CHALMERS
UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG

User Study for In-Vehicle Displays

Exploring Driver Behaviour and the Influence of HUD During Traffic Navigation Using Eye-Tracking

Master's thesis in Computer science and engineering

Eva Le, Xiaoyan Shi

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2025

Master's thesis 2025

User Study for In-Vehicle Displays

Exploring Driver Behaviour and the Influence of HUD During Traffic Navigation Using Eye-Tracking

Eva Le
Xiaoyan Shi

UNIVERSITY OF
GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2025

User Study for In-Vehicle Displays
Exploring Driver Behaviour and the Influence of HUD During Traffic Navigation Using Eye-Tracking
Eva Le, Xiaoyan Shi

Supervisor: Ilaria Torre, Computer Science and Engineering
Advisor: David Hermann, Volvo Car Corporation
Examiner: Michael Heron, Computer Science and Engineering

Master's Thesis 2025
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Illustrated system setup including in-vehicle displays, webcams, and eye-tracking cameras.

Gothenburg, Sweden 2025

User Study for In-Vehicle Displays
Exploring Driver Behaviour and the Influence of HUD During Traffic Navigation Using Eye-Tracking
Eva Le, Xiaoyan Shi
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

This master's thesis investigates how Head-Up Displays (HUDs) influence driver behaviour in real-world traffic, with a focus on attention, distraction, and cognitive load. The study examines a conventional HUD installed in a Volvo XC60 using eye-tracking technology. A mixed-methods approach was employed, combining quantitative analysis of gaze dwell time and NASA-TLX scores with qualitative insights from thematic analysis of participant feedback. The research compares driver behaviours in HUD and non-HUD conditions. Findings indicate that while the presence of a HUD reduces visual attention to Head-Down Displays (HDDs) and is generally perceived as enhancing safety, usability concerns were raised—especially regarding the navigation interface. Although the HUD did not significantly affect the overall cognitive load, some participants reported visual discomfort. The study highlights both the potential benefits and challenges of HUD implementation and calls for further research with more diverse participants and multimodal HUD designs to ensure safe and user-friendly automotive interfaces.

Keywords: Head-Up Display, Eye-Tracking, Human-Computer Interaction, Interaction Design, User Experience, Usability, Automotive, Driver Behaviour, Attention, Distraction, Cognitive Load, Safety.

Acknowledgements

We would like to express our deepest gratitude to our supervisor, David Hermann from Volvo Car Corporation (VCC). His generous support and extensive knowledge of display systems greatly guided our work throughout this project. His mentorship has been invaluable to the understanding and development of this thesis.

We also sincerely thank our academic supervisor, Ilaria Torre, at Chalmers University of Technology, for providing us with insightful feedback on research design and statistical analysis. Your guidance has been essential in shaping the direction and quality of this thesis.

We gratefully acknowledge the financial support from the VINNOVA project "SCREENS II" (Project number: Dnr 2023-00755). Thank you to all members of the "SCREENS II" project, especially Kjell Brunnström, Jonas Andersson, Daban Rizgary from RISE, and Henrik Lind, Emil Rolander from Smart Eye. We appreciate the engaging meetings and thoughtful discussions that helped us improve and reflect on our work from different perspectives.

We are also thankful for the support from the Vehicle Engineer, Advanced Engineering Workshop, and Instrument Support departments at VCC, whose assistance was essential in preparing the equipment used in our study.

We would like to extend our appreciation to Smart Eye and iMotions for their technical support in the eye-tracking setup and integration.

We would like to thank all the participants who contributed their time and effort to our user study. Your involvement made this research possible.

Lastly, we are grateful for the resources and support we have received from Chalmers University of Technology. The guidance and encouragement from the teachers and faculty members in the Interaction Design and Technology division have contributed greatly to our academic development.

Eva Le, Xiaoyan Shi, Gothenburg, 2025-10-28

Contents

List of Figures

List of Tables

List of Abbreviations

ADAS - Advanced Driver Assistance Systems
AOI - Area Of Interest
AR - Augmented Reality
AR-HUD - Augmented Reality Head-Up Display
CLT - Cognitive Load Theory
CSD - Center Stack Display
DIM - Drivers Information Module
ECG - Electrocardiogram
EDA - Electrodermal Activity
EEG - Electroencephalography
FoV - Field of View
GDPR - General Data Protection Regulation
GPS - Global Positioning System
HDD - Head-Down Display
HMI - Human-Machine Interaction
HRGR - Heart Rate Growth Rate
HRV - Heart Rate Variability
HUD - Head-Up Display
IR - Infrared
IVIS - In-Vehicle Information Systems
IxD - Interaction Design
NASA-TLX - National Aeronautics and Space Administration-Task Load Index
NHTSA - National Highway Traffic Safety Administration

OCEs - Online Controlled Experiments

RISE - Research Institute of Sweden

SA - Situation Awareness

SD - Standard Deviation

SotA - State-of-the-Art

SWAT - Subjective Workload Assessment Technique

TA - Thematic Analysis

UI - User Interface

UX - User Experience

VCC - Volvo Car Corporation

WCS - World Coordinate System

1

Introduction

This master's thesis is conducted within the framework of the "SCREENS II" research project (RISE, 2023), coordinated by the Research Institutes of Sweden (RISE), in collaboration with Volvo Car Corporation (VCC), AB Volvo, Scania, and Smart Eye. While "SCREENS II" addresses the broader challenges of increasing automation in driving environments to enhance road safety, this thesis focuses on Interaction Design (IxD) and Human-Computer Interaction (HCI) in automotive interfaces, specifically in-vehicle displays. The study investigates how the presence or absence of a Head-Up Display (HUD) influences driver behaviour, using a conventional HUD installed in a Volvo XC60 and employing eye-tracking technology to collect behavioural data.

1.1 Context

The integration of advanced technologies in vehicles is transforming how drivers interact with their vehicles and the road environment. In the automotive industry, manufacturers are continuously developing Advanced Driver Assistance Systems (ADAS) to enhance vehicle safety and driving comfort (Gulino et al., 2025; Langlois, 2013; Neumann, 2024; Nidamanuri et al., 2021). HUDs are one such technology integrated with ADAS, designed to improve driving safety and efficiency by utilising information from both the vehicle and its surroundings (Maroto et al., 2018). As an interaction-based visual assistance technology, the HUD projects driving information onto the windshield, enabling the driver to see it directly within the physical driving scene (Khedkar et al., 2015; Maroto et al., 2018; X. Wang and Qin, 2014), as illustrated in Figure 1.1. The area marked in blue is referred to as the eye box, which defines the angle within which the virtual image is visible. Although HUDs can vary in the design of their components across manufacturers, their primary goal remains the same: enhancing driver comfort and safety (Maroto et al., 2018).

In recent years, the development of HUDs has advanced alongside emerging technologies, with Augmented Reality Heads-Up Displays (AR-HUDs) drawing significant attention in User Experience (UX) research (Zhou et al., 2024). However, despite the promise of AR-HUDs, challenges such as increased production costs (Zhou et al., 2024), design and implementation complexities (Guo et al., 2020; Zhou et al., 2024), and driver adaptation (Xia et al., 2023) limit their widespread adoption. This study focuses on conventional HUDs instead of AR-HUDs due to greater standardisation, lower variability in implementation, and clearer study parameters for evaluating driver behaviour using eye-tracking technology.

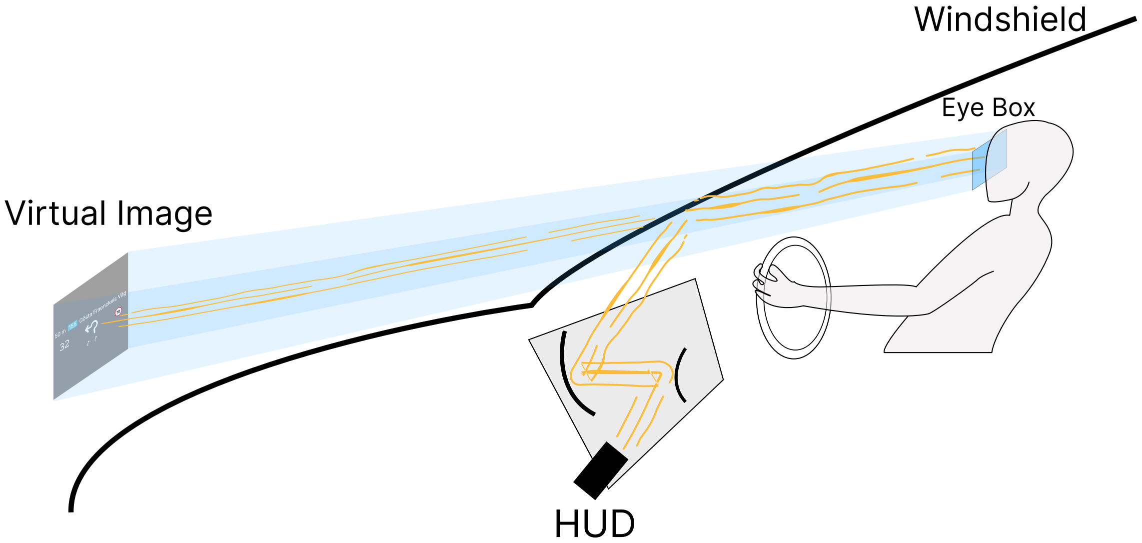


Figure 1.1: The HUD image is projected on the windshield creating a virtual image. Inspired by Maroto et al. (2018). Illustrated by Eva Le.

Regardless of the growing implementation of HUDs in vehicles, it remains unclear how they influence driver behaviour in real traffic, especially in terms of attention, distraction, and cognitive load, when compared to Head-Down Displays (HDDs). Although prior studies on test tracks (AblaSSmeier et al., 2007) and in simulators (Cheng et al., 2023; Liu and Wen, 2004) have shown that HUDs can significantly influence driver behaviour, their impact in real-world driving scenarios remains unclear. In the context of this study, HDDs consist of Driver Information Module (DIM) and Center Stack Display (CSD), as shown in Figure 1.2. Compared to HDDs, HUDs are designed to provide driving information directly within the drivers line of sight, reducing the need to look away from the road, which is generally considered more advantageous in terms of situational awareness (SA) (Thalen, 2006). However, studies indicate that first-time users often experience discomfort while using HUDs, as they need time to familiarise themselves with the tool and integrate it into their driving routine (Liu and Wen, 2004). Additionally, depending on the complexity and type of information displayed on the HUD, drivers might experience increased cognitive load associated with switching their attention from different display sources (Cheng et al., 2023; Maroto et al., 2018).

