicle used in the experiment

Figure 3.3: Interior of the WoZ car: back side

WoZ vehicle and the key elements are shown in Figure 3.3 and Figure 3.2 and listed below:

Figure 3.2: Interior of the WoZ car: front side

1. Television screen
   A 43" Sony television screen (KD-43XF7596) was mounted on the partition facing the back cabin. It mimicked the function of the windshield and showed the front view to the participant by connecting to the GoPro camera mounting on the windshield in the front cabin.
2. Cameras
   There were two cameras in the WoZ car. Camera 1 (shown in Figure 3.2) was a GoPro mounted on the windshield to record the road view and send it to the television screen by cable. Camera 2 (shown in Figure 3.3) was a motion camera from Nikkie set on the top of the partition facing the back cabin to record the participant.
3. Data acquisition system
   We adapted the system using the Arduino board with the accelerometer ADXL335 sensor and GPS Adafruit Ultimate GPS sensor to log the movement data, including speed, acceleration (sampled at 100Hz) and the location (sampled at 1 Hz) of the vehicle. All these data, together with the time-stamp of the non-critical cue, were stored in the SD card using Arduino Mega with Data Logging Shield and synchronized by the real-time clock.
4. Instrument Panel (IP)
   A IP is made of MDF board after laser cut and assembled in the car. Then we covered the instrument panel with fabric. A display showing the mode of the AV is placed in the centre of the IP. An emergency button is also set on the IP. When

Figure 3.4: The sitting area of the participant

the participant feels too sick to continue the experiment, they could press the button, and it will beep in the front cabin to inform the experimenter.

5. Mode display
   A mode display was installed in the centre of IP, which showed the automatic or manual mode of the WoZ car during the experiment.
6. Steering wheel
   A Logitech steering wheel was attached to the IP, acting as a puppet. While driving, the Logitech steering wheel would rotate simultaneously with the real steering wheel.
7. Pedals
   Pedals were also placed on the back side of the cabin. The pedals, together with the Logitech steering wheel, were not functional during the journey. They helped create the realism of a level 3 AV and reminded the participant of the driver's role.

3.2.2 Prototype

We used a 1027 disc vibration motor embedded in a wristband to create a vibrotactile prototype and implemented it in the experiment car to research the non-critical cue's influence on the driver's experience. The motor was connected to a button in the front cabin by wire. When the experimenter pressed the button, the wristband would vibrate, and the timestamp would be stored in the data acquisition system.

Choosing the vibration-based interface was related to the NDRT. We asked the participants to watch videos with a provided tablet as the NDRT. Therefore the prototype in the audio or visual modality, which involve listening and watching respectively, would conflict with the NDRT. Meanwhile, the repetition auditory warning sound is widely used in TORs (Sanghavi, Zhang, and Jeon, 2020; Edworthy, Loxley, and Dennis, 1991; Hellier, Edworthy, and Dennis, 1993). We hoped to make a distinction between the non-critical cue and TORs, and this is another reason why we created the vibrotactile prototype.

It should be noted that we do not evaluate the design quality of the prototype. Our main focus is to explore the concept of the non-critical cue behind the prototype.

3.2.3 Experiment Route

We only conducted the experiment during the daytime on predefined routes, avoiding adverse weather conditions (e.g., fog, snow, icy road).

The experiment route included highways (N2 and A50) and motorways (Tilburgseweg and John F Kennedylaan) in the Eindhoven area as shown in Figure 3.5.

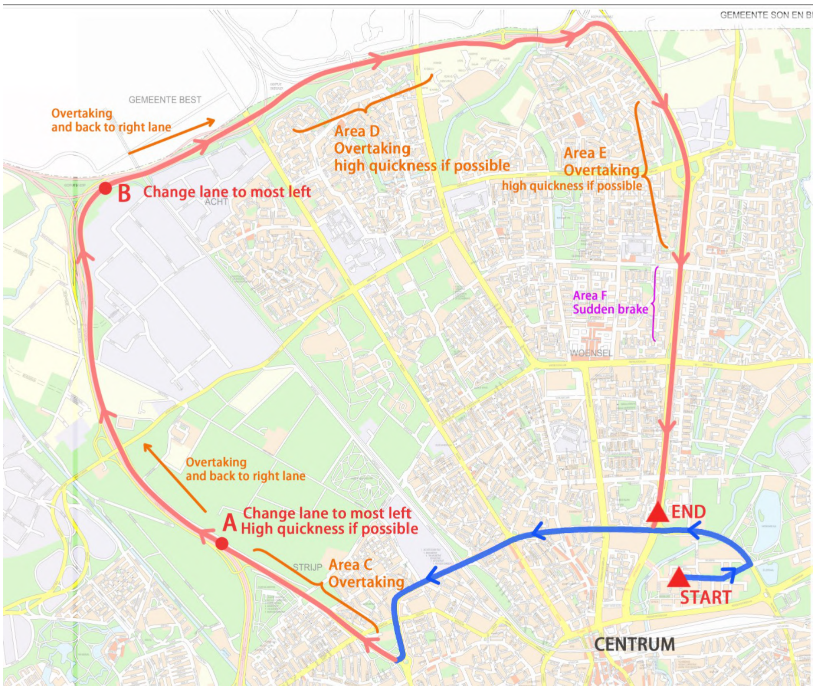

Figure 3.5: Experiment Route

The route consisted of two segments. In the first segment shown in blue in Figure 3.5, the WoZ car started from the parking lot under the Atlas building in the TU/e campus to the Tilburgseweg via Fellenoord, Boschdijk and Beukenlaan. During this period, the participants could get familiar with the experiment setup and the environment. From the Tilburgseweg on, the experiment started officially. Then the car would drive from Tilburgseweg via highway N2 and A50 to Kennedylaan, and this segment was in red in Figure 3.5. The experiment ended at the cross of Kennedylaan and Onze Lieve Vrouwestraat (marked as "END" in Figure 3.5). Then the WoZ car would drive back to the blue segment and enter the next round or head back to the start point if both rounds were finished.

The first segment took around 10 to 15 minutes, according to the traffic. From the Tilburgseweg, the official experiment phase lasted 12 to 15 minutes. One round of the journey took about 30 minutes, and the whole experiment would take up to 1.5 hours, including two rounds on the road and an interview.

3.2.4 Driving Manoeuvres

In case the events in the environment and the vehicle movement are too subtle for the participant to detect, we set some special manoeuvres along the route. The area where the manoeuvres would be performed is shown in Figure 3.5. All the manoeuvres should be conducted under safe conditions. If the traffic condition were too complex for the driver to perform the manoeuvres safely, the manoeuvres would be abandoned. The requirements of each manoeuvre are explained below.

Lane changing

To keep the defensive driving style, the driver should conduct the low quickness lane changing shown in the right of Figure 3.6 (Bellem et al., 2016). The wizard driver was required to change the lane at spots A, and B and areas C, D and E (Figure 3.5) with overtaking. However, at spot A and areas D, E shown in, the wizard driver should perform high quickness lane changing shown in the left in Figure 3.6 if possible. At spots A and B, the driver is supposed to change continuously to the most left lane.

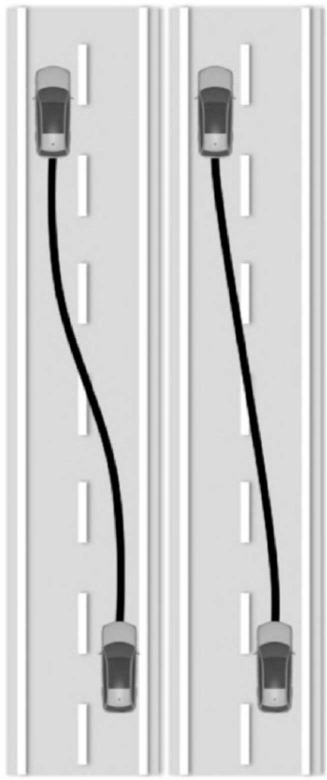

Figure 3.6: Lane Changing Quickness. Left: high quickness; right: low quickness (Bellem et al., 2016)

Overtake

The driver should conduct at least three overtaking manoeuvres during the whole journey. In the area after spots A, B and areas C, D, and E shown in Figure 3.5, overtaking should be conducted. Because the manoeuvre overtaking included lane-changing behaviour, the requirement for overtaking is the same as for lane-changing. In areas D and E, the manoeuvre should be conducted with high quickness if possible.

Sudden Deceleration

The sudden deceleration should have been done only when there was no car following closely. Around the area shown in Figure 3.5, the driver should reduce the speed suddenly from 75km/h to about 65 to 70 km/h at the speed camera there if possible. Due to the road sign showing the change in the speed limit around the area and the speed camera, we assumed that it would not be too confusing for the participant to understand the sudden deceleration if the signs were noticed.

3.2.5 Wizard Driver Training

A study conducted by Yusof et al. (2016) showed that the human driver preferred the defensive and the LRT and defensive driving style because they gave participants senses of comfort, pleasantness and safety. We would like to make sure the wizard driver drive with LRT or defensive driving style. Meanwhile, it is essential for the objectivity and reliability of the experiment that the wizard driver can reproduce the consistent driving style for different participants (Müller, Weinbeer, and Bengler, 2019).

So we conducted the driver training session with every wizard driver to ensure (1) they conducted consistent defensive or LRT driving style for every participant; (2) the driving styles of the WoZ car in the trials conducted by different wizard drivers were consistent. In the sessions, an Automatic Acceleration and Data controller (AUTOAccD) developed by Karjanto et al. (2017) was used to assist the wizard drivers in keeping LRT and defensive style. And the wizard drivers would practice performing the scripted manoeuvres mentioned above with the WoZ car in the training session. Instruction will be provided to the driver during the training, which is in Appendix A.

3.3 Participants Requirements

We required the participants to be older than 18 years old and have valid European driver's licenses. People very susceptible to motion sickness were excluded from participating. We screened people with the Motion sickness susceptibility questionnaire short-form (MSSQ-Short) online from Golding (2006), shown in Appendix B.1. Based on the questionnaire result, people who scored below 50% in the percentage conversion statistics of the research result from Golding (2006) were selected to participate in the experiment.

61 people responded to the motion sickness questionnaire, and 42 of them were qualified. In the end, 24 participants took part in the experiment.

3.4 Procedure

Once the participant signed up for the study, the motion sickness susceptibility questionnaire (created by *Qualtrics XM // the leading experience management software* (2022)) would be sent to the participant via email. The participant who passed the motion sickness test would receive the pre-experiment questionnaire and the email with the link to the time slot. By choosing the time slot, the participant confirmed the participation.

When the participants came to the experiment site, the experimenter greeted them on the ground floor in Atlas, TU/e. The experimenter would then explain the study, and ask the participants to sign the consent form, which explained the data being collected and the potential risks of this experiment. It would be emphasized verbally to the participants that they could always stop the experiment without reasons and consequences.

Then they would be led to approach the WoZ vehicle from the back. To easier facilitate the imagination of the experiment car that could drive automatically, the presence of the wizard driver should be obfuscated and hidden from the participants until the end of the experiment. Therefore, the wizard driver should already be sitting inside and not visible through the car's windows.

After the participants sat in the car, the experimental setup was briefly introduced. Firstly, we informed the users that the automated driving function could only work in a particular area. That is to say, the car would drive manually in the first ten minutes, and meanwhile, the participant could get used to the environment and the experiment setup. Speech audio would be played to the participant when the experiment officially started. We also told the participants that they could not control the vehicle with the steering wheel and pedal, but this was to facilitate the automated driving experience.

The experimenter went to sit on the front passenger seat and was hidden by the partition. After confirming all the equipment works properly, the wizard driver would start driving. When the car reached the Tilburgseweg (the red segments in Figure 3.5), the experiment started officially.

During the trial, the participant was required to watch the video on the tablet. When each round ended, the participant would fill in the post-experiment questionnaire in the car. After two rounds, the wizard driver would drive the car back to the starting spot. Then we conducted the interview in the Atlas.

Each experiment lasted about one hour and a half, including 15 minutes of introduction, 50 to 55 minutes of the on-road experiment and questionnaire, and 20 minutes of the interview.

3.5 Measurements

In this experiment, we collected both qualitative and quantitative data. All the data we measured was shown in the Figure 3.7.

Introductory questionnaire

Prior to the experiment, the participants were asked to fill in an introductory questionnaire, which collected demographic information, and the propensity to trust the automated vehicles. The questionnaire was created by *Qualtrics XM // the leading experience management software* (2022). The participants would receive the introductory questionnaire after they passed the motion sickness screening via email and were requested to fill it out before the experiment.

The detailed contents are listed in Appendix B.2 and explained below:

Demographic data In this section, the age, gender and nationality of the participants were collected. Meanwhile, the driving condition of the driver, such as the driving frequency, was collected as well.

Propensity to trust the automated vehicles Six 5-point Likert-scale questions, adapted from Merritt et al. (2013), were used to measure the participant's tendency to trust automated vehicles in general.

Post-experiment questionnaire

After the participants had experienced each round, they were asked to fill in the post-experiment questionnaire. The post-questionnaire is a modified version of the User Experience Questionnaire (UEQ) (Laugwitz, Held, and Schrepp, 2008). Four scales were assessed to evaluate the user experience of the round with and without the non-critical cue provided. The four scales are attractiveness (likability and overall impression), perspicuity (how easy to learn

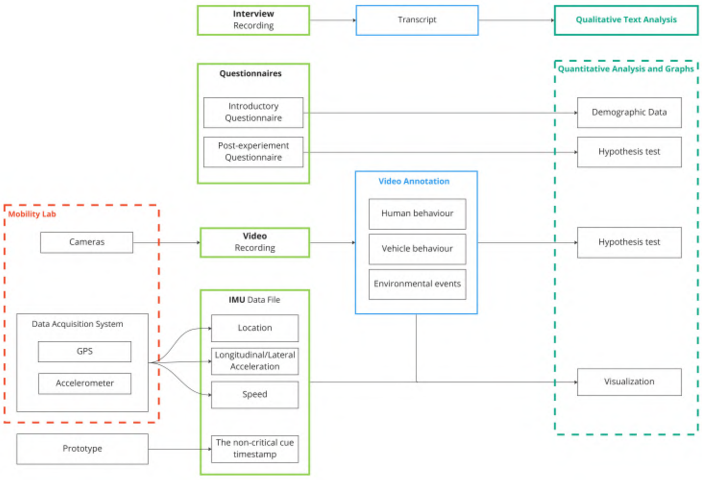

Figure 3.7: Data Collected in the Experiment

and understand), efficiency (whether interacting without unnecessary effort), and dependability (whether feeling in control of the interaction). The questionnaire contains 18 items, 6 of which belong to the attractiveness scale. The other scales include four items each. Each item is a seven-staged scale with the form of a semantic differential. The order of the items is randomized. Half of the items start with the negative term, and the other half start with the positive (Laugwitz, Held, and Schrepp, 2008). The questionnaire is listed in Appendix B.3.

Vehicle IMU Data File

The vehicle Inertial Measurement Unit (IMU) data, including the speed, longitudinal and lateral acceleration, and location, was collected by the data acquisition system in the mobility lab, including the Adafruit Ultimate GPS and Accelerometer ADXL335. All the data were synchronized by the real-time clock and stored. When the experimenter triggered the cue by pressing the button, the timestamps were also stored in the same file.

Video Recording

During the experiment, the front view was recorded by a GoPro from the windshield. From this view, we could see the vehicle's behaviour, including lane change and braking, and the events in the environment, such as trucks and road construction. Participants' behaviour and sound inside the car were recorded as well.

Interview

After the participant completed two rounds of the experiment, we conducted a semi-structured interview with the participant. The interview contains ten questions. We required the participants to compare two rounds when answering the questions. The interview questions could be categorised into five main domains. These are general feelings about the journey, automation surprise, distraction, opinions on the prototype, and the experiment set-up.

We didn't directly ask the participants about the automation surprise. However, we discussed the three qualities to measure the automation surprise with the participants. The three qualities are observability, predictability and timeliness, defined by Carsten and Martens (2019). The interview questions are listed in Appendix C.1.

Data Analysis 4

In this chapter, we will elaborate on the analysis of the data we measured as mentioned in section 3.5.

4.1 Questionnaire

4.1.1 Introductory Questionnaire

As was mentioned in the previous section 3.3, 24 participants took part in the experiment. Experiments of 20 participants out of the 24 succeeded and received valid data. All the following analyses were based on the 20 participants (16 males and 4 females; median age = 27, SDage = 6.24, range =23 to 54). All the participants had driver's licenses valid in Europe. 40% of the participants drove a few times per week during the three months before the experiment. Only 2 participants had never driven during that period. 75% of the participants had over three years of driving experience. More than half of the participants had heard about the Advanced Driver Assistant System (ADAS) but never tried before, as shown in Figure 4.1.

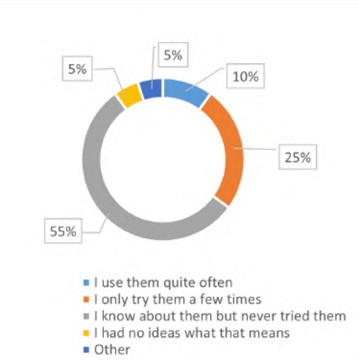

Figure 4.1: Knowledge about ADAS

The propensity to trust the automated vehicle questionnaire includes six 5-point Likert-scale questions. The mean of the answer is 3.477 (max = 4.83, min = 2.00).

4.1.2 Post-experiment Questionnaire

As mentioned in section 3.5, the post-experiment questionnaire includes 18 semantic differential items belonging to four scales. We analyzed the data from the post-experiment questionnaire with the analysis tool provided by Hinderks, Schrepp, and Thomaschewski (2018), who is also the provider of the UEQ. The scale means corresponding to 5% confidence intervals of two rounds are shown in Table 4.1, Table 4.2 and Figure 4.2.

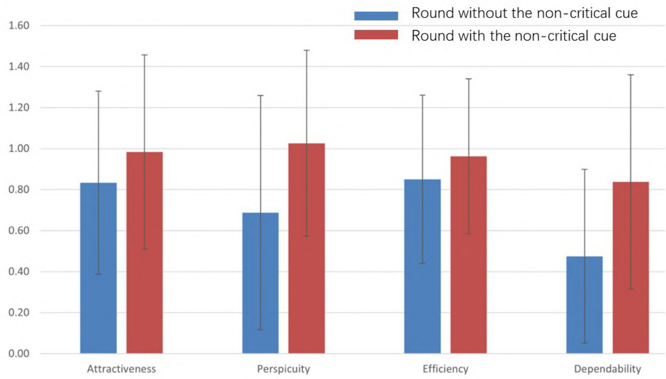

Figure 4.2: Scale means corresponding to 5% confidence intervals.

Table 4.1: Scale means corresponding to 5% confidence intervals: the round without the cue

	Mean	STD	Confidence	Confidence Interval	
Attractiveness	0.83	1.02	0.45	0.39	1.28
Perspicuity	0.69	1.31	0.57	0.12	1.26
Efficiency	0.85	0.94	0.41	0.434	1.26
Dependability	0.48	0.97	0.42	0.05	0.90

Table 4.2: Scale means corresponding to 5% confidence intervals: the round with the cue

	Mean	STD	Confidence	Confidence Interval	
Attractiveness	0.98	1.08	0.47	0.51	1.46
Perspicuity	1.03	1.04	0.45	0.57	1.48
Efficiency	0.96	0.86	0.38	0.58	1.34
Dependability	0.84	1.19	0.52	0.31	1.36

A sample t-test was conducted, and the result is listed in Table 4.3

Table 4.3: The sample t-test result of UEQ (alpha=0.05)

Scale	P-value
Attractiveness	0.6542
Perspicuity	0.3709
Efficiency	0.6951
Dependability	0.2979

Statistically speaking, there is no significant difference between the user experience in the round with and without the non-critical cue. However, from the Figure 4.2, the mean score of the four scales in the round with the non-critical cue are higher than the other round.

4.2 Video Data

The video recording analysis includes two main procedures. In the first procedure, the videos were coded by two coders in the video analysis software, and the Inter-rater Reliability (IRR) was calculated. In the second procedure, the video annotation results from procedure one were analyzed.

Subject	Behaviour	Key	Type of the Behaviour	Criteria: Start Point	Criteria: Stop Point	Comments	View
Human	Look up	U	State	The participant's eyes moved away from the tablet.	The participant's eyes moved back to the tablet.	• The "look up" behaviour should be less than three seconds. • The "look around" behaviour should be more than three seconds. • The gap between two behaviours should be more than 0.5 seconds.	Inside view
	Look around	A	State				
	Other behaviours	O	State	The behaviour started.	The behaviour ended.	• smile, yawn or other meaningful behaviours from the coder's opinion	Inside view
Vehicle	Lane changing	L	State	The road centerline started to move.	The vehicle started to go straightly.	-	Front view
	Sudden Deceleration	D	State	The speed on the dashboard started and stoppedchanging		• This vehicle's behavious is the one specifically happened at the speed camera on the Kennedylaan.	Front view
Environment	Turn	S	State	The road branch started.	The road branch ended.	-	Front view
	Truck	T	State	The end of the truck reaches the frame of the front camera view.	The front of the truck reaches the frame of the front camera view.	• Only the trucks driving on the next lane would be counted.	Front view
	Speed camera	M	Point	The speed camera on the Kennedylaan reaches the frame of the front camera view.		• The coordinate of the speed camera is 51 28.3365 N,5 29.2588 E	Front view
	Construction	C	State	The car was three-second away from the construction road sign	The road centerline obstructed by dashboard in the camera front view.		Front view

Figure 4.3: Video coding system

4.2.1 Procedure 1: Video Annotation

Before coding the videos, we pre-processed all video clips. The continuity of the videos was checked. Then we deleted the parts before, between and after each round of each video and separated one video into two according to different rounds. As a result, for one participant, there were two videos of the front view and two videos of the inside cabin view, paired according to rounds. In total, 40 pairs of videos were generated. Next, we used the open-source event-logging software BORIS for video annotation(Friard and Gamba, 2016).

Two coders were involved in the coding process. One was from this study, and another was an external researcher. Coder1 has already read through all the video data in the pre-processing procedure. Then the ethograms and coding scheme was generated inductively by coder 1. The coding scheme and the coding criteria are shown in Figure 4.3. The coding system includes three subjects and behaviours under each subject, which are the human (look up-/around, others), the vehicle (lane changing, sudden deceleration) and the environment (road construction, truck passing by, turning zone, speed camera). It is worth mentioning the "behaviour" under the subject "Environment" represents the events that happened in the environment. And the behaviours under the "Vehicle" subject were the scripted driving manoeuvres.

Then coder1 explained the coding system to the second coder and discussed it with coder2. Then coder1 coded all the videos. Inter-rater Reliability (IRR) was performed on ten pairs of videos (25% of the database) belonging to the experiments with five participants (P02, P17, P18, P24, P29) by the second coder (O'Connor and Joffe, 2020). The videos rated by coder2 were randomly selected by participant number. After the two coders completed the video coding, IRR between coders was calculated using the BORIS with Cohen's kappa statistic in the time unit of 1.000 second (Landis and

Koch, 1977). Then the iteration of the video coding was conducted by coder1 (Lazar, Feng, and Hochheiser, 2017). During the iteration, coder1 checked the video annotation files of two coders and revised some coder1's annotations. The video coding data from coder2 was not edited.