In addition, previous research shows that a driver experiencing fatigue or distraction has contributed to as much as 20% of accidents on monotonous roads, and these numbers could have been reduced if driver behaviour had been effectively monitored and predicted (Kang, 2013). The increasing use of In-Vehicle Information Systems (IVIS) and introduction of in-vehicle technology, such as navigation systems and mobile phones, have led to greater driver distractions, both visual and cognitive (Liang, 2009).

To evaluate the effectiveness of HUD applications, eye-tracking technology offers

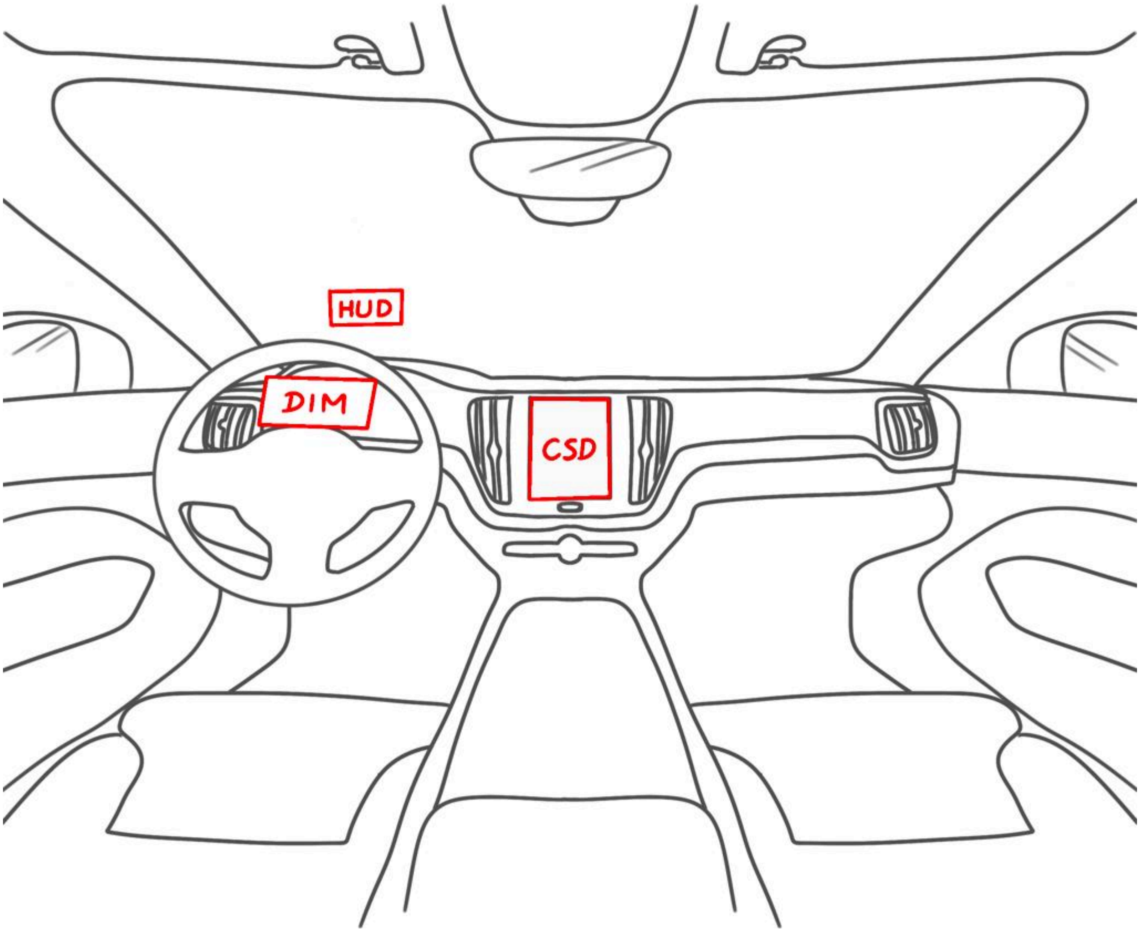


Figure 1.2: The placement of different displays (HUD, DIM, CSD) in vehicle.

valuable insights, as it is a powerful tool for analysing human vision and perception (Krassanakis et al., 2014). By capturing metrics such as gaze patterns, fixation durations, and saccades, eye-tracking offers a deeper understanding of driver attention (Hurtado and Chiasson, 2016), behaviour, and decision-making processes (You et al., 2017). This makes eye-tracking an essential tool for assessing how drivers interact with various in-vehicle interfaces. However, unlike screen-based eye-tracking studies—where high data quality can be maintained as long as critical distances and angles are kept constant and calibration follows a standard procedure (Vehlen et al., 2021)—conducting eye-tracking in real-road traffic presents greater challenges. The accuracy of eye-tracking data can be influenced by varying conditions, including lighting changes, weather fluctuations, and the dynamic nature of driving environments. These factors may lead to issues such as tracking loss, increased end-to-end latency, or sensitivity to participants' movements (Orquin and Holmqvist, 2018). As a result, effectively analysing driver behaviour and identifying differences between display types can become more difficult. The need to understand these dynamics is critical in enhancing automotive interface design, ensuring both safety and usability.

1.2 Stakeholder

The primary stakeholders of this project begin with the master's students, Eve Le and Xiaoyan Shi, who are responsible for conducting the research and authoring the thesis, encompassing the design, execution, and analysis of the study. Following this, Chalmers University of Technology acts as the academic supervisor and examiner, providing essential guidance and oversight throughout the research process. Another key stakeholder is the "SCREENS II" research initiative, which offers the broader research context and facilitates collaboration with industry partners who contribute valuable expertise and technology to the investigation of automotive interfaces and driver safety. Finally, the stakeholders extend to the drivers who interact with these systems, as well as automotive engineers, designers, and researchers who will benefit from the findings to advance the design and implementation of vehicle display technologies.

1.3 Goals and Research Question

HUDs are designed to enhance SA by keeping critical driving information within the drivers line of sight. However, they may also introduce certain visual and attentional challenges that could compromise driving safety.

The primary aim of this project is to explore how the digitalisation of the drivers environment impacts driver behaviour. Several important factors are considered when it comes to ensuring safety and usability in HUD applications. These include attention, distraction, and cognitive load. To address these factors, eye-tracking is employed as a primary tool for analysing driver behaviour under two conditions: with and without a HUD. Based on this, the following research question is proposed:

What are some of the most significant differences in driver behaviours when it comes to HUD and non-HUD conditions?

2

Background

This chapter reviews key background knowledge and related work on Human-Machine Interaction (HMI) in the automotive industry, focusing on in-vehicle displays and their impact on driver safety and behaviour. It covers industry practice, different types of in-vehicle displays, driver distraction, cognitive effects such as cognitive capture, and the role of eye-tracking and cognitive load in evaluating HMI systems.

2.1 Human-Machine Interaction in Automotive Industry

Over 130 years of evolution, the automotive industry has increasingly emphasised software and UX, with HMI playing a pivotal role in modern automotive product design (Gong, 2025; J. Ma and Gong, 2024; Niu and Tang, 2024). Automotive HMI systems are inherently more complex than those in smartphones or household appliances, as they encompass multifaceted design objectives rather than simply straightforward engineering solutions (J. Ma & Gong, 2024). As summarised by J. Ma and Gong (2024), the design objectives in automotive HMI systems can be divided into three levels of increasing abstraction: *functionality*, *usability*, and *imaginability* (see Figure 2.1).

- *Functionality* represents the functions of the automotive HMI system that can be achieved based on the integration of the software and hardware.
- *Usability* typically involves the interaction between users, products, tasks, and the environment. In the context of automotive HMI, usability is assessed using multiple indexes as shown in Figure 2.1. Among these, achieving both safety and efficiency is particularly challenging compared to other industries. Unlike other HMI evaluation criteria, safety is a distinctive and essential metric specific to automotive HMI. Efficiency refers to the relevant resources (e.g., task time, visual attention, and physical movements) utilised by the user and HMI system when completing a specified interaction task. Unlike in consumer electronics like smartphones, where users can focus entirely on the interface, automotive HMI design must prioritise driving safety as drivers are not supposed to fully focus on interacting with the HMI system. As a result, enhancing efficiency without compromising safety is a sig-

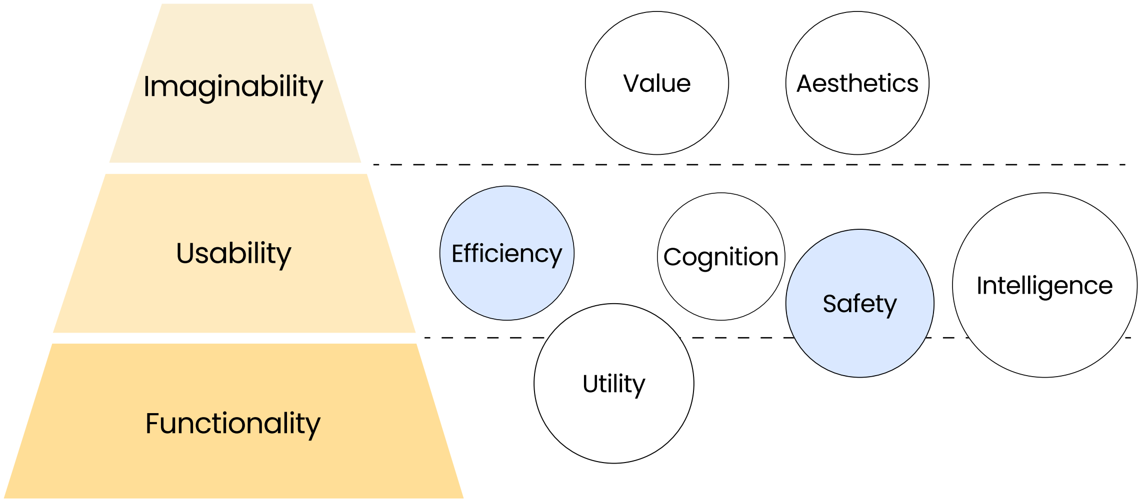


Figure 2.1: Automotive HMI objectives and corresponding evaluation indexes. Inspired by J. Ma and Gong (2024). Illustrated by Eva Le

nificantly more difficult task for automotive interfaces (J. Ma and Gong, 2024).