The thresholds of the Kappa and its relation to the strength of agreement are shown in Table 4.4. According to Landis and Koch (1977), a kappa between 0.61 to 0.80 is substantial and above 0.80 is an almost perfect agreement.

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Table 4.4: kappa Statistics and the Related Strength of Agreement(Landis and Koch, 1977)

The final result of IRR iteration is shown in Figure 4.4. The overall IRR before iteration was 0.843, and after iteration was 0.864, both of which were almost perfect agreement according to Landis and Koch (1977). The IRR under each subject was also calculated, and the scores were substantial generally. However, a few scores were lower than 0.60 or decreased after the iteration. These scores were marked red in the Figure 4.4 and explained below.

P02-without The IRR under the environment category decreased after iteration. It was because coder1 and coder2 annotated the starting point of the second turn on the route mistakenly to the same wrong spot. Then when iterating, coder1 corrected the mistake; however, the file from coder2 stayed the same, which led to the decrement in the IRR under the category of environment and overall score.

P02-with The IRR of the vehicle category decreased because neither the first nor second coder annotated the sudden deceleration event. Coder1 added the sudden deceleration in iteration.

P17-without We set the criteria that the two consecutive behaviours (e.g. the participant looking up twice in a row) would be annotated as one if the gap between the two behaviours was

		Human		Environment		Vehicle		Overall	
		Original	Iterative	Original	Iterative	Original	Iterative	Original	Iterative
P02	Round without the prototype	0.844	-	0.925	0.677	0.738	0.781	0.874	0.851
	Round with the prototype	0.788	0.816	0.699	0.983	0.763	0.751	0.849	0.885
P17	Round without the prototype	0.590	0.652	0.890	-	0.711	-	0.748	0.774
	Round with the prototype	0.582	0.651	0.994	-	0.766	0.774	0.767	0.798
P18	Round without the prototype	0.745	0.770	0.890	0.971	0.671	0.742	0.830	0.854
	Round with the prototype	0.676	0.798	0.982	-	0.566	0.721	0.781	0.870
P24	Round without the prototype	0.785	0.822	0.961	0.978	0.758	0.782	0.875	0.896
	Round with the prototype	0.853	0.856	0.989	-	0.820	-	0.913	0.914
P29	Round without the prototype	0.792	0.797	0.989	-	0.843	0.853	0.895	0.902
	Round with the prototype	0.856	0.859	0.981	-	0.881	0.890	0.899	0.900
	Average	0.751	0.787	0.930	0.968	0.752	0.784	0.843	0.864

Figure 4.4: Inter-rater Reliability of the Video Coding

less than 0.5 seconds. However, if the gap is too close to the 0.5s, it would be hard to recognise accurately, even with BORIS. Due to the reason above, in this case, the original IRR was lower than 0.60 because some consecutive "look up" behaviours were coded as one by coder1 but two by coder2.

P17-with The original IRR score is lower than 0.6 because of the annotation of behaviour named "Other behaviour". The criterion for this behaviour was that the coders could annotate the meaningful behaviours in their own opinion, including but not limited to smiling and yawning. This is subjective and could cause some deviation between the two coders' annotations. If we do not count the "other behaviour" when calculating IRR, the score will rise to 0.748 after iteration.

P18-with Coder1 assigned several "Lane Changing" behaviours, which should have belonged to the "Human" subject, wrongly to the "Environment". The mistake was corrected by coder1, and the IRR of both "Environment" and "Vehicle" increased.

4.2.2 Procedure 2: Video Annotation Data Analysis

After the video coding, the video annotations of every participant were exported as CSV files, including the start/stop time and duration of all the behaviours. This video annotation data were analysed together with the IMU data. When the experiment officially started, the speaker would give a beep sound, which decided the starting point of video annotation. Meanwhile, the experimenter in the front seat would press the button to timestamp the moment in the IMU data. The two data sets were synchronized manually by matching these two timestamps and combining the two CSV files into one with the same timeline.

Then the data set was visualized by another researcher involved in the study. The graphs of each participant give an overview of the driving profile, the events as well as the participant's behaviours of each round. One visualization example of P14 was shown in Figure 4.5. By comparing the two rounds, we can see that the general driving profile in the two rounds was similar. However, in the round with the non-critical cue, there was no huge truck passing by, whereas there were several in the round without the cue. Regarding the participant's behaviour, P14 would look up or around almost every time when he was cued, or when there was a lane change. He also kept looking around more in the round without the cue, compared with the other round.

Figure 4.5: Visualization of video data

In addition, we statistically compared the difference in look-up number, and look-around duration means between the round with and without the non-critical cue. The t-test was conducted by STATA 16 (StataCorp, 2019). The normality was checked, and the p-value is listed in Table 4.5. There is no statistically significant difference between the round with and without the non-critical cue for the look-up number, and for the look-around duration means.

	P-value
Look-up number	0.7397
Look-around duration mean	0.8761

Table 4.5: The t-test result of video annotation data (alpha=0.05)

4.3 Interview

Firstly, We recorded the interviews and transcribed them into texts. Then the thematic qualitative text analysis method was applied to interview analysis in *MAXQDA* (2022) (Kuckartz, 2014). To ensure coding quality, we implemented a cooperative coding method named consensual coding. In this project, two coders from the research team were involved in the interview coding process, which is adapted from Consensual Qualitative Research (CQR) developed and updated by Hill, Thompson, and Williams (1997) and Hill et al. (2005). We followed five steps when executing the consensual coding method in the thematic analysis.

For the first step, the interview recordings were transcribed by one researcher who is involved in the study (coder1). Then coder1 formed six main categories corresponding to the questions asked in the interview: general, observability, predictability, timely, automation surprise, distraction and setup. The starting point also considered grouping into two groups, with or without cue. Coder1 coded the interview transcriptions with these main categories. During the coding by coder1, inductive themes emerged (Lee et al., 2019).

Coder1 then generated a coding scheme and shared it with a second coder (coder2). Coder2 individually coded the interview data with the coding scheme from the first coder, during which new categories and sub-categories emerged.

Next in the third stage, the two coders read through the interview transcription together and discussed the difference and similarities of their coding (Hill et al., 2005). Then they fine-tuned the definition of the categories and regrouped them with a higher level of abstraction (Kuckartz, 2014). Some categories were merged, divided or merged (Hill, Thompson, and Williams, 1997).

The result of this stage is that the two coders found consensual and generated the first version of the interview coding scheme. In the fourth stage, we included an external researcher as an auditor, going through the categories and finalizing the coding scheme together with coder1 and coder2. The full version of the final coding scheme is included in the appendix C.2. (Appendix Coding Scheme: categories and sub-categories are well described and explained with examples)

The last step is charting the results of the fourth step (Hill, Thompson, and Williams, 1997). Relations between the categories were depicted visually in a coding map (Figure 4.6) (Hill et al., 2005). The findings and results from the interview coding and the coding map will be explained in the next chapter.

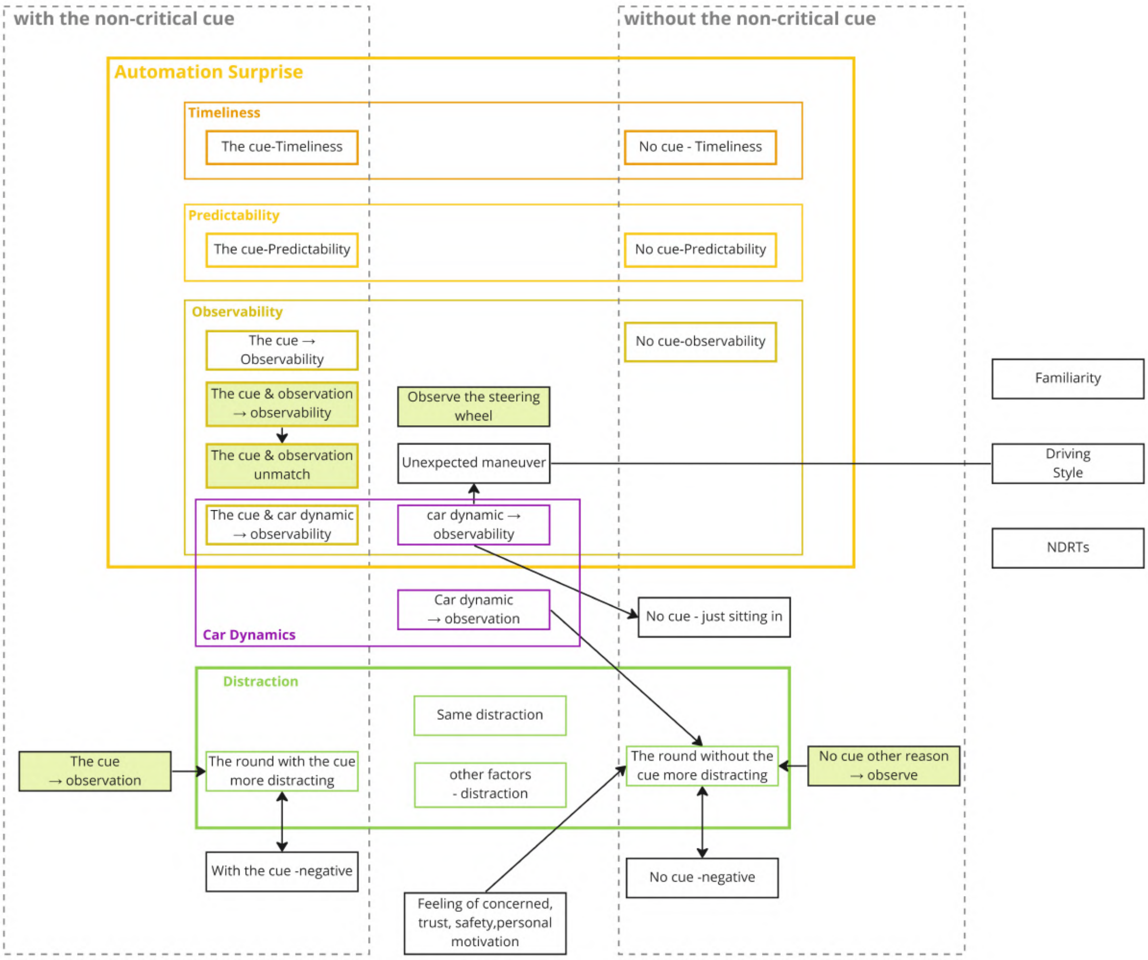

Figure 4.6: Coding map: relations in the interview coding scheme

5 Results and Findings

In this chapter, we delve into what we obtained from data analysis and provide an interpretation of the interview qualitative analysis results and video recordings. We also discuss the main findings and how they relate to the research questions and objectives outlined in the introduction. It should be noted that "the driver" mentioned in this chapter below refers to the participants as well as the user of the AV.

5.1 Cue and Automation surprise

Automation surprise, as explained in the introduction, is the negative feeling such as shock or surprise that the driver experiences in a conditionally AV when conducting NDRTs. As was explained in the previous chapter, we developed and tested the non-critical cue to investigate if it could contribute to reducing AS without interrupting the driver conducting NDRTs in the conditionally automated vehicle. Meanwhile, we also explore the influence of inherent cues, such as the car dynamics and the steering wheel turning direction. And we adapted three qualities, observability, predictability and timeliness of HMI to measure if they could minimise automation surprise (Carsten and Martens, 2019). This section explains the influence of the non-critical cue and the inherent cues on automation surprise from these three perspectives.

Cue and Timeliness

As stated in the previous chapter, a HMI in the AV with good timeliness could provide information to the driver early enough, so that the driver could take proper action (Carsten and Martens, 2019). This means the non-critical cue should hint to the driver sometime before the movements of the car or the events in the environment to reduce automation surprise effectively.

According to the interview, some participants felt the non-critical cue coming before the movement of the car was able to make them "aware something would happen in advance and give the driver longer time to react"(P04). Also, the participants were able to prepare themselves or take some actions, for example, look up to check when receiving the cue:

> "With the prototype, the signal will come maybe before, just before the car will do some action, or something

> will happen for the vehicle. It will drag your focus on the vehicle when you are watching the video, it makes you more focused on the status of the vehicle so that you will get the changes of the status of the vehicle way faster, than with no prototype." - P11

In the round without the non-critical cue, there was no warning before the car's action, which means the user could only be interrupted by the car movement itself and then "try to understand the intention" (P30) after the movement already happens. For example, P25 expressed that he got no time to react in the round without the non-critical cue because there was no warning.

Comparing the participants' experience in the rounds with and without the non-critical cue, the non-critical cue could potentially help reduce automation surprise from the timeliness perspective, because it notices the driver sometime before the car movement, and gives the driver a chance to take some action.

Cue and Observability

Here, observability means whether the participants could detect and understand the car's intention and movement. With good observability, they could be aware that the car will make some movement with its intention. The interview results revealed three ways that how the non-critical cue could potentially help reduce AS in observability.

Firstly, the non-critical cue could directly contribute to the increment of observability. From the interview, more than half of the participants described that they could straightly know "something was going to happen" (P14), "the following steps" (P23) or understand the intention of the car with the non-critical cue. The participants did not express evident thinking processes or additional actions taken when achieving this, according to the interview. P01, for example, saw the non-critical cue as a signal and could link the signal directly to "something about to happen", such as "changes on the road" or "changing lanes" by the car. P04 described the non-critical cue as the AV telling him the decision it made, and he could understand the reason behind it, making him "not surprised of it".

In addition, the non-critical cue is possible to reduce automation surprise by triggering the driver to observe more to increase observability. According to the interview, some participants mentioned when they received the cue, they would "move the attention out of the video" (P04) and "look in the environment" (P29). Then they could understand why the car conducted the manoeuvre:

> "In part two, with the help of the vibration, I know what's going on. And when it vibrates, I will observe what's going on around me, then it helps me to have a better understanding, and I can predict what the car is gonna do." - P23

However, there are also a few circumstances when the observation of the surroundings from the participants may not align with their understanding of the non-critical cue. The cue cannot contribute to increasing observability in this situation. For example, P34 was curious about what the car would do because he saw nothing when he looked up from watching the video on the tablet due to the cue. Similarly, P26 felt there was "not much to pay attention to" when he received the cue and looked around. In the video recordings, P26 received the cue due to lane changing. Then he turned around to look through the rear window in the round with the non-critical cue (at around 352.87 seconds of the recording), which showed he did not understand the non-critical indicated the lane change manoeuvre and was wondering what was going on when he received the non-critical cue.

Besides observing the vehicle and the environment when receiving the cue, the user could also gain more observability by combining the non-critical cue with the car dynamic they sensed. In this case, the non-critical cue and sensed car dynamic work mutually to reduce AS. In other words, it is probable that the cue coming before the car movement prepared the driver for the sensation of car dynamic, and vice versa; the sensation proves the driver's understanding of the cue:

> "... know what the information is, very gently. It's just a lane change, and it (the non-critical cue) is a tiny tap. And then maybe if it's in combination with the car going around a corner or something, then you know what's happening without having to look up. "-P09

Besides working together with the non-critical cue, the car dynamic could increase the observability to reduce automation surprise individually. The driver could sense and understand the car's movement by the sensation. In this condition, when the driver senses the gentle car movement, the driver could perceive the car's behaviour and the environment, and then understand the intention of the automated vehicle:

> "I felt I could sense the situation, just from feeling how the car was moving. And that was just as good as the vibration for me. Yeah, you could feel the car accelerating, decelerating, moving left, right." -P21

However, the car had some aggressive scripted manoeuvres during the journey. The participants perceived these movements as "harsh braking" (P01), "sudden slowdown" (P11), "sharp lane change" (P17), etc. Under this condition, when the car moves aggressively, the car dynamic could be very harsh rather than a "tiny tap" to the driver. It could be sudden, unexpected, and may cause disturbance to the driver conducting NDRTs or other negative feelings:

> " ... sometimes the vehicles will slow down suddenly, without any information, so it will make me nervous. Because maybe it will cause the accident if there's a sudden slowdown for the vehicle. " - P11

Car dynamics' performance in timeliness is poor, as the inherent cue in the AVs playing a role on AS. Because the driver could only sense it at the same time when the car starts moving. Especially, when car dynamics are due to aggressive properties, it would perform terribly in timeliness and can no longer be able to function as a non-critical cue.

Correspondingly, when the car dynamic is too subtle for the participants to notice, it cannot function as well as the non-critical cue either. According to the interview, when there was no non-critical cue, some participants were not aware of being in the AV because they did not notice any signals from their surroundings. They felt "sitting on the train" (P34) or "more at home" (P25) and were very involved in the NDRTs. P26 expressed that the experience in the no non-critical cue round was the same as sitting in the back seat of a manually driving car as a passenger. Similarly, P04 felt he was in a taxi in the round without the non-critical cue. These examples demonstrate that when the AV drives smoothly, and the car dynamic is too subtle to be noticed, the participant would perceive it as being in another common mobile scenario, such as train commuting, in which they are often out of the loop, especially when focusing on NDRTs. One of the results could be the driver feels overwhelmed or surprised when switching the attention back to the surroundings after long-time immersing in the NDRTs:

> " The second part, because there is no vibration (non-critical cue), then I will watch the video, watch a little longer than the first part, ..., but when I look up I find that it is completely different (environment), yes I do not know where it took me, ..., a little overwhelmed kind of feeling." -P06

Overall, according to the interview, only a proper level (not aggressive or too subtle) of car dynamic itself could function similarly as a non-critical cue in the automated vehicle. However, the car dynamics are not ideal from the perspective of timeliness to reduce automation surprise.

Similar to the car dynamic, the turning direction of the steering wheel could also potentially work as a non-critical cue to help reduce automation surprise from the observability perspective. As mentioned before, in the conditionally automated driving scenario, the steering wheel would rotate by itself when the vehicle heading in a different direction. In the interview, three participants (P04, P11 and P23) indicated that the steering wheel functioned as a guide for them to grasp the vehicle's purpose. The observability could be improved by the participants observing the steering wheel alone or in conjunction with the environment. For example, P23 considered the steering wheel a good sign, because he could know if the car was turning left or right by it. P11 combined the non-critical cue and the steering wheel to distinguish if the car would change the lane or slow down because the former manoeuvre needed steering. Besides showing the vehicle's intention, the steering wheel is also useful to discriminate if the AV is in automated mode or not. Because the steering wheel would only turn by itself when the automated on mode within its ODD.

In contrast to the round with the non-critical cue, however, the round without the non-critical cue is more likely to make the driver experience the automation surprise. This is almost certainly due to poor observability when the non-critical cue was not provided. In the interview, when talking about the round without the non-critical cue, more than half of the participants felt the absence of information from the surroundings. For example, P1 felt "there was no information coming and not receiving any information" and "less in the loop". P18 described that he "did not have a proper view of what's happening" in the round without the non-critical cue. The driver may not be aware of the AV's intentions and status during the journey without the non-critical cue:

> "...so compare with the part two (without the non-critical cue), when there's no prototype means that there's no way for me to know the future, and the intent of the vehicle." - p11

Overall, from the aspect of observability, the non-critical cue potentially has a positive effect on reducing automation surprise in the conditionally AVs. The car dynamics as the inherent cue in conditional AVs could help reduce AS with limited conditions. The turning direction of the steering wheel could be a good sign of the AV's intentions (driving direction, manoeuvres relating to steering) and status (automated mode on or off).

Cue and Predictability

Regarding predictability, it means that the driver could gain enough observability and be able to plan his/her actions accordingly with the help of the non-critical cue (Carsten and Martens, 2019).

The non-critical cue could directly enhance predictability and reduce automation surprise. For example, P18 could adjust his behaviour according to the cue:

> "when it wants to do a quick manoeuvre, I could already be prepared for the quick manoeuvre, with the prototype." - P18

In contrast, in the round without the non-critical cue, the participants usually could not adjust their behaviour or self-prepare. This is mainly because, as mentioned in the previous section, the participants do not gain sufficient observability. Then it results in much less predictability in the round without the non-critical cue. The driver could "just go with the flow" (P34) or "just trust" (P04) because there was no way for them to predict.

5.2 Cue and Distraction

In the context of this research, the driver would conduct the NDRTs during the journey. The NDRTs are their primary task, which could be interfered with car movement and environmental events. When this happened, we would say the driver was distracted from conducting NDRTs. Many participants described in the interview that when receiving the non-critical cue, they would look up and check the car and the environment, as indicated in the previous section. For example, P34 felt the non-critical was telling him something was about to happen, and then he just immediately looked up. Moreover, this interference to the NDRTs by the non-critical cue could lead to contrasting feelings and experiences.

More than half of the participants felt the non-critical cue distracts them from conducting NDRTs. Many of them explicitly expressed that they "got more distracted" (P12) in the round with the non-critical cue but did not feel distracted and could focus on watching the video in the other round:

> "Because I was distracted by the vibration mostly. (...) In Part one, without any vibration, and any interaction between the car, I was more focused on the task." - P31

Some participants also described their feelings which shows how the non-critical cue mentally distracted them. For example, P04 said he was bothered by the cue. The non-critical cue may even put some mental pressure on the participants; as P24 said, the cue "forced him to observe more during the driving".

These participants felt the non-critical cue was distracting because the cue constantly triggered them to look up. Another reason these participants perceive the non-critical cue as a distraction is that they could not figure out the exact meaning of the non-critical cue instantly. They would think about the meaning of the non-critical cue when it appeared. This cognition process would distract them from conducting the NDRTs:

> " Sometimes it (the vehicle) would turn left or right. Sometimes it tells me some different situation, like the car passing by, so I am not sure what kinda info from the cue, so I need to combine (...) the evidence together, and know what will happen. That interrupts me." -P04

This could be a limitation of the implicit HMI, and we will discuss it in the following chapter.

However, it is essential to note that the non-critical cue is more distracting does not necessarily mean that the participants would perceive it negatively. For instance, P31 acknowledged that the round without the non-critical cue was more distracting yet felt more at ease. Additionally, P24 described the non-critical cue as "nice regarding safety", although he would switch his attention to the road whenever a non-critical cue was present.

On the contrary, close to half of the participants felt the round without the non-critical cue was more distracting than the other round. When there is no non-critical cue, they might feel more concerned and need to pay more attention to the road and the car. As a result, they would check what is going on actively and observe more, rather than focus on the NDRTs. For example, during the interview, P11 expressed that he "cannot really focus on the tablet" in the round without the non-critical cue. Because he was worried about the car and had to switch his attention back to the surroundings from time to time to "check if the state of the car was normal". For P18, even though he did not always look up and check, he still felt distracted without the non-critical cue:

> " Even though I was not looking up actively. I was consciously paying attention to my peripherals. even though I was watching it (the video). I could not concentrate on it." -P18

These participants didn't perceive the interference from the cue as a distraction but as positivity. The non-critical cue made them more relaxed because they didn't have to worry about or take care of the car's movement by themselves, but the vehicle would warn them:

> " I did not need to ... keep looking, observing. I thought the car was already warning me. I find it quite useful ... Then I actually spend more time watching the video because I knew that later the car would warn me." - P25

The bad observability in the no non-critical cue round, as indicated previously, could be the reason for this experience. P14 felt the need to look up and thus pay less attention to the tablet because he "didn't know exactly what was going to happen" without the cue. This means he did not have sufficient knowledge about the intention of the automated vehicle, which indicates poor observability.