- *Imaginability* describes the extent to which users can envision how a product will be used or what it offers, beyond just its basic functionality and usability. Automotive HMI systems have a wide range of designs and features, which can vary significantly across different brands of vehicles. Due to this diversity, it is nearly impossible for users to fully and accurately understand a vehicle's HMI system before purchasing it. Moreover, even after owning the vehicle for several years, many users tend to utilise only a limited subset of the available HMI functions.

While both the number and intricacy of HMI elements continue to grow, automotive HMI not only facilitates interaction between the driver and vehicle but also connects to the external environment. Consequently, its design approach is shifting from a technology-driven focus to a user-centred paradigm (Yardm and Pedgley, 2023; X. Zhang et al., 2022).

VCC has consistently prioritised safety in automotive design and actively invests in advanced technologies aimed at enhancing driver awareness and minimising risk. According to Volvo's safety strategy (Eugensson and Ivarsson, 2022), improving ergonomics and human-machine interfaces is essential for reducing driver distraction and maintaining attention during the driving task. Enhancements to navigation systems also form part of this strategy, aiming to alleviate driver stress and cognitive burden. Effective HMI design can contribute to reduced cognitive load, improved user comprehension, and increased comfort and confidence during vehicle operation.

2.2 In-Vehicle Displays

In-vehicle displays play a crucial role in modern automotive HMI, providing drivers with essential information while ensuring minimal distraction. These displays have evolved significantly to accommodate advancements in technology, offering a wide range of functionalities that enhance both safety and UX. The following sections outline the key types of in-vehicle displays, describing their characteristics and applications.

2.2.1 Development and Application of HUDs in the Automotive Industry

HUD technology was initially developed in 1940s for military aviation, where pilots needed real-time information without looking away from their flight path (Cameron, 2015; Okabayashi et al., 1989; Ward and Parkes, 1994). The world's first application of a HUD in a production vehicle was introduced with the 1988 Nissan Silvia model in the Japanese Domestic Market (Okabayashi et al., 1989). In the same year, General Motors also implemented a HUD in their Oldsmobile Cutlass Supreme (Weihrauch et al., 1989). Since the early 2000s, automotive HUD technology has advanced significantly (Jin et al., 2016). Cadillac introduced thermal imaging to enhance visibility, while General Motors developed the first four-colour display system, transitioning from monochrome to colour projection. Pioneer Corporation later introduced an Augmented Reality (AR) navigation system that projected animations onto the driver's visor. Following these developments, manufacturers such as BMW and Volkswagen developed their own HUD systems. BMW created technology capable of reading road signs and projecting information, like temporary speed limits and hazard alerts, directly onto the windshield. Over time, the automotive industry adopted HUDs to improve SA, minimise distractions, and enhance road safety (Betancur et al., 2018).

Automotive HUDs rely on key optical properties to function effectively. As illustrated in Figure 1.1, these key optical properties include:

- **Virtual Image**: The HUD projects an image through the windshield that appears to float at a distance in front of the driver, outside the vehicle. In conventional HUDs, the symbols in the virtual image remain fixed, while in AR-HUDs, the symbols dynamically adapt to real-world objects.
- **Field of View (FoV)**: The extent of the displayed information visible within the drivers direct line of sight.
- **Projection Distance**: The distance at which the virtual image is perceived, typically ranging from 2 to 4 metres in conventional HUDs, and from 8 to 10 metres in AR-HUDs (see Figure 2.2).

Modern vehicles incorporate HUDs as either conventional HUDs or AR-HUDs. Conventional HUDs typically display speed, navigation arrows, and warning indicators. Figure 2.3 illustrates example use cases of conventional HUDs, with the *Volvo*

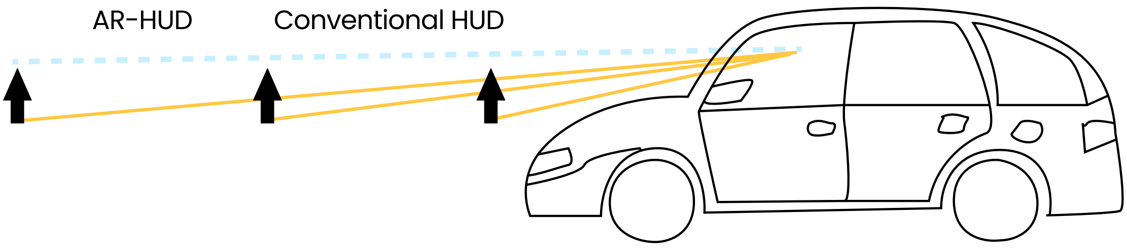


Figure 2.2: Projection distance of conventional HUD and AR-HUD.

XC60's HUD displaying different types of information. While AR-HUDs offer a longer projection distance, allowing them to overlay real-world objects with contextual digital information, Figure 2.4 illustrates how the AR-HUD in the *Mercedes-Benz S-Class* provides a wider FoV compared to conventional HUDs.

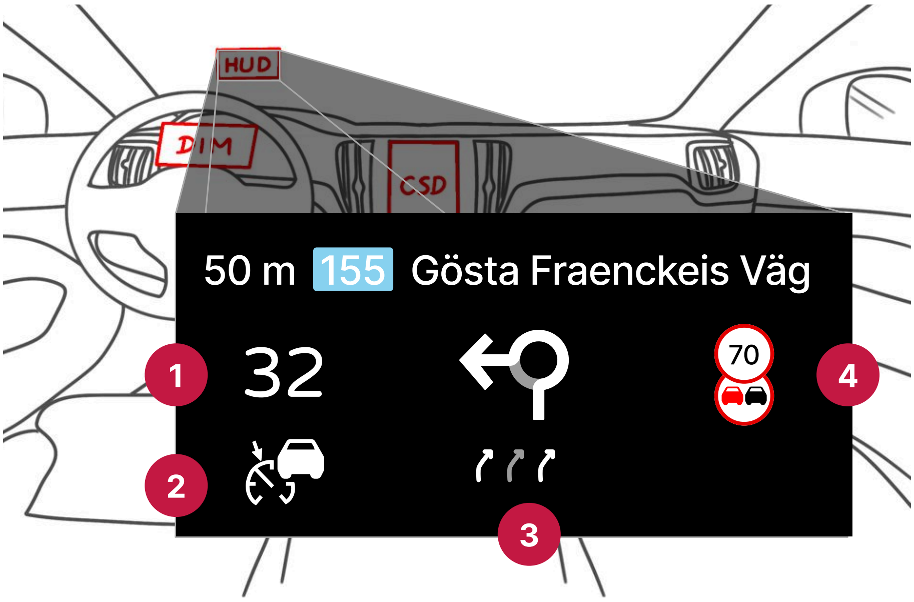


Figure 2.3: The conventional HUD in the Volvo XC60 showing 1) Speed, 2) Cruise Control, 3) Navigation, and 4) Road Signs.

Figure 2.4: AR-HUD in the Mercedes-Benz S-Class, providing a wide FoV and dynamic symbols.

2.2.2 Head-Down Displays

HDDs, unlike HUDs positioned within the drivers FoV, require drivers to look down, diverting their gaze from the road. Liu and Wen (2004) emphasise the necessity for drivers to look away to read the displayed information, while Smith et al. (2016) highlight that HDDs make drivers alternate their focus between the road and the dashboard or central console.

Different automotive manufacturers use various terms to describe HDDs. For instance, VCC specifically defines HDDs as the DIM and the CSD, with their positions illustrated in Figure 1.2.

DIM serves as the primary interface between the driver and vehicle systems, typically located behind the steering wheel. It presents real-time data, including speed, fuel level, engine status, and warning indicators, ensuring drivers remain informed about vehicle performance. Modern DIMs vary from traditional analogue instrument clusters to fully digital displays, which provide customisable interfaces and adaptive layouts. Some advanced DIMs incorporate secondary information, such as navigation directions and ADAS alerts.

CSD is positioned in the middle of the dashboard and serves as the central control interface for infotainment, climate control, and vehicle settings. These displays range from touchscreen-based systems to multi-layered interfaces incorporating physical controls for enhanced usability. CSDs provide access to navigation, media playback, phone connectivity, and vehicle diagnostics, often integrating voice commands and gesture recognition for improved interaction.

Other manufacturers have HDD functions comparable to those of VCC. For example, Mercedes-Benz uses *Digital Instrument Cluster* for speed and status information, along with a *MBUX Central Display* located on the centre stack.

2.3 Driver Behaviour vs Driving Behaviour

It is essential to distinguish between driver behaviour and driving behaviour, as these two categories capture different but interrelated aspects of the driving experience (Kang, 2013; Singh and Kathuria, 2021). Table 2.1 presents a summary of their features and common examples.

Category	**Features**	**Examples**
Driver Behaviour	Visual Feature	Eye closure, blinking, yawning, head pose, facial expressions
	Non-Visual Feature	Heart rate, pulse rate, brain activity
Driving Behaviour	Vehicle-based Feature	Steering, accelerating, braking

Table 2.1: Summary of various features of driver and driving behaviour with common examples.

Driver behaviour includes both visual and non-visual features. Visual features include eye closure, blinking, yawning, head pose, and facial expressions, all of which can often be detected using in-vehicle monitoring or eye-tracking systems (Hayley et al., 2021; Victor et al., 2005; You et al., 2017). Non-visual features involve physiological indicators such as heart rate, pulse rate, and brain activity, which are often used to assess mental workload, fatigue, drowsiness, or emotional reactions (Lin et al., 2007; Yang et al., 2010).

In contrast, *driving behaviour* refers to driver's actions related to vehicle control. This includes lane-keeping performance, speed variations, steering movements, and pressure applied to the acceleration pedal. These behaviours are often monitored using vehicle telemetry data (Iatropoulos et al., 2024; Y. Zhang et al., 2021), or assessed through performance-based metrics such as reaction time and hazard response (Cheng et al., 2023; Liu and Wen, 2004). Such measures provide objective indicators of driving performance and are essential for evaluating the effectiveness of in-vehicle systems on real-road safety.

2.4 Safety in Driving

Safety, in the context of automotive HMI, refers to the system's ability to minimise driver distraction and support safe driving while the driver engages in HMI-related tasks during driving (J. Ma and Gong, 2024). Driver inattention and driver distraction have been widely recognised as critical factors in road safety, and are among the leading contributors to traffic crashes (Beanland et al., 2013; Stutts et al., 2001).

2.4.1 Primary Task vs Secondary Task

Understanding the impact of HMI systems on driver behaviour, especially regarding attention, distraction, and cognitive load, makes it necessary to identify and define primary and secondary tasks in driving. The primary task refers to driving-related activities, such as steering and monitoring road hazards, while the secondary task involves non-driving activities (Cheng et al., 2023; Gong, 2025; Y. Ma et al., 2018).

As discussed in Section 2.1, safety is a key criterion for evaluating automotive HMI systems. This emphasis arises from the fact that many commonly used automotive HMI functions involve secondary tasks that drivers must perform concurrently with driving, without significantly compromising road safety (J. Ma and Gong, 2024).

According to the automotive HMI human factors model proposed by Gong (2025), primary and secondary tasks occur simultaneously and influence each other. Importantly, the execution of secondary tasks should not introduce significant driver distraction, as excessive cognitive load can compromise driving safety. This concern is echoed by Kang (2013), who highlights that extensive use in IVIS, such as navigation systems and mobile phones, can lead to visual and cognitive distractions for the driver. Similarly, Gong (2025) emphasises the limited nature of a driver's attentional capacity, highlighting that engaging in secondary tasks can lead to decreased driving performance due to distraction.

In a broad context, secondary tasks include activities such as engaging in conversations with passengers, talking on a cell phone, eating, smoking, and so on. However, these fall outside the scope of this discussion as they are not directly related to automotive HMI. As noted by National Highway Traffic Safety Administration (NHTSA) (2012), secondary task interactions may pertain to driver comfort, convenience, communication, entertainment, information access, or navigation.

2.4.2 Types and Effects of Driver Distraction

Evaluating distraction as a component of driver behaviour requires a brief conceptual overview. Since secondary tasks are a common source of distraction, it is important to classify the types of distraction they may cause. According to NHTSA (2010), driver distraction can be categorised into three main types: *visual distraction*, *manual distraction*, and *cognitive distraction*.

- Visual distraction: Tasks that require the driver to look away from the roadway to visually obtain information.
- Manual distraction: Tasks that require the driver to take a hand off the steering wheel and manipulate a device.
- Cognitive distraction: Tasks that are defined as the mental workload associated with a task that involves thinking about something other than the driving task.

2.4.3 Cognitive Capture

Cognitive capture is a hypothesis proposed by Gish and Staplin (1995) to describe the inefficient allocation of attention between a HUD and the primary driving task. According to this hypothesis, the HUD may draw the driver's attention away from the road, resulting in delayed responses and missed external cues. Notably, the time required to redirect attention from the HUD back to the road exceeds the time initially needed to attend to the HUD, thereby posing a potential safety risk.

In Russell et al. (2023), the authors examined how cognitive capture influences drivers' responses to traffic scenarios that required immediate action. Their findings suggest that distraction caused by tasks performed using a HUD can have negative effects comparable to those associated with HDDs.

2.4.4 HUD Superimposition—"Masking Effect"

While HUDs are intended to reduce eyes-off-road time and enhance SA, several studies have raised concerns about the potential for HUD content to mask critical visual information in the real world. This phenomenon is often referred to as the "HUD masking effect" (Kiefer and Gellatly, 1996; Oh et al., 2016; Tangmanee and Teeravarunyou, 2012). This effect occurs when virtual images projected by the HUD are superimposed onto real-world objects, potentially obstructing the driver's ability to detect pedestrians, vehicles, or traffic signs. Such visual interference may affect both the visibility of HUD icons or text, increase driver's display glance duration, and thereby contribute to driver distraction and reduce driving safety.

Previous research has identified multiple factors contributing to the HUD superimposition issue. For instance, Okabayashi et al. (1991) found that increased luminance of the HUD image can reduce a driver's ability to recognise objects in the external environment. In addition, Oh et al. (2016) demonstrated that higher levels of superimposition negatively influenced glance duration, regardless of driver age. Furthermore, Tangmanee and Teeravarunyou (2012) reported that while the shape of HUD graphics had minimal impact, the spatial placement of visual elements (e.g., arrows in navigation) significantly affected driver performance.

Notably, these studies were conducted in controlled laboratory settings and may not fully represent the real-world environment. In contrast, Kiefer and Gellatly (1996) examined the effects of HUD superimposition on a closed test track. The study concluded that when the HUD is configured at an appropriate angle, its benefits in reducing visual distraction outweigh the relatively minor risks associated with superimposition. This is because the projected virtual image is positioned below the driver's primary visual field and does not interfere with critical forward-scene elements.

Nevertheless, it is important to note that most existing studies on HUD superimposition have been conducted in simulated or controlled environments, which may not fully capture the complexity and variability of real-world driving scenarios. Further research in naturalistic driving contexts is needed to validate these findings and inform the development of safer HUD designs.

2.5 Eye-Tracking in Behavioural Study

Eye-tracking is a valuable experimental method for recording eye movements and gaze locations over time and across tasks. Its use has increased significantly in recent years. B. T. Carter and Luke (2020) showed that publications involving eye-tracking rose sharply between 1968 and 2018. The study also noted that many books and academic papers have explored the application of eye-tracking in various fields, which include reading, memory, decision-making, UX design and usability research. The origins of eye-tracking can be traced back to Bell (1823), who described the physiological connection between eye movements and the nervous system, linking ocular motion to neurological and cognitive processes.

Eye-tracking is widely used as a primary research tool in behavioural studies and spans a broad range of topics. For instance, researchers have examined visual attention in healthcare settings (Hofmaenner et al., 2021) and social interactions (Vehlen et al., 2021), and used it to assess cognitive load (Cheng et al., 2023; Zagermann et al., 2016), marketing, neuroscience, and HMI (Blascheck et al., 2014).

In the context of HUD research, eye-tracking is commonly employed to analyse driver behaviour. The types of eye-tracking devices used in these studies vary. Some rely on standard cameras to record eye movements for later analysis (Russell et al., 2023). Others use eye-tracking glasses (Bitkina et al., 2021; Oh et al., 2016), head-mounted devices (AblaSSmeier et al., 2007), or stand-alone systems that can collect a range of gaze metrics without requiring participants to wear any equipment (Dombrovskis, 2010).

Although naturalistic driving studies with eye-tracking present challenges, as discussed in Section 1.1, they remain a valuable method for unobtrusive data collection (Singh and Kathuria, 2021). In particular, stand-alone eye-tracking systems are well-suited for capturing authentic driver behaviour without interfering with the driving task. Wege et al. (2013) highlighted the advantages of real-world testing, noting its strong construct and face validity due to the complexity and realism of the environment. However, in their study, eye movements were recorded using in-cabin and face-view cameras and manually coded frame-by-frame.

2.6 Cognitive Load Measurement in Driving

HCI often investigates the cognitive load placed on individuals while interacting with systems in dynamic environments such as driving (Demberg et al., 2013; Kosch et al., 2023; Stojmenova and Sodnik, 2015; von Janczewski et al., 2022). Section 3.6.1 presents various methods for measuring cognitive load in HCI. Among these, NASA-Task Load Index (NASA-TLX) has been widely used in experiments to estimate drivers cognitive load, providing a reliable basis for comparing different experimental conditions and tasks. Such applications involve navigation conditions study (Harms and Patten, 2003) and interface type evaluation (Jakus et al., 2015).

Galy et al. (2018) proposed a model describing the relationship between Cognitive Load Theory (CLT) categories, the different workload dimensions measured by NASA-TLX, and driver's mental load factors. Driver's mental load factors include driving situation, driving experience, and driver's functional state. CLT and NASA-TLX are explained in greater detail in Section 3.6. Figure 2.5 presents an adoption of the model. The figure illustrates that self-reported human cognitive resources are divided into three parts: obligatory, available and remaining. These resources are dependent on various factors such as the CLT categories and driver's functional state (e.g., alertness and tension).

Starting with the *obligatorily used* cognitive resource. The resource is affected by both the intrinsic and extraneous mental load. These are in turn affected by the complexity of the driving situation, which can be subjectively determined by the mental, physical and temporal demand dimensions of the NASA-TLX. The *available* cognitive resource depends on the driver's functional state, which can be measured by the frustration dimension. The *remaining* cognitive resources influence the germane cognitive load, which in turn affects task efficiency and impacts SA. More information about SA is provided in Section 3.5. The performance subscale of NASA-TLX can be used as a subjective measure of SA. The driver's experience affects both SA and germane mental load. For experienced drivers who have remaining cognitive resources, these resources contribute to germane mental load, resulting in enhanced driving performance and SA.