In addition, the participants who felt more distracted in the round without the non-critical cue usually expressed their feeling of unsafety during that round. Most of these participants were afraid of hazards and became worried, nervous or concerned. P18 expressed that he checked the car and the road actively "to see if there are any hazards" in the round without the non-critical cue. These participants were not distracted by outside information but by their inner feelings. The unsafe feeling for the AVs forced them to check the car and the environment. For example, P31 described that he was "more distracted by himself". The original propensity to trust the automated vehicle of the participants could relate to this. The mean value of participants who rate the round without the non-critical cue more distracted is 3.39, lower than all participants' average AV trust propensity.

However, the feeling of safety is not only related to their prior trust of AVs. After the participants got in the car, their trust would change based on their experience in the car and the performance of the car. And vice versa, this "after trust" would also influence their experience of the left journey in the car. For example, P29 doubted if he could trust the car when there were some harsh movement. Then he felt the no cue round was more distracted although his propensity to trust AVs was 3.67, which is higher than the average.

Furthermore, other factors could also influence the experience of distraction, except for the non-critical cue. Returning briefly to the car dynamic, it could function as a non-critical cue when it is present at a proper level. It could also be very harsh to the driver. Whether gentle or aggressive, car dynamics could trigger the participants to observe their surroundings. For example, P06 would look up both when there was a hard brake and a significant turning. The former is aggressive, and the latter is normal. Further, it could be a distraction sometimes. For example, P23 explicitly describe the car movement as a distraction:

> " When the car moves, of course, it's a distraction. Cause I was watching the video, when there is a movement, I would instinct to see what's going on here." -P23

5.3 Other factors and the experience

Besides the factors relating to the non-critical cue, other factors also play a role in the experience of the conditionally automated driving scenario.

The driving style of the automated vehicle could influence the driver's experience of automation surprise.

Firstly, the unmatched driving style between the driver and the car would negatively influence the automation surprise from the perspective of observability. P12 has his expectation of the movement of the car, and only when the vehicle meets his expectation could the cue help with observability. P18 could not predict the car's movement in the round without the non-critical cue because he "would behave in a different way than the car", and thus became nervous. Trust on the AV would also be affected. P02 doubted whether she could trust the car because she would keep a larger distance than the AV while driving. Additionally, the automated vehicle's driving style that deviates from good driving norms may make it difficult to trust and understand:

> " Driving style, not following the norms and good practices, doesn't secure or reliable, then not at ease to enjoy NDRAs,(...). I feel the need to constantly look on the road." -P25

According to the interview, many participants realized they were more familiar with the automated vehicle and the environment during the experiment. The familiarity increased both during and between the two rounds of the experiment. Meanwhile, this familiarity influences the participants' understanding of the intention of the vehicle and their perception of the environment. For example, P17 expressed he looked up less to the surroundings because he got more used to the set-up of the experiment. P30 stated that he had a clear mind in the second round because he found out it would be the same route.

From the video recording, we could also find the influence from the participants' familiarity to the experiment route. If the participants are more familiar with the environment, they will understand the intention of the car easier. For example, P24 stated that he was very familiar with the experiment route, including the speed limitation and the speed camera. During the experiment, When the WoZ car suddenly braked at the speed camera in the first round, P24 looked around for about 8 seconds after the deceleration manoeuvre was completed. From his obvious smiling facial expression right before he looked back to the tablet, we assumed that he realized the sudden brake was due to the speed camera. Then in the second round, he only kept observing for three more seconds after the sudden deceleration ended.

5.4 Video Recording Findings

From coding all the videos, there are also some interesting findings related to the users' general experience. Some of the findings may not be able to answer the research questions directly. However, we list and discuss the findings here because they may provide insights into future design and research or be in line with the interviews.

When coding the video, we differentiated the "look up" and "look around" behaviour by three seconds. According to the interview, we knew that the participants would possibly look up or look around because of the car dynamic or receiving the non-critical cue. This could also be seen from the video recordings.

In the video recordings, when there are movements of the car or environmental events, we can sometimes see the participants would move their eyes from the tablet to the screen on the partition or the windows. Depending on the intensity of the car dynamics or the type of events in the environment, different behaviours were conducted by different participants after they moved their eyes.

Taking lane change as an example, sometimes the participant would have a glance at the front view right after the car's movement. Sometimes they did not look up at all. As mentioned before, we required the wizard driver to drive defensively but to perform high-quickness lane changes in specific areas. This caused the different quickness of lane changing in the experiment trials, to which the participants reacted differently. The phenomenon is consistent with the interview, where some participants said they looked up when they felt the movement of the car and some of them mentioned they sometimes could sense the car dynamic and felt no need to look up. However, when the car dynamics were harsh, the participants would look around, i.e., gazing at the surroundings for more than three seconds, sometimes including the side and rear windows. For example, all the participants looked around at the sudden deceleration point (if the manoeuvre was conducted).

A possible reason behind this difference in gazing time could be the different environmental events and AV's behaviours cause different cognitive workloads of the participants. Studies showed that gaze behaviour is an indicator of cognitive workload (Gold et al., 2016). In the interview, a majority of the participants also implied their cognitive process after perceiving the events or the car's behaviours. When talking about the sudden acceleration and the harsh lane changes, they expressed that they were curious about what was happening and could not figure it out after observing the surroundings.

Furthermore, there is another interesting finding from the video recordings. The participants could be possibly amused by the NDRT in the experiment, which was watching the funny videos on the provided tablet. And when they were amused, they would smile and glance at the surroundings. However, none of the participants mentioned this look-up behaviour in the interviews. Thus, we assumed there was an unconscious break from focusing on the funny videos when the participant was amused. If the future HMI in the AVs could provide information via this focusing break, it may help maintain the attention level of the participants without disturbing them.

Discussion 6

In this chapter, we compare and contrast our results with those of other studies in the same area. We further discuss the possible reasons for the findings in the last chapter. We also examine the limitations of our study and suggest areas for future research.

6.1 Timing of the Non-Critical Cue

As indicated in the previous chapter, the non-critical cue could potentially reduce AS from the perspective of timeliness, because it prompted the user sometime before the vehicle's movement and the events in the environment. However, according to this experiment, it is not clear how much time before the car's movement the cue should be provided to the user.

A similar concept in the automated driving scenario is the take-over time budget, which refers to the period from the Take-over Request (TOR) to the automation system limitation (Zhang et al., 2019). This time budget should be more than the Take-over Time (TOT), which covers a process from the user perceiving the stimulus to taking actual action (Zhang et al., 2019). Similarly, regarding the non-critical scenario of this research, the user should also be able to raise a proper attention level within the duration between the non-critical cue and the events. Ideally, the user should stay in the attention level "Change blind" defined by Riener et al. (2017) as mentioned in the previous chapter. Different from the TOT, when at this attention level, no immediate action would be taken by the user (Riener et al., 2017). Therefore, measuring the user's response time to the non-critical cue could be more difficult than measuring TOT, which could use the driver's behaviour such as hands-on time, braking or steering as reference (e.g., Zhang et al. (2019) , Gold et al. (2013),Kerschbaum, Lorenz, and Bengler (2015)).

Regarding the 'time budget' in the non-critical conditionally automated driving scenario, the cognitive process after receiving the non-critical cue still needs to research in future work.

6.2 Car Dynamics and Driving Style

In the previous chapter, we found that when the car dynamic is at a proper level, it could be functional to AS similarly as the non-critical cue. Meanwhile, the driving style of the automated vehicle could influence the driver's experience in the non-critical conditionally automated driving scenario. There are many connections between the car dynamic and the driving style, which could also be seen from the interview. When the participants compared the driving style of the vehicle with that of themselves, the examples they mentioned were usually the vehicle manoeuvres such as harsh brakes and sudden lane changing (P21, P23, P25).

According to Sagberg et al. (2015), Driving style is a subcategory of driving behaviours, which were classified into three levels: strategic, tactical, and operational by Michon (1985). Driving style may belong to all these three levels, for example, steering or acceleration habits at the operational level, habitual choice of speed and headway at the tactical level, and habitual route at the strategic level (Sagberg et al., 2015). The car dynamics (i.e., the car movements or motions) are the results of some driving behaviour habits such as steering or acceleration. Theoretically, the car dynamic could also connect to driving style at the operational level. That is to say, the driving style could influence the AS at the operational level by means of car dynamics.

Findings from Bellem et al. (2018) suggest minimising acceleration and jerk during acceleration manoeuvre, and for lane changes, both small accelerations and early motion feedback are advised. However, we find that if the car dynamic is too subtle to be sensed, the human driver may feel unsafe and nervous during automated driving. Meanwhile, very aggressive driving properties also disturb the users. This is consistent with the findings from Dettmann et al. (2021). They found that a defensive automated driving approach with specific acceleration, deceleration and speed parameters would be the most promising to improve users' driving experience (Dettmann et al., 2021). Thus, we argue that the car dynamic should generally be under the defensive driving style at the operational level. However, the specific intensity of car dynamics to contribute to reducing automation surprise needs further discussion.

Furthermore, the driving style could possibly influence the AS at the tactical and strategic level. For example, P25 mentioned in the interview that he had difficulty understanding why the vehicle constantly changed lanes, which he would not do. This is a comparison of their driving styles at the tactical level by the participant. And this difference in driving style at the tactical level result in bad observability of the participant. P34 and P18 were familiar with driving in the experiment area. So when the car switched lanes, they were not surprised because they could expect the deviation of the route. At the strategic level, if the route that the AV choose is also the habitual route of the human driver, there will be better predictability and less possibility of experiencing AS.

6.3 The Design of the Non-critical Cue

In this study, we focused on exploring and testing the concept of the non-critical cue rather than designing of the non-critical cue. After the experiment, we obtained various insights for improving the design of the non-critical cue. And the concept could be better researched and developed in future work.

The prototype of the non-critical cue we developed was a vibrotactile wristband that only had a single mode of vibration to prompt the user. However, there are various AV's behaviours and events in the environment. This single-mode vibration could represent varieties of behaviours and events in the experiment, which usually confuses the users. The user needs to think about the exact meaning of the cue, which increases the mental workload of the users. Consequently, there would be higher chances for the user to feel disturbed or even interrupted by the cue while conducting NDRTs.

In addition, in this experiment, the prototype prompted the participants at every vehicle manoeuvre. This may be redundant because some of the subtle car dynamics do not need to be noticed by the user. For example, P04 was slightly bothered by the prototype because it frequently vibrated at events that he did not care about. P26, as mentioned in the previous chapter, was cued at the low-quickness lane change that he did not notice, which resulted in his confusion. And he even turned around to check the rearview, wondering if there was anything abnormal in the environment.

Furthermore, the ambiguous meaning of the prototype and its vibrating at all events could also potentially trigger some instinct behaviours, which were even opposed to the goal of the design. For example, P21 was prompted by the prototype while looking around, and he instinctively started to watch the video, which is opposite to the goal of the cue. We assume that this was because the exact meaning of the cue was unclear to him. Meanwhile, he had the mindset that his primary task was watching the tablet, which he was not following at that time. Thus he naturally switched back to the NDRT when receiving the cue.

In future work on improving the non-critical cue and exploring more on the concept more, we suggest designing the cue with more severity or other variations, which could help the user distinguish the different conditions to some extent. However, there is a trade-off. If the prototype is too complex, matching the cue variety and the specific event would also increase the cognitive workload. Additionally, the cue could only pass information to participants at selected conditions rather than every vehicle's behaviours and environmental events.

Limitations

First of all, there are some limitations to the experiment setup. As a conditionally AV, the interior should be in line with the manual vehicle. However, the rear view was lacking compared to the manual driving car, which could influence the user to observe the environment directly. Meanwhile, because of the partition and the noise, the participants could not hear the sound of the turning indicator clearly. The role of indicator in the HAD scenario could not be fully explored. In addition, the TV screen showing the front view attracted more attention from the participants than the front view should. For example, P21 felt "zoomed in" to the front view because of the TV screen. One of the possible reasons could be the light from the screen is different from the natural light source. This deviation influenced the participants' attention and their observing behaviours, especially when it was dark outside, such as on rainy days. In further studies, the mobility lab could be improved based on the above limitations before conducting the research.

Secondly, the limitation of the WoZ method would influence the prototype implementation. The way of applying the WoZ method in this experiment was the experimenter pressed the button and sent the signal to the participants. However, the experimenter sitting in the front of the car did not have a proper rear view of the vehicle, as well as of the wizard driver's side. If a truck were passing the experiment car from behind, especially from the left lane (driver's side), there would be a great possibility that the experimenter could not prompt the participant timely.

Furthermore, there are also some limitations of the on-road study, that is, we cannot precisely control all the influential factors same in both rounds of the experiment. For example, in the experiment of this study, although two rounds were conducted consecutively within 1 hour, the traffic intensity can still be different. P14, as mentioned in the previous chapter, experienced eight times huge trucks passing by but zero in the other round. The scripted manoeuvres also possibly cannot be performed as they could only be conducted safely under certain traffic conditions. For instance, the sudden brake could only be conducted when no other cars were behind closely.

Conclusion 7

The main objective of this study was to research how the non-critical cue influences the driver's experience in the HAD scenario. We aimed to explore the effect of the non-critical cue experience from the aspects of the AS and whether the cue would disturb the user conducting NDRTs. The role that the inherent cues from the conditionally AV, such as the car dynamics, play in the scenario was also investigated.

An on-road within-subject experiment on the highway and motorway near the city was performed. We collected valid experiment data, including the questionnaires, IMU data, video recordings and interview recordings, from trials with 20 participants. The data was analyzed using the qualitative-oriented mixed method.

The results indicate that the non-critical cue could potentially impact the user experience on AS from three perspectives: timeliness, observability, and predictability. Regarding timeliness, the cue can possibly help avoid AS because it is provided sometime before the movement of the AV and events in the environment. However, the specific duration that the cue should be given in advance may relate to the cognitive process of the user in this scenario, which needs further research. Then speaking of observability, the non-critical cue is possible to reduce AS by directly increasing observability. In addition, the non-critical cue could also trigger the user to observe the surroundings or work mutually with the car dynamics to increase observability and contribute to the reduction of AS. The increment of predictability could come from the acceptable level of observability or be provided by the non-critical cue directly.

In addition, the non-critical had a great possibility of distracting the users by trigging them to observe the surroundings. However, this does not necessarily mean the user would feel disturbed or be interrupted by the cue from conducting NDRT. In contrast, when the non-critical was absent, users may have felt nervous or unsafe due to the limited information available. These feelings would urge them frequently check their surroundings and cannot focus on the NDRTs.

Regarding the inherent cues, the car dynamics could function similarly as a non-critical cue to reduce AS from the observability aspect, if it results from the defensive driving style of the AV at the operational level. If the car dynamic is too subtle or too harsh, it could not contribute to avoiding AS. The driving styles at the tactical and strategic levels also affect AS in observability and predictability. The turning steering wheel in the conditionally AVs could improve the observability by showing the intentions and the status of the AV.

In future works, the design of the non-critical cue could be improved from its variety. And the more specific topics related to the non-critical cue in the HAD could be investigated, such as how much time the cue should prompt the user in advance or what is the cognitive process when receiving a non-critical cue.

Appendix

A

Wizard Driver Instruction

A.1 General Driving style

You are supposed to keep a defensive driving style in general, expect for when asked to conduct special tasks. The general requirements for defensive driving styles are as follows:

1. Always referring to the AUTOAccD, where the red bar indicates assertive driving behaviours and the green bar indicates defensive ones.

NOTE Keep the bars in green to maintain a defensive driving style.

2. Look "Left–Right–Left" before proceeding through intersections.
3. Do not go through intersections on a yellow light; pick a point of decision and don't get caught in the middle of the intersection.
4. Check the road 12 - 15 seconds ahead on the highway; 4 - 6 seconds in city driving (about two blocks).
5. Check mirrors every 3 - 5 seconds in the city, and 5 - 8 on the highway.
6. When following the front car, the following time should be 3 seconds from the vehicle in front of you, depending upon the speed and whether in city or highway driving.
7. Make sure that you leave yourself plenty of space so you can come to a stop gently rather than all at once.
8. Count a 2-second delay at intersections to allow following appropriate time from the car in front of you is also required. Mind the stopping distance requirements for the vehicle you're driving, including the effects of inclement weather conditions.

A.2 Driving procedures and manoeuvres

A.2.1 Before start

- Hidden from participants The driver should be sitting in the car before the participants come and keep quiet to hide from participants.
- Check before driving Adjust the mirrors to the maximum visibility which suits you. Check instruments and gauges. A prepared card with the text "start" on it would be shown to you when you are supposed to start driving. Make sure the hand brake is released before leaving.

A.2.2 On-road Manoeuvre Requirements

The manoeuvres you would be required to conduct today are as follows:

- Lane change(left,right)

- Overtake
- Deceleration

Meanwhile, certain special manoeuvres should be done during the journey when the experimenter on the passenger seat shows the task instruction on a prepared card to you. When you are going to do the manoeuvres, please show a gesture signal to me. This is only required in the experiment part with the prototype.

A.2.3 Emergency Situation

We also provide an emergency button for the participants for abortion of the study at any time without stating a reason. In case of driving on an urban road, you need to stop safely as soon as possible. In case of driving on a highway, you need to exit the highway and find a proper place to stop safely as soon as possible.

A.2.4 After the experiment

After the experiment, you should drive the car back to the same parking spot. If needed, you could only get off the car without seeing by the participants and leave the car from the front end. After the interview, you could greet with the participants if you wish.

A.3 Reminders

- All the driving behaviours are supposed to conduct under a safety conditions!
- All the required manoeuvres can only be done if you feel possible and safe to do so.
- During the experiment, you would drive in a certain consistent driving style and would be required to perform several manoeuvres during the journey.
- Due to the modification of the experiment car, there will be extended blind spots. You cannot glance over the shoulder in the direction you plan to move and the middle rear-view mirror in the car is disabled because of the partition. Keep that in mind and be more careful when you are going to manoeuvre.
  Here are some tips. Take care to avoid putting yourself in the blind spots of other vehicles. A good rule of thumb is that if you can't see a trucker's mirrors, the trucker can't see you.

B Questionnaires

B.1 Motion Sickness Susceptibility Questionnaire

Your childhood experience only(before 12 years of age), for each of the following types of transport or entertainment please indicate

1. As a child (before age 12), how often you felt sick or nauseated(tick boxes)

Table B.1: Motion sickness susceptibility questionnaire short-form (MSSQ-Short) 1

	Not applicable - Never Travelled	Never Felt Sick	Rarely Felt Sick	Sometimes Felt Sick	Frequently Felt Sick
Cars					
Buses or Coaches					
Trains					
Aircraft					
Small boats					
Ships, e.g. Channel Ferries					
Swings in playgrounds					
Roundabouts in playgrounds					
Big Dippers, Funfair Rides					
	t	0	1	2	3

Your experience over the last 10 years (approximately), for each of the following types of transport or entertainment please indicate

2. Over the last 10 years, how often you felt sick or nauseated(tick boxes)

Table B.2: Motion sickness susceptibility questionnaire short-form (MSSQ-Short) 2

	Not applicable - Never Travelled	Never Felt Sick	Rarely Felt Sick	Sometimes Felt Sick	Frequently Felt Sick
Cars					
Buses or Coaches					
Trains					
Aircraft					
Small boats					
Ships, e.g. Channel Ferries					
Swings in playgrounds					
Roundabouts in playgrounds					
Big Dippers, Funfair Rides					
	t	0	1	2	3

B.2 Introductory Questionnaire

B.2.1 Demographic information

In this part, we will collect demographical data such as age, gender, driving experience and etc.

1. Participant ID:_________

2. Age:_________

3. Gender:

◯ Male

◯ Female

◯ Other

4. Nationality:_________

5. How often do you drive?

◯ Never

◯ A few times per year

◯ A few times per month

◯ A few times per week

◯ Almost everyday

◯ Everyday

7. How long have you been driving?

◯ Less than one year

◯ 1-2 years

◯ 2-3 years

◯ More than 3 years

8. Have you had any experience with Advanced Driver Assistant System (ADAS, such as adaptive cruise control, lane keeping, and automatic parking.)?

◯ Yes, I use them quite often

◯ Yes, but only try them a few times

◯ No, I know about them but never tried them

◯ No, I had no idea what that means

◯ Other:_________

B.2.2 Propensity to Trust of Automated Vehicles

The following questions assess an individual's propensity to trust machines. Answer the following questions regarding your current behaviours, not what you want your behaviours to be. Your responses are completely anonymous. Your answers to the following questions will NOT be shared or linked to any identifying information.

8. I usually trust automated vehicles until there is a reason not to.

○ Strongly disagree

○ Somewhat disagree

○ Neither agree nor disagree

○ Somewhat agree o Strongly agree

9. For the most part, I distrust automated vehicles.

○ Strongly disagree

○ Somewhat disagree

○ Neither agree nor disagree

○ Somewhat agree

○ Strongly agree

10. In general, I would rely on an automated vehicle to assist me.

○ Strongly disagree

○ Somewhat disagree

○ Neither agree nor disagree

○ Somewhat agree o Strongly agree

11. My tendency to trust automated vehicles is high.

○ Strongly disagree

○ Somewhat disagree

○ Neither agree nor disagree

○ Somewhat agree

○ Strongly agree

12. It is easy for me to trust automated vehicles to do their job.

    ○ Strongly disagree

    ○ Somewhat disagree

    ○ Neither agree nor disagree

    ○ Somewhat agree

    ○ Strongly agree

13. I am likely to trust an automated vehicle event when I have little knowledge about it.

    ○ Strongly disagree

    ○ Somewhat disagree

    ○ Neither agree nor disagree

    ○ Somewhat agree

    ○ Strongly agree

B.3 Post-experiment Questionnaire

Measure the following attributes.

Table B.3: User Experience Questionnaire (UEQ)

annoying	◯	◯	◯	◯	◯	◯	◯	enjoyable
not understandable	◯	◯	◯	◯	◯	◯	◯	understandable
creative	◯	◯	◯	◯	◯	◯	◯	dull
easy to learn	◯	◯	◯	◯	◯	◯	◯	difficult to learn
valuable	◯	◯	◯	◯	◯	◯	◯	inferior
boring	◯	◯	◯	◯	◯	◯	◯	exciting
not interesting	◯	◯	◯	◯	◯	◯	◯	interesting
unpredictable	◯	◯	◯	◯	◯	◯	◯	predictable
fast	◯	◯	◯	◯	◯	◯	◯	slow
inventive	◯	◯	◯	◯	◯	◯	◯	conventional
obstructive	◯	◯	◯	◯	◯	◯	◯	supportable
good	◯	◯	◯	◯	◯	◯	◯	bad
complicated	◯	◯	◯	◯	◯	◯	◯	easy
unlikable	◯	◯	◯	◯	◯	◯	◯	pleasing
usual	◯	◯	◯	◯	◯	◯	◯	leading edge
unpleasant	◯	◯	◯	◯	◯	◯	◯	pleasant
secure	◯	◯	◯	◯	◯	◯	◯	not secure
motivating	◯	◯	◯	◯	◯	◯	◯	demotivating
meet expectations	◯	◯	◯	◯	◯	◯	◯	does not meet
expectation inefficient	◯	◯	◯	◯	◯	◯	◯	efficient
clear	◯	◯	◯	◯	◯	◯	◯	confusing
impractical	◯	◯	◯	◯	◯	◯	◯	practical
organized	◯	◯	◯	◯	◯	◯	◯	cluttered
attractive	◯	◯	◯	◯	◯	◯	◯	unattractive
friendly	◯	◯	◯	◯	◯	◯	◯	unfriendly
conservative	◯	◯	◯	◯	◯	◯	◯	innovative

C

Interviews

C.1 Interview questions

The following questions would be asked for both rounds of the experiments. For each question, the participants need to talk about both parts of the experiment separately and then compare them.