However, Galy et al. (2018) argues that effort, a dimension in NASA-TLX, is a consequence of task demand and should therefore not be viewed as equal to other

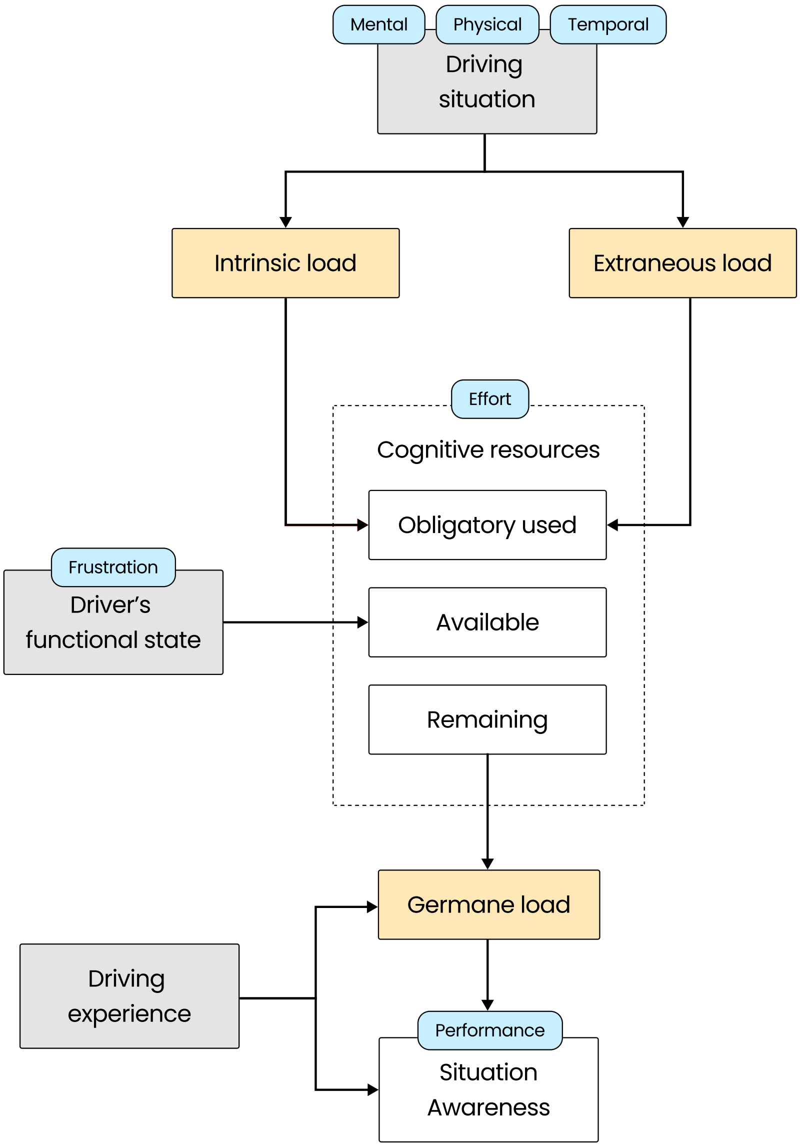


Figure 2.5: Graphical representations of proposed relationships between cognitive load factors, mental load factors, and each dimension of NASA-TLX. Inspired by Galy et al. (2018). Illustrated by Eva Le.

dimensions of workload. According to the author, effort is the difference between obligatory used and available cognitive resources. When available resources are high compared to those required, the perceived effort is low. Conversely, when available resources are limited and most of them are consumed, the perceived effort is high. Thus, the author suggests that the different dimensions of NASA-TLX workload are all sensitive to distinct factors, and relying entirely on the overall workload score can be misleading, as the load can refer to different CLT categories.

3

Theory

This chapter outlines the theoretical foundation and methodological framework guiding this study. It introduces key concepts from Interaction Design (IxD), UX research, with a particular focus on eye-tracking as a method for examining user behaviour. The chapter also discusses relevant theoretical frameworks, including human information processing, cognitive load, and SA. It further details the approach to design, conduct, and analyse in UX research.

3.1 Interaction Design

IxD is a multidisciplinary field that focuses on shaping digital and physical interactions between users and technology. As shown in Figure 3.1, IxD sits at the intersection of UX, HCI, and Human Factors, drawing from diverse disciplines such as psychology, cognitive science, and ergonomics to enhance user engagement. The study of usability, a key factor in ensuring a quality UX, remains a crucial aspect of the broad and multidisciplinary field of HCI (Hartson and Pyla, 2012).

3.2 User Experience Study

According to Mahlke (2005), it is important to take into account various aspects in studying UX. The basic UX process model presented in the work shows how different factors—such as system qualities and user emotions—contribute to users' judgements and behavioural consequences. This framework is particularly relevant in evaluating how interactive technologies influence UX and decision-making. Similarly, Baxter et al. (2015) explain that "UX research is about understanding people, the domain, and technology." They emphasise that UX methods can be broadly applied to study human behaviour, perceptions, needs, and concerns in various contexts involving technology.

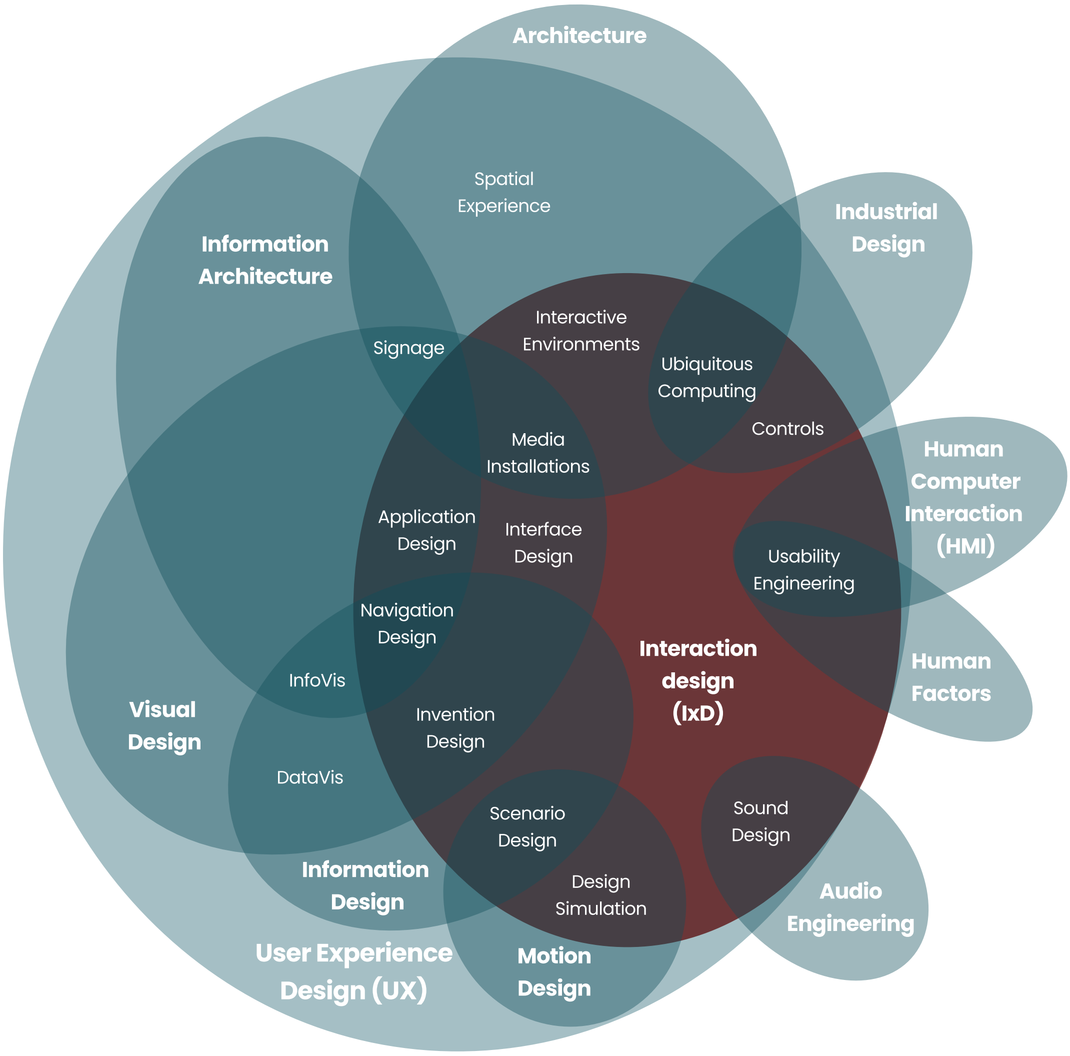


Figure 3.1: The variety of disciplinary knowledge and skills involved in contemporary design of IxD. Inspired by Carroll (2009). Illustrated by Eva Le.

3.3 Eye-Tracking Technology

This section describes the functionality of eye-tracking technology, the types of eye-tracking data it generates (such as gaze points, fixations, and saccades), and the methods used to analyse this data to assess drivers visual attention and cognitive load.

3.3.1 Functionality of Eye-Tracking

Eye-tracking has become a popular and promising method in HCI and UX research, offering valuable insights into users' visual attention and cognitive load (Punde et al., 2017), as well as supporting usability evaluation in User Interface (UI) (B. Albert and Tullis, 2013; Jacob and Karn, 2003). The general explanation of how an eye-tracking camera works is by measuring the angle between pupil centre and corneal reflections to get gaze direction (B. T. Carter and Luke, 2020). By directing a camera together with a light source towards the eyes, it is possible to capture pupil position and corneal reflection as seen in Figure 3.2. The light used is typically infrared (IR) light.

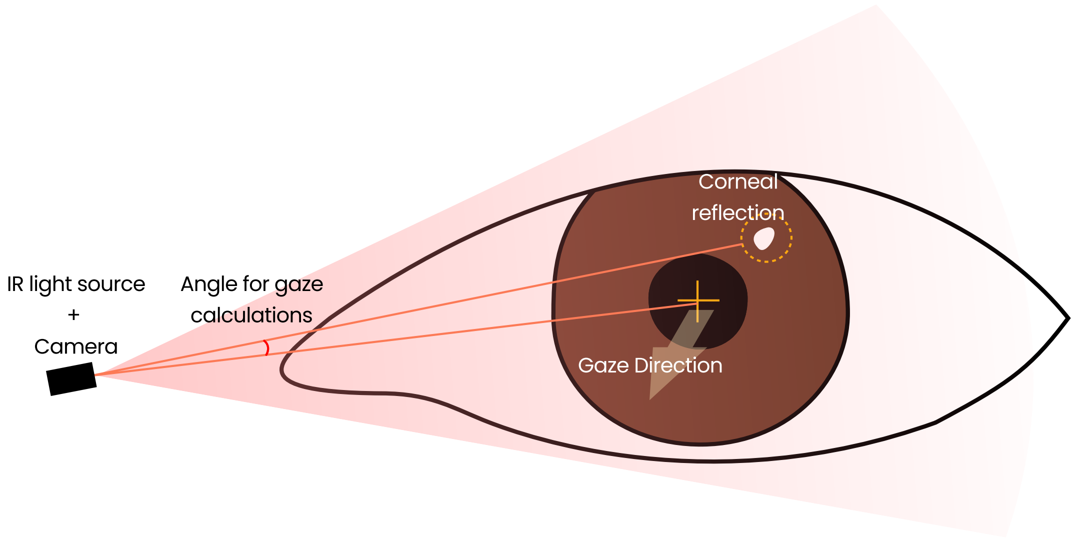


Figure 3.2: Eye-tacking principle. Inspired by iMotions (2025). Illustrated by Eva Le.