1. How do you feel about the journey? Can you compare the two rounds? Why?

2. What do you think about the communication/mutual understanding with your vehicle? Can you compare the two rounds? Why?

3. How easy do you find it to predict the behaviour of the car? or the events in the environment? Can you compare the two rounds? Why?

4. How timely do you find it to understand the behaviour of the car? And for the events in the environment?

5. From one to seven, how much do you feel distracted/intruded when conducting NDRTs in each round? If yes, what disturbed you?

6. Is there anything from the vehicle's behaviour or the environment that surprised/shocked you? If yes, what?

7. Was there anything you expected to find that was not there? If yes, what?

8. How do you feel about the cue? What modality do you prefer?

9. How do you understand the experiment setup, and how automatic the car is with the setup?

C.2 Interview Coding Scheme

Code		Definition	Example
OBSERVABILITY		can the participants understand/detect the car's intention/movement, and be aware of something is going to happen in the environment	
	cue-observability	the non-critical cue itself improved observability directly	On the first one (with cue), I could sense that there was something about to happen.(P18)
	no cue-observability	no cue decreases observability	Yeah, so compare with the part two, so when there's no prototype means that there's no way for me to know the future. intent of the of the vehicle (P11)
PREDICTABILITY		based on sufficient observability, so that the participants can adjust/plan their behavior, or self-prepare	
	cue-predictability	the non-critical cue itself improved predictability	And if you have a vib for instance, you can say oki now it's gonna do something. Then I can track it doing it, then it's... I can grab the steer or something, or push the button. (P29)
	no cue-predictability	no cue descreased predictability	in the first phase (no cue), I didn't really look up and notice stuff, so I didn't really have to predict anything. Just go with the flow, I guess. (P34)
TIMELINESS		the relavant info being provided timely enough for the participants to form undersanding or take action	
	cue-timeliness	the non-critical cue improved timeliness	P2 (with cue), you are aware something would happen in advance. It gives you longer time to react it, I think it makes it easier to understand. (P04)
	no cue-not timeliness	no cue decreased timeliness	(when without cue) Only when there is a sudden change of the acceleration, you can know that something is happening. (P11)
OBSERVE		Observe refers to the participants physically look up or look around, check their surrondings, environment, steer and car movement, or just general where they are and what is going on	
	cue cause observ	cue cause observation	So every time it's vibrates, I was really checking what's happened. (P02)
	cue & observ>observability	the non-critical cue cause observation, which result in good understanding	it's trying to communicate with me, tells me his decision, what will happen. I expect what would happen, and I understand why. When I move my attention out of the video, watch the screen, I can understand why it trys to give me the signal. (P04)
	no cue-other reason cause observe	no cue, but still actively observe. This could be caused by other factors	For p1, that's my personal motivation, I would like to check where I am I , the speed, the travel situ, my subjective motivation. (P04)
	cue & observation-unmatch	the partcipants don't understand why there is a cue, even with observation	When it vibrated, then I thought maybe it indicates that I need to pay attention to something. But then I looked around, there was not much to pay attention to. Maybe there was like a car changing lane in front of our car. (P26)
	observe steer	observe the steering wheel	I notice the wheel is running by itself, it's a really good sign, because then I know it's turning left or right. (P23)
CAR DYNAMIC		car dynamic means the car's movement, turning left turning right, brake, etc. related to maneuver and driving style. But this code focus on the car dynamic cause observation or being observed.	
	car dynamic cause observ	car's movement cause observation	And if the car made some movement, I am also observing what's happening. (P30)
	car dynamic-observability	car's movement improved observability or predictability	In p1, when it starts moving, or makes a move, then I know it will change the lane. (First Round No Cue\P23: 20) you know, maybe know what the information is, very gently. It's just a lane change and it's a tiny tap. And then maybe if it's in combination with the car going around a corner or something then you know what's happening without having to look up. (P09)
DISTRACTION		being distracted from conducting NDRAs (watching the video)	
	same distracting	the distraction level of the two rounds are the same	When the car moves, ofc it's a distraction, cause I was watching the video, when there is a movement, I would instinct to see what's going on here. In p2 it's kinda the same, even with the vibration, and every vibration and the movement, I would check what's going on. (P23)
	with cue more distracting than no cue	feel more distracted in with cue round, or the cue itself is perceived as distracting feel less distracted in no cue round	Yeah, I feel distracted. When I got the vibration I was like, directly checking what's going on. (First Round No Cue\P02: 17-18) Bc P1 (no cue) I don't know what's going to happen, there is no vib to interrupt my watching. P1 like when it moves, I start to look up what's happening. (P30)
	no cue more distracting than with cue	feel more distracted in no cue round feel less distracted in with cue round	But in p1, it was more like a constant, look around to see if there are any hazards. Actively looking for hazards in the first one. (P18)
WITH THE CUE		Any other ideas or penomenon relate to with the cue round	
	with the cue-negative	negative thoughts about the round with the cue	But if some emergent situ that the automatic car feel it doesn't have any confidence, it needs my attention, put my hands on the vehicle. It can give me some vib, because vib is a on-body feeling, I think it's a more high level inform, comparing to other visual signal. (P04)
	the cue trigger something else	cue trigger other behavior rather than look up/look around	because I was watching this, this video. And there was even at one stage, I was watching the screen, because the video was just repeating the same kind of videos. So I was watching the screen. And then the vibration went off. And my instinct was to look at the video instead. So it gave me a trigger, something was up. But I was already looking at the road. So look at this. So my mind in that instinct, misinterpreted the cue, told me the wrong thing. Or at least, it got me to do the wrong thing.(P21)
WITHOUT THE CUE		Any other ideas or penomenon relate to without the cue round	
	no cue-negative	nervous, worried, confusing	without prototype I was way more, how you say it, not relaxing. (P02)
	no cue - just sitting in	- the participants feel that they are being driven by the car, which is in full control or drives well - no attentive to anything	On the second part, while there was no information coming, telling us if there was any change, so I cannot really say anything. I just didn't feel like I was receiving any information. The car was just driving, and I was there. (P01)
TACTICAL MANEUVER		emphasize the decision the car made to change lane, over take or brake. Emphasize the it may cause automation surprise.	
	unexpected maneuver	the maneuvers are - unexpected - perceived as unnecessary - not understood by the participants - suddenly	Well, some harsh braking. I did not understand why that was happening because I didn't see any car in front. (P01)
DRIVING STYLE		emphasize the general experience of the driving	
	same driving style in two rounds	perceive the two rounds had the same driving style	
	general positive driving style	- match with 'my' driving style, expectation, norms - moderate, calm...	
	general negative driving style	- not match with 'my' driving style, expectation, norms - reckless, dangerous	
FAMILIAR WITH THE ENVIRONMENT		familiar with the environment or get familiared during the trials	The more familiar I am with the environment, the more relaxed I am. Because this area is a little bit familiar with me, so I am sort of knowing what to expect in terms of the road and the environment, (P18)
NDRAs		the selected NDRAs (watching the video) has influence on the experience	I didn't want to look up from the video because I wanted to watch the video. So I didn't really check like whatever if there was a vehicle passing that's why we maybe changed lanes. (P17)
FEELINGS		feeling of concerned, trust, safety, personal motivation, curious	But in p1, it was more like a constant, look around to see if there are any hazards. Actively looking for hazards in the first one. (P18)
PROOF OF CONCEPT		Proof of concept is the proposed vibrotactile cue, the code includes two aspects that is targeted at the concept (cue), rather than the haptic modality design	
	severity of the situ decides the needs for the cue	1. the cue was not needed, or cue too much 2. hazard situ, more cue needed (warning, missed cue)	On the highway, there was a car gets way too close to our car. That might even be the other car's fault, but I would expect a warning. (P25)
	Prototype too simple	The prototype should have more variations for different events and behaviours.	Also the vib, is just a signal, the info is too simple to me. So that make it more complicated to understand the situ, but normally I found out the situ is not dangerous, I don't need to be aware of it. (P04)

Figure C.1: Interview Coding Scheme

Bibliograph

Carsten, Oliver and Marieke H Martens (2019). 'How can humans understand their automated cars? HMI principles, problems and solutions'. In: *Cognition, Technology & Work* 21.1, pp. 3–20 (cited on pages 1, 3, 4, 9–12, 23, 32, 37).

Detjen, Henrik, Sarah Faltaous, Bastian Pfleging, Stefan Geisler, and Stefan Schneegass (2021). 'How to increase automated vehicles' acceptance through in-vehicle interaction design: A review'. In: *International Journal of Human–Computer Interaction* 37.4, pp. 308–330 (cited on page 1).

Kyriakidis, Miltos, Joost CF de Winter, Neville Stanton, Thierry Bellet, Bart van Arem, Karel Brookhuis, Marieke H Martens, Klaus Bengler, Jan Andersson, Natasha Merat, et al. (2019). 'A human factors perspective on automated driving'. In: *Theoretical Issues in Ergonomics Science* 20.3, pp. 223–249 (cited on page 1).

SAE International (n.d.). *J3016C: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. URL: https://www.sae.org/standards/content/j3016_202104 (cited on pages 1, 2, 5, 6, 8).

Zhang, Bo, Joost De Winter, Silvia Varotto, Riender Happee, and Marieke Martens (2019). 'Determinants of take-over time from automated driving: A meta-analysis of 129 studies'. In: *Transportation research part F: traffic psychology and behaviour* 64, pp. 285–307 (cited on pages 2, 43).

Ayoub, Jackie, Feng Zhou, Shan Bao, and X Jessie Yang (2019). 'From manual driving to automated driving: A review of 10 years of autoui'. In: *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications*, pp. 70–90 (cited on pages 2, 10).

Dehais, Frederic, Vsevolod Peysakhovich, Sébastien Scannella, Jennifer Fongue, and Thibault Gateau (2015a). '" Automation Surprise" in Aviation: Real-Time Solutions'. In: *Proceedings of the 33rd annual ACM conference on Human Factors in Computing Systems*, pp. 2525–2534 (cited on page 3).

Victor, Trent W, Emma Tivesten, Pär Gustavsson, Joel Johansson, Fredrik Sangberg, and Mikael Ljung Aust (2018). 'Automation expectation mismatch: Incorrect prediction despite eyes on threat and hands on wheel'. In: *Human factors* 60.8, pp. 1095–1116 (cited on pages 3, 11).

Weiser, Mark and John Seely Brown (1996). 'Designing calm technology'. In: *PowerGrid Journal* 1.1, pp. 75–85 (cited on pages 4, 7).

National Highway Traffic Safety Administration (n.d.). *Automated Vehicles for Safety*. URL: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety#the-topic-road-to-full-automation (cited on page 5).

Richards, Dale and Alex Stedmon (2016). 'To delegate or not to delegate: A review of control frameworks for autonomous cars'. In: *Applied ergonomics* 53, pp. 383–388 (cited on page 5).

He, Xiaolin, Jork Stapel, Meng Wang, and Riender Happee (2022). 'Modelling perceived risk and trust in driving automation reacting to merging and braking vehicles'. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 86, pp. 178–195. DOI: https://doi.org/10.1016/j.trf.2022.02.016 (cited on page 6).

Weiser, Mark and John Seely Brown (1997). 'The coming age of calm technology'. In: *Beyond calculation*. Springer, pp. 75–85 (cited on pages 7, 10).

Matthews, Tara, Tye Rattenbury, and Scott Carter (2007). 'Defining, designing, and evaluating peripheral displays: An analysis using activity theory'. In: *Human–Computer Interaction* 22.1-2, pp. 221–261 (cited on pages 7, 8).