Eye-tracking systems record raw eye movement data, or gaze point, which is then aggregated into fixations and saccades for analysis (Blascheck et al., 2014). *Fixation* is the aggregated gaze points based on a specified timespan (usually 200 to 300 ms) and area (typically 20 to 50 pixels). *Saccade* refers to the rapid eye movement from one fixation to another. This is the fastest movement performed by the human body which lasts around 30 to 80 ms. Vision is essentially suppressed during saccades, even though this is not consciously perceived (B. Albert and Tullis, 2013). Figure 3.3 illustrates the relationship between gaze, fixation and saccade.

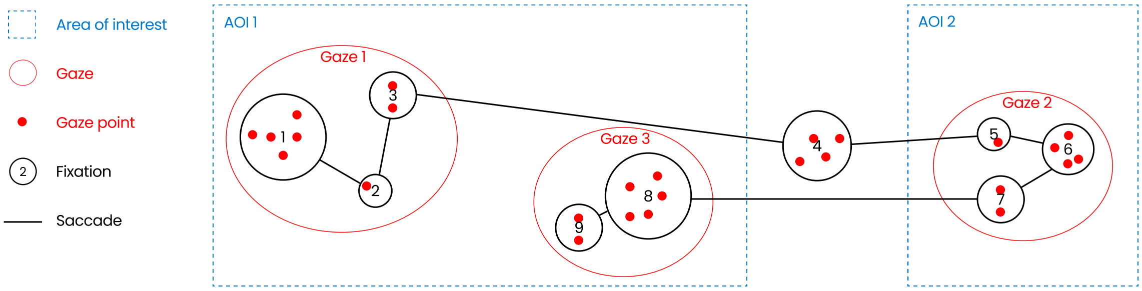


Figure 3.3: Relationship between gaze, fixation and saccades. Inspired by Blascheck et al. (2014). Illustrated by Eva Le.

The initial step before collecting eye-tracking data involves calibrating the system (B. Albert and Tullis, 2013). Calibration includes asking participants to focus on a series of predefined points over an Area of Interest (AOI) for the system to subsequently interpolate the gaze location. An AOI refers to specific elements or regions that researchers are interested in measuring visual attention. Calibration quality is typically expressed in the deviation from x and y planes in degrees. An acceptable calibration is typically within a deviation of less than one degree in controlled experiments. The calibration result defines how reliable the eye-tracking data will be. Therefore, eye-tracking data with poor calibration result should not be included in the analysis. For a high quality calibration, participants should minimise movements and blinking, remain still, and avoid looking away. The quality of eye-tracking data is usually described in accuracy and precision, or in other words, validity and reliability (B. T. Carter and Luke, 2020; Reingold, 2014), see Figure 3.4.

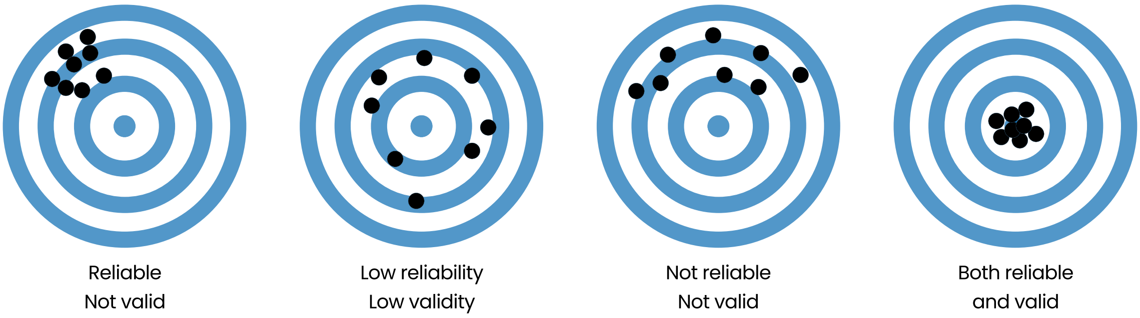


Figure 3.4: Reliability and Validity.

3.3.2 Eye-Tracking Data

There are different kinds of eye-tracking data metrics that are useful for different purposes: dwell time, number of fixations, fixation duration, sequence, time to first fixation, revisits, and hit ratio (B. Albert and Tullis, 2013). These are some of the most common eye-tracking metrics used by UX researchers.

- *Dwell time* refers to the total time spent looking within an AOI, including fixations, saccades and revisits. The metric is commonly used to pinpoint the level of interest using AOI. Generally, dwell time greater than 500 ms is considered as participants have been able to process the information. Dwell time is usually reported as either absolute time (ms) or as a percentage (%).

- *Number of fixations* refers to the number of times a fixation was made within an AOI . As this number is usually strongly correlated to dwell time, just reporting dwell time is often sufficient.

- *Fixation duration* refers to the average time of fixations; normally ranging from 150 to 300 ms.

- *Sequence* refers to the order of first fixation on the AOIs. With several participants, this number is usually an estimation of the average order in which each AOI was visited and does not represent all participants' sequences.

- *Time to first fixation* refers to the time it takes until the participants fixates on an AOI. This number is usually the average time of all the participants' first fixation.

- *Revisits* refers to the number of times returning to an AOI after first fixation.
- *Hit ratio* refers to the percentage of participants who fixated on the AOI at least once.

Dwell time, number of fixations and fixation duration are similar in a way that they all represent the relative engagement with an AOI.

3.3.3 Eye-Tracking Analysis

The most common way to visualise eye-tracking data from multiple participants is through heat maps (B. Albert and Tullis, 2013). Heat maps visualise the density of fixation with a scale of colours. The most common way to analyse eye-tracking data is through AOI analysis. It is recommended to include heat maps while presenting the AOI analysis data, as the question about where the participants looked within the AOI tend to come up. Binning charts offer another way of analysing eye movements by showing the percentage of time spent fixating on each AOI during a given time interval. Gridded AOIs provide another useful form of visualisation.

Pupillometry is commonly used in usability testing with eye-tracking (B. Albert and Tullis, 2013). It involves measuring pupillary responses, such as dilation (increased pupil size) and constriction (decreased pupil size). While ambient light is the primary factor influencing pupillary response, other contributing factors includes emotional arousal, cognitive workload, and increased interest (B. Albert and Tullis, 2013; iMotions, 2024). Higher levels of arousal or interest typically result in pupil dilation. However, interpreting pupillary responses can be challenging, as they correlate with various mental and emotional states.

When analysing eye-tracking data, it is important to consider both the risks and benefits of the system. In a study conducted by W. Albert and Tedesco (2010) on “Reliability of Self-Reported Awareness Measured Based on Eye tracking”, 10% of participants who reported that they had “definitely seen” an element showed no fixations in the eye-tracking data. Another 5% of participants reported having “spent a long time looking at an element”, yet the data did not register any fixations. These findings highlight that, while eye-tracking can be a valuable tool, it is important to remain mindful of possible inaccuracies and aim to ensure the highest data quality.

3.3.4 Ethical Implications of Using Eye-Tracking

Recent research has raised concerns about the growing ethical implications of using eye-tracking technologies. Korkmaz and Gulsecen (2023) classify biometric data into two main categories: physiological (e.g., fingerprints, facial features, iris patterns) and behavioural (e.g., handwriting, gait, keystrokes, and notably, eye movements). Eye-tracking falls within the behavioural category, capturing subtle, often involuntary data patterns that are unique to individuals. Importantly, individuals may not always have control over the sharing of such implicit information, unlike with explicit disclosures in open information sources, such as social media or online surveys.

Korkmaz and Gulsecen (2023) also introduce the concept of *cognitive identity*, which refers to the unique way individuals synthesise knowledge from past experience and real-time perceptual input. Eye-tracking technology has become a valuable tool for studying such cognitive processes, including those related to education, measurement, and decision-making. The capture of gaze patterns, therefore, may inadvertently reveal information tied to this internal cognitive identity structure.

Similarly, Kröger et al. (2020) demonstrate that data commonly captured by eye-trackers can be used to infer a wide range of sensitive personal attributes. These include attributes such as eye opening and closure, eye movements (e.g., gaze, fixations, and saccades), eye condition (e.g., redness, dryness, watering), pupil attributes, iris characteristics (e.g., colour, texture), and various facial attributes (e.g., wrinkles, eye shape, expressions). From these data points, it is possible to infer biometric identity, age, gender, emotional state, and even certain health conditions. Kröger et al. (2020) categorise these types of inferences and emphasise the potential for unintended disclosure of personal information from seemingly innocuous gaze recordings.

3.4 Human Information Processing

As summarised by Kosslyn (1985), based on the work of Marr (1982), the “canonical theory” of human visual information processing includes three general phases: forming perceptual images, organising information in short-term memory, and interpreting the information through long-term memory (see Figure 3.5).

The first phase is the formation of perceptual images, often referred to as early visual processing. This phase involves the initial conversion of light into neural signals as it reaches the retina. At this stage, the raw visual input is detected without meaning but is only seen as visual patterns (e.g., lines, marks, colours, textures, and movement). The output of the processing in the first phase is organised into “perceptual units”.

The second phase involves short-term memory, where perceptual units from the previous phase are temporarily held. These unit are operated in short-term memory, which has limited capacity. Basic visual patterns are grouped into meaningful shapes (e.g., three enclosing lines are perceived as a triangle).

The final phase is supported by long-term memory, where prior knowledge is used to interpret and understand visuals. For example, knowing how to read charts or graphs to understand the information they convey.

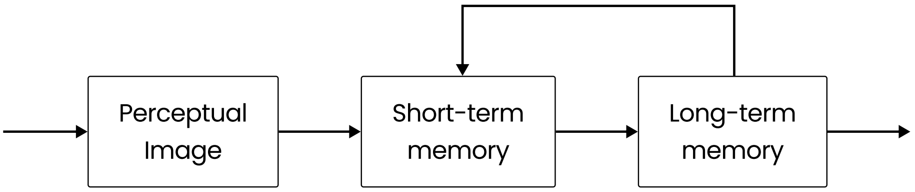


Figure 3.5: Three levels of visual information processing (Kosslyn, 1985).

3.4.1 Attention in Information Processing

According to the definition provided by Proctor and Vu (2007), attention refers to the increased awareness focused on a specific event or action, allowing it to undergo increased processing. This focused processing can lead to a deeper understanding of the event, enhanced performance of the action, or improved memory for the event. By enabling us to filter out irrelevant information, attention helps us concentrate on aspects that are most relevant to our goals. Multiple influential models have been developed to explain how attention is processed in the human cognitive system.

Filter Theory: In early research, the dichotic listening study by Cherry (1953) demonstrated that individuals could accurately repeat the message they focused on, while recalling little from the unattended message beyond basic physical features like the speaker's gender. To explain this, Broadbent (1958) proposed the filter theory, suggesting that the nervous system operates as a single-channel processor. According to this model, attention acts as a filter that selects stimuli based on physical attributes, such as spatial location, allowing only the selected input to proceed for further processing.

Attenuation Theory: Later studies, including Treisman (1964), showed that unattended information could still be partially processed beyond its physical characteristics. For instance, people might still recognise names or notice unexpected events. This lead to the development of "filter-attenuation theory", which proposes that the filter reduces, rather than completely block, the strength of unattended message. If a stimulus has a low identification threshold. it can still be recognised despite the attenuation.

Late-Selection Theory: Deutsch and Deutsch (1963) proposed that all stimuli, including unattended ones, are fully identified and processed for meaning, and that selection occurs at a later stage of processing. Unlike attenuation theory, which suggests limited identification of unattended information, late-selection theory assumes that the semantic content of all inputs is analysed before attention determines which information reaches awareness or influences behaviour.

Load Theory: To address the debate between early and late selection, Lavie et al. (2004) proposed load theory, which suggests that attentional selection depends on task demands. When the perceptual load is high, that is, the perceptual system is under heavy demand, irrelevant stimuli are filtered out early. In contrast, when the memory load is high, cognitive control over distractions is reduced. This means that the likelihood of being distracted by irrelevant information depends on how much of our mental "capacity" is being used, either during the perceptual stage or at the cognitive (memory) stage.

Multiple Resource Theory: In the unitary resource model proposed by Kahneman (1973), attention is viewed as a single pool of limited capacity that is allocated based on task demands and voluntary control. Building on this idea, Wickens (1984) introduced multiple resource theory, which suggests that performance on multiple tasks improves when the tasks draw on different sensory modalities or response systems.

3.4.2 Visual Attention

Visual attention is evident in many everyday activities, helping individuals concentrate on the most relevant aspects of their surroundings. According to Kahneman (1973), it enables us to focus on important elements within a visual scene while disregarding less significant details. This view aligns with the findings of Evans et al. (2011), who emphasise that a central function of visual attention is to filter information by either suppressing irrelevant stimuli or selecting potentially important ones, which allows the brain to allocate its limited processing resources more efficiently.

A frequently used metaphor for visual attention is a spotlight, which is thought to direct attention to everything within its field (Posner, Cohen, et al., 1984). The direction of attention does not always match the direction of gaze, as the attentional spotlight can be shifted independently from where the eyes are fixed (Proctor and Vu, 2007). Studies have shown that when attention is directed to a specific location where a target is expected, but a different stimulus appears elsewhere, a spatial gradient forms around the attended area. As a result, stimuli that appear closer to the focus of attention are processed more quickly and accurately than those that are farther away (Yantis et al., 2000).

Proctor and Vu (2007) explained that the attentional spotlight can be shifted by two types of cues: *exogenous* and *endogenous*. Exogenous cues are sudden external events that automatically draw attention, while endogenous cues require interpretation and involve a voluntary shift of attention.

3.4.3 Distraction and Perceptual Load

According to the distraction-conflict model summarised by Baron (1986), distraction is viewed as something that directs attention away from some ongoing activity. Lavie (2005) argued that maintaining focus on a task is essential for effective cognitive functioning, particularly in the presence of irrelevant distractions. However, individuals are frequently diverted by task-irrelevant stimuli, as shown by everyday examples such as a fly disrupting reading or a billboard capturing a driver's attention.

Distraction is closely linked to perceptual load. According to Lavie (2010), tasks with high perceptual load significantly reduce the impact of distractors, leading to a form of inattentional blindness where individuals may fail to recognise stimuli they have just seen. In such conditions, people may "look but not see." The study also highlights that distractions can arise from both external sources and internal thoughts.

Additionally, age and perceptual load together affect how easily individuals are distracted. Theories of age-related cognitive change suggest that information-processing capacity develops during childhood and declines in older age (Lavie, 2005). Lavie (2005) found that, compared to mature (young to middle-aged) adults, both younger children and older adults experienced more distraction when performing tasks with very low perceptual load.

3.5 Situational Awareness

Researchers have for long been under misalignment with the definition of the terminology of *situation awareness* and have proposed various definitions (Fracker, 1988). However, the core idea still remains the same across all definitions: “knowing what is going on” (M. R. Endsley, 1995). M. R. Endsley (1988) introduced a general definition of SA that applies across various task domains, in contrast to earlier, more limited definitions: *“Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.”.* This definition comprises three levels of SA: perception, comprehension, and projection (M. R. Endsley and Connors, 2008), see Figure 3.6. SA is an essential foundation for effective decision-making and performance.

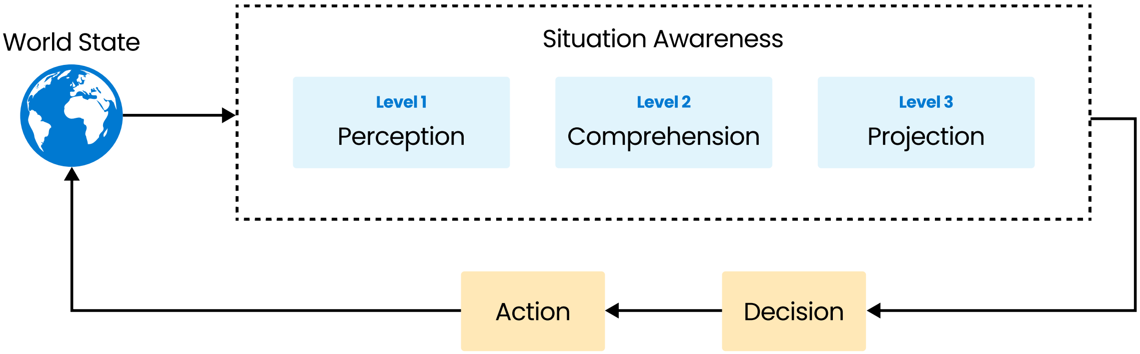


Figure 3.6: The three levels of situation awareness are essential for effective decision-making. Inspired by M. R. Endsley and Connors (2008). Illustrated by Eva Le.

Perception is the first level in achieving SA; perception of the elements in the environment (M. R. Endsley, 1995). It involves detecting meaningful cues through the senses by perceiving relevant elements in the environment in terms of their status, attributes, and dynamics. M. R. Endsley (1995) provided an example where driver’s perception involves knowing where other vehicles and obstacles are in their surroundings, how they are moving, and be aware of the current state and motion of their own vehicle.

Comprehension is the next level in achieving SA. At this stage, simply being aware of the elements is not enough; comprehension of the current situation is required. Comprehension involves understanding the importance of the elements and how they impact the operator’s goal. For example, in a situation where vehicles ahead suddenly slowed down and a long queue formed, a driver demonstrates comprehension by interpreting this as a possible indication of an accident, traffic congestion, or an obstacle ahead.

Projection is the highest level of SA (M. R. Endsley, 1995). It is achieved by integrating knowledge from perception and comprehension to predict future states based on the current situation. If building on the previous example, through seeing vehicles in front slowing down and understanding that something must have happened, drivers may engage in projection by anticipating emergency vehicles to arrive or the traffic may come to a complete stop.

M. Endsley (2020) emphasises the importance of SA as a fundamental to successful driving. Systems should be designed to assist operators in maintaining and developing SA for efficient decision-making (M. R. Endsley and Connors, 2008). A common pitfall is adopting a technology-centred design approach rather than a user-centred one when aiming to support human work. In the context of automation of vehicles, M. Endsley (2020) expressed that cars are yet to become fully automated, and suggests guidelines for semi-automation in vehicles. For instance, M. Endsley (2020) explained that driver-assist automation with protective features (e.g., automation limiters) are more effective than semi-automated features (e.g., adaptive cruise control) as they keep drivers actively engaged.

3.6 Cognitive Load

Cognitive Load Theory (CLT) explains how humans process information based on the idea that working memory is limited and handles new information, while long-term memory is unlimited and stores knowledge more permanently (Sweller et al., 1998). CLT was proposed as a guideline for instructional design aimed at reducing the load on working memory to support schema development in long-term memory for an improved learning outcome (Sweller et al., 2019). Cognitive load was divided into three types of load: intrinsic, extraneous, and germane load.

Intrinsic cognitive load refers to the complexity of the information being processed depending on learner's current knowledge, the nature of the information, and the level of element interactivity involved (Sweller et al., 2019). Element interactivity refers to the number of elements in a task that must be processed simultaneously in working memory, and the degree to which these elements interact with each other. *Extraneous cognitive load* refers to the additional mental effort caused by ineffective instructional design or a high level of element interactivity. *Germane cognitive load* refers to the load required to handle intrinsic cognitive load, which depends on the amount of resources going to unwanted extraneous cognitive load.

3.6.1 Cognitive Load Measurement in HCI

As noted by Kosch et al. (2023), HCI studies commonly use the terms “cognitive load”, “cognitive workload”, and “mental workload” interchangeably to describe task-related mental demands. According to von Janczewski et al. (2022), cognitive workload is defined as the portion of an individual's information processing capacity or resources that is devoted to performing tasks in order to satisfy system demands. Tools for measuring specific types of cognitive load can be classified into three main categories: *subjective measures*, *performance measures*, and *psychophysiological measures*, with the latter two falling under the broader category of objective measures (Galy et al., 2012; von Janczewski et al., 2022).