De Guzman, Edward S, Margaret Yau, Anthony Gagliano, Austin Park, and Anind K Dey (2004). ‘Exploring the design and use of peripheral displays of awareness information’. In: *CHI’04 extended abstracts on Human factors in computing systems*, pp. 1247–1250 (cited on page 7).

Endsley, Mica R (1995). ‘Toward a theory of situation awareness in dynamic systems’. In: *Human factors* 37.1, pp. 32–64 (cited on page 8).

Sanquist, TF, BR Brisbois, and M Baucum (2016). ‘Attention and situational awareness in first responder operations’. In: *Pacific Northwest National Laboratory, Alexandria, VA* (cited on page 8).

Riener, Andreas, Myounghoon Philart Jeon, Ignacio J. Alvarez, and Anna-Katharina Frison (2017). ‘Driver in the Loop: Best Practices in Automotive Sensing and Feedback Mechanisms’. In: (cited on pages 8, 9, 43).

Beyer, Gilbert, Gian Mario Bertolotti, Andrea Cristiani, and Shadi Al Dehni (2010). ‘An adaptive driver alert system making use of implicit sensing and notification techniques’. In: *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Springer, pp. 417–424 (cited on page 8).

Boelhouwer, Anika, Jelle Dijk, and Marieke Martens (June 2019). ‘Turmoil Behind the Automated Wheel: An Embodied Perspective on Current HMI Developments in Partially Automated Vehicles’. In: pp. 3–25. doi: 10.1007/978-3-030-22666-4_1 (cited on pages 9, 10).

Karatas, Nihan, Soshi Yoshikawa, Shintaro Tamura, Sho Otaki, Ryuji Funayama, and Michio Okada (2017). ‘Sociable driving agents to maintain driver’s attention in autonomous driving’. In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 143–149 (cited on page 10).

Telpaz, Ariel, Brian Rhindress, Ido Zelman, and Omer Tsimhoni (2015). ‘Haptic seat for automated driving: preparing the driver to take control effectively’. In: *Proceedings of the 7th international conference on automotive user interfaces and interactive vehicular applications*, pp. 23–30 (cited on page 10).

Walker, Francesco, Oliver Morgenstern, Javier Martinez Avila, Marieke Martens, and Willem Verwey (2021). ‘A Novel Technique for Faster Responses to Take Over Requests in an Automated Vehicle’. In: *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. AutomotiveUI ’21. Leeds, United Kingdom: Association for Computing Machinery, 159–164. doi: 10.1145/3409118.3475152 (cited on page 10).

Liu, Weimin, Qingkun Li, Zhenyuan Wang, Wenjun Wang, Chao Zeng, and Bo Cheng (2022). ‘A Literature Review on Additional Semantic Information Conveyed from Driving Automation Systems to Drivers through Advanced In-Vehicle HMI Just Before, During, and Right After Takeover Request’. In: *International Journal of Human–Computer Interaction*, pp. 1–21 (cited on page 10).

Dehais, Frederic, Vsevolod Peysakhovich, Sébastien Scannella, Jennifer Fongue, and Thibault Gateau (2015b). ‘"Automation Surprise" in Aviation: Real-Time Solutions’. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. Seoul, Republic of Korea: Association for Computing Machinery, 2525–2534. doi: 10.1145/2702123.2702521 (cited on page 11).

Sarter, Nadine B and David D Woods (1995). ‘How in the world did we ever get into that mode? Mode error and awareness in supervisory control’. In: *Human factors* 37.1, pp. 5–19 (cited on page 11).

Monsaingeon, Noé, Loïc Caroux, Sabine Langlois, and Céline Lemercier (Oct. 2021). ‘A systematic review of mode awareness measurements for automated driving’. In: (cited on page 11).

Kelley, John F (2018). ‘Wizard of Oz (WoZ) a yellow brick journey’. In: *Journal of Usability Studies* 13.3, pp. 119–124 (cited on page 12).

Salber, Daniel and Joëlle Coutaz (1993). 'Applying the wizard of oz technique to the study of multimodal systems'. In: *International Conference on Human-Computer Interaction*. Springer, pp. 219–230 (cited on page 12).

Habibovic, Azra, Jonas Andersson, Maria Nilsson, V Malmsten Lundgren, and J Nilsson (2016). 'Evaluating interactions with non-existing automated vehicles: three Wizard of Oz approaches'. In: *2016 IEEE intelligent vehicles symposium (IV)*. IEEE, pp. 32–37 (cited on page 12).

Ruijten, Peter AM, Jacques MB Terken, and Sanjeev N Chandramouli (2018). 'Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior'. In: *Multimodal Technologies and Interaction* 2.4, p. 62 (cited on page 12).

Baltodano, Sonia, Srinath Sibi, Nikolas Martelaro, Nikhil Gowda, and Wendy Ju (2015). 'The RRADS platform: a real road autonomous driving simulator'. In: *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 281–288 (cited on pages 12, 13).

Karjanto, Juffrizal, N YUSOF, Jacques Terken, Frank Delbressine, Matthias Rauterberg, and Muhammad Zahir Hassan (2018). 'Development of on-road automated vehicle simulator for motion sickness studies'. In: *Int. J. Driv. Sci* 1.1, pp. 1–12 (cited on pages 13, 14).

Detjen, Henrik, Bastian Pfleging, and Stefan Schneegass (2020). 'A wizard of oz field study to understand non-driving-related activities, trust, and acceptance of automated vehicles'. In: *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 19–29 (cited on page 13).

Sanghavi, Harsh, Yiqi Zhang, and Myounghoon Jeon (2020). 'Effects of anger and display urgency on takeover performance in semi-automated vehicles'. In: *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 48–56 (cited on page 17).

Edworthy, Judy, Sarah Loxley, and Ian Dennis (1991). 'Improving auditory warning design: Relationship between warning sound parameters and perceived urgency'. In: *Human factors* 33.2, pp. 205–231 (cited on page 17).

Hellier, Elizabeth J, Judy Edworthy, and IAN Dennis (1993). 'Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency'. In: *Human factors* 35.4, pp. 693–706 (cited on page 17).

Bellem, Hanna, Thorben Schönenberg, Josef Krems, and Michael Schrauf (Aug. 2016). 'Objective metrics of comfort: Developing a driving style for highly automated vehicles'. In: *Transportation Research Part F Traffic Psychology and Behaviour* 41, pp. 45–54. DOI: 10.1016/j.trf.2016.05.005 (cited on page 18).

Yusof, Nidzamuddin Md, Juffrizal Karjanto, Jacques Terken, Frank Delbressine, Muhammad Zahir Hassan, and Matthias Rauterberg (2016). 'The exploration of autonomous vehicle driving styles: preferred longitudinal, lateral, and vertical accelerations'. In: *Proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications*, pp. 245–252 (cited on page 19).

Müller, Andrea Isabell, Veronika Weinbeer, and Klaus Bengler (Sept. 2019). 'Using the Wizard of Oz Paradigm to Prototype Automated Vehicles: Methodological Challenges'. In: *Adjunct Proceedings - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019*. Association for Computing Machinery, Inc, pp. 181–186. DOI: 10.1145/3349263.3351526 (cited on page 19).

Karjanto, Juffrizal, Nidzamuddin Md Yusof, Jacques Terken, Frank Delbressine, Muhammad Zahir Hassan, and Matthias Rauterberg (2017). 'Simulating autonomous driving styles: Accelerations for three road profiles'. In: *MATEC web of conferences*. Vol. 90. EDP Sciences, p. 01005 (cited on page 19).

Golding, John F (2006). ‘Predicting individual differences in motion sickness susceptibility by questionnaire’. In: *Personality and Individual differences* 41.2, pp. 237–248 (cited on page 19).

Qualtrics XM // the leading experience management software (2022). URL: https://www.qualtrics.com (cited on pages 20, 21).

Merritt, Stephanie M, Heather Heimbaugh, Jennifer LaChapell, and Deborah Lee (2013). ‘I trust it, but I don’t know why: Effects of implicit attitudes toward automation on trust in an automated system’. In: *Human factors* 55.3, pp. 520–534 (cited on page 21).

Laugwitz, Bettina, Theo Held, and Martin Schrepp (2008). ‘Construction and Evaluation of a User Experience Questionnaire’. In: *HCI and Usability for Education and Work*. Ed. by Andreas Holzinger. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 63–76 (cited on pages 21, 22).

Hinderks, Andreas, Martin Schrepp, and Jörg Thomaschewski (2018). *User experience questionnaire*. URL: http://www.ueq-online.org/ (cited on page 24).

Friard, Olivier and Marco Gamba (2016). ‘BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations’. In: *Methods in ecology and evolution* 7.11, pp. 1325–1330 (cited on page 26).

O’Connor, Cliodhna and Helene Joffe (2020). ‘Intercoder reliability in qualitative research: debates and practical guidelines’. In: *International journal of qualitative methods* 19, p. 1609406919899220 (cited on page 26).

Landis, J Richard and Gary G Koch (1977). ‘The measurement of observer agreement for categorical data’. In: *biometrics*, pp. 159–174 (cited on pages 26, 27).

Lazar, Jonathan, Jinjuan Heidi Feng, and Harry Hochheiser (2017). *Research methods in human-computer interaction*. Morgan Kaufmann (cited on page 27).

StataCorp (2019). *Stata Statistical Software: Release 16* (cited on page 29).

MAXQDA (2022). URL: https://www.maxqda.com/ (cited on page 29).

Kuckartz, Udo (2014). *Qualitative text analysis: A guide to methods, practice and using software*. Sage (cited on pages 29, 30).

Hill, Clara E, Barbara J Thompson, and Elizabeth Nutt Williams (1997). ‘A guide to conducting consensual qualitative research’. In: *The counseling psychologist* 25.4, pp. 517–572 (cited on pages 29, 30).

Hill, Clara E, Sarah Knox, Barbara J Thompson, Elizabeth Nutt Williams, Shirley A Hess, and Nicholas Ladany (2005). ‘Consensual qualitative research: An update.’ In: *Journal of counseling psychology* 52.2, p. 196 (cited on pages 29, 30).

Lee, Minha, Sander Ackermans, Nena Van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn (2019). ‘Caring for Vincent: a chatbot for self-compassion’. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13 (cited on page 29).

Gold, Christian, Moritz Körber, David Lechner, and Klaus Bengler (2016). ‘Taking over control from highly automated vehicles in complex traffic situations: The role of traffic density’. In: *Human factors* 58.4, pp. 642–652 (cited on page 42).

Gold, Christian, Daniel Damböck, Klaus Bengler, and Lutz Lorenz (2013). ‘Partially automated driving as a fallback level of high automation’. In: *6. Tagung Fahrerassistenzsysteme* (cited on page 43).

Kerschbaum, Philipp, Lutz Lorenz, and Klaus Bengler (2015). ‘A transforming steering wheel for highly automated cars’. In: *2015 ieee intelligent vehicles symposium (iv)*. IEEE, pp. 1287–1292 (cited on page 43).

Sagberg, Fridulv, Selpi, Giulio Francesco Bianchi Piccinini, and Johan Engström (2015). ‘A review of research on driving styles and road safety’. In: *Human factors* 57.7, pp. 1248–1275 (cited on page 44).

Michon, John A (1985). 'A critical view of driver behavior models: what do we know, what should we do?' In: *Human behavior and traffic safety*, pp. 485–524 (cited on page 44).

Bellem, Hanna, Barbara Thiel, Michael Schrauf, and Josef F Krems (2018). 'Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits'. In: *Transportation research part F: traffic psychology and behaviour* 55, pp. 90–100 (cited on page 44).

Dettmann, Andre, Franziska Hartwich, Patrick Roßner, Matthias Beggiato, Konstantin Felbel, Josef Krems, and Angelika C Bullinger (2021). 'Comfort or not? Automated driving style and user characteristics causing human discomfort in automated driving'. In: *International Journal of Human–Computer Interaction* 37.4, pp. 331–339 (cited on page 44).