Among subjective measures, two of the most widely used tools are the National Aeronautics and Space Administration-Task Load Index (NASA-TLX) (Hart and Staveland, 1988) and the Subjective Workload Assessment Technique (SWAT) (Reid and Nygren, 1988). These tools rely on participants' self-reports to evaluate perceived workload across different dimensions, such as mental and physical demands. Performance measures assess cognitive workload by evaluating observable behaviours such as response accuracy and response time (Galy et al., 2012). Psychophysiological measures capture physiological signals that are interpreted in relation to mental workload. Common examples include electrocardiogram (ECG), heart rate variability (HRV), electroencephalography (EEG), electrodermal activity (EDA), and various indicators of eye activity such as pupil dilation and blink rate (von Janczewski et al., 2022).

3.6.2 Subjective Workload Assessment using NASA-TLX

The definition of workload has been widely debated in psychology among "experts", highlighting the complexity of the concept (Hart and Staveland, 1988). Each respondent is also likely to have their own interpretations, associating the term "workload" with different personal experiences. Hart (2006) explains the term *workload* as the "cost of accomplishing mission requirements for the human operator", emphasising that this definition adapts an human-centred rather than a task-centred approach. Understanding workload helps reveal task demands that may not be visible through performance alone, offering a more complete picture of operator experience.

NASA-TLX is a multi-dimensional workload rating technique used to quantify subjective assessment of overall workload (Hart and Staveland, 1988). NASA-TLX consists of six single-dimensional, or multi-dimensional rating subscales reflecting a relatively distinct group of variables: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Component ratings like this provide diagnostic value by identifying the sources of workload (Hart, 2006). These subscales were selected after a thorough analysis on the subjective experience of workload assessment in a simple laboratory task of flying an aircraft. The selected dimensions also happened to align with existing theories of workload, including the effects workload imposes on operators and how operators respond to task demands. Hart and Staveland (1988) explains the NASA-TLX procedure, where operators are asked to rate each subscale on a scale from 1 to 20. These subscales are also compared pairwise, resulting in 15 pair comparisons. Each time a subscale is selected in a pairwise comparison, it receives one point of weight. Weighting more accurately reflects the contribution of each subscale to the overall workload by reducing between-rater variability and increasing sensitivity to relevant variables. The overall workload score is calculated as a weighted average of the subscale ratings. Figure 3.7 illustrates an example of a weighted subscale and the resulting overall workload.

Over 20 years since NASA-TLX was proposed, the use of the tool has been translated, modified, administered in various formats, and compared to other workload assessment tools (Hart, 2006). Hart (2006) could not guarantee the quality of the translations due to limited language skills, but most of the work were considered successful. The tool was shown to be applicable across various domains, as the

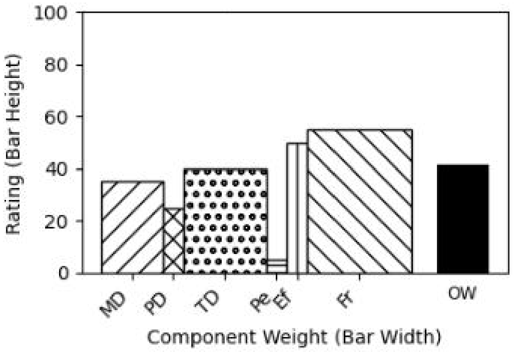


Figure 3.7: Graphic representation of weighted subscale ratings and an overall workload score.

focus is on the basic human activities. However, several studies have applied modifications to the NASA-TLX, such as relabelling, adding, or removing subscales, to better align with specific study contexts. This practice is approached with caution by Hart (2006), who explains that major modifications require the establishment of validity, sensitivity, and reliability, as the resemblance to the original tool is minor.

On the other hand, Hart (2006) views the flexible use and analysis of NASA-TLX positively. In some studies, the weighting process was eliminated (Raw TLX), or the weighted subscales were analysed individually. These approaches were compared to the original NASA-TLX and found to be either more sensitive, equally sensitive, or less sensitive depending on the specific study. Other studies chose to analyse the subscales instead of or in addition to the overall workload score, an approach supported by Hart (2006), who emphasised the diagnostic value of the subscales. However, users have noted significant correlations between the subscales, which Hart (2006) considers natural, as they all assess different dimensions of the same underlying construct. Hart (2006) also addresses context and anchor effects related to subjective ratings in general. To control for these effects, anchors (explicit experiences or instructions) could be provided to operators.

3.7 Literature Review

According to the guidelines suggested by Wadsworth (2020), it is essential for researchers to immerse themselves in the topic prior to conducting the study, as this deepens contextual understanding and informs the research direction. A well-founded user study required a thorough review of relevant literature on key subjects.

A thorough literature review and an understanding of state-of-the-art (SotA) research are fundamental to academic research, as they establish a foundation for identifying research gaps, contextualising previous work, and positioning new studies within the existing body of knowledge. Once this groundwork is laid, clear research objectives can be defined to determine what data to collect, select appropriate methodologies, and analyse results effectively (Robson and McCartan, 2016). This structured approach ensures that the user study is both feasible and aligned with current practices.

3.7.1 Systematic Literature Review

A structured review of existing research is necessary to identify key findings. According to Nightingale (2009), unlike traditional literature reviews, systematic reviews are conducted to gather and synthesize prior research, which aim to identify all studies, including published and unpublished ones (Nightingale, 2009). The author also highlights that literature reviews can be affected by selection bias when authors include only studies from a specific area or those supporting their opinions. Therefore, conducting systematic reviews that incorporate all types of research can help minimise bias and enhance multi-dimension discussion. Four key areas are needed in terms of review literature in a systematic approach given by Badger et al. (2000): (1) definition of the problem; (2) the search strategy; (3) criteria for the evaluation of studies; (4) data extraction.

3.7.2 State-of-the-Art Literature Review

A SotA review is particularly useful in the context of emerging technologies, as it provides a chronological perspective on technological advancements and their applications. Barry et al. (2022) demonstrates that SotA reviews offer researchers the insights of *where the field currently stands*, *how it has evolved over time*, and *the potential future directions.*

3.8 Study Design

A research design outlines the overall strategy for addressing the study's research questions. It ensures that the evidence gathered is relevant, credible, and reliable, while also considering potential challenges and alternative perspectives (Wadsworth, 2020).

The practice of conducting user studies has long been applied in the medical field, commonly referred to as clinical trials in "evidence-based medicine". One approach to this is A/B testing, also known as Online Controlled Experiments (OCEs) or split tests, used for comparing two variations of a system by exposing it to the end-users (Gupta et al., 2019; Kohavi et al., 2020; Siroker and Koomen, 2015), see Figure 3.8. Variation A is usually known as the control group, and variation B is known as the treatment group (Kohavi et al., 2020). In practice, A/B testing is widely used across various fields, including psychology, pharmaceutics, education, software engineering, and automotive. However, there is a lack of comprehensive, empirically grounded studies on the SotA research (Quin et al., 2024).

3.9 Data Gathering

According to Sharp et al. (2019), a framework for data gathering consists of five key steps. The first step is *setting goals.* Establishing clear research goals is essential,

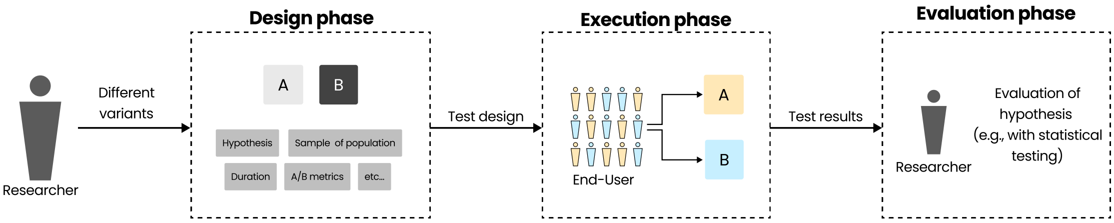


Figure 3.8: Example of a general A/B testing process. Inspired by Quin et al. (2024). Illustrated by Eva Le.

as these goals determine what data to collect, which methods to employ, and how to analyse the results (Robson and McCartan, 2016). The second step is *identifying participants*. The selection of participants, also referred to as *sampling* (Sharp et al., 2019), plays a crucial role in the data gathering process and will be discussed in more detail in Section 3.9.1. The third step is *formalising the relationship with participants*. In this phase, obtaining informed consent is essential, as ethical considerations are central to responsible research. Participants are typically asked to sign consent forms that ensure they understand the purpose of the study and their rights, including the right to withdraw at any time. This process also permits researchers to use the collected data for analysis, presentation, and publication, as noted by Sharp et al. (2019). The fourth step is *aiming for triangulation*. To enhance the reliability and validity of the data, triangulation is often employed. This involves cross-verifying findings from multiple perspectives (Jupp, 2006; Olsen, 2004). Jupp (2006) identifies four types of triangulation: *data triangulation*, *investigator triangulation*, *theory triangulation*, and *methodological triangulation*. Finally, the fifth step is *conducting pilot studies*. Pilot studies serve as a preparatory stage before the main study, allowing researchers to identify and address potential issues in data collection, such as flaws in the experimental setup or challenges in participant interaction. This process contributes to the overall feasibility and reliability of the full-scale investigation (van Teijlingen and Hundley, 2002).

3.9.1 Sampling

According to C. R. Carter et al. (2023), sampling process should follow inclusive recruitment strategies that value varied backgrounds and perspectives. In addition, choosing an appropriate sampling method is essential to aligning with the study's objectives. Various sampling techniques exist, each suited to different research contexts and constrains (Sharp et al., 2019).

Saturation sampling, which involves accessing the entire target population, is uncommon in most of the research context. More typically, only a subset of the population is available, in which case researchers typically choose between two main sampling strategies: *probability sampling* or *nonprobability sampling*. In the case of probability sampling, commonly used methods include *random sampling* and *stratified sampling*. For nonprobability sampling, common approaches involves *convenience sampling* or *volunteer panels*. The key difference between probability and nonprobability sampling methods is that probability methods allow for the application of statistical tests and enable generalisation to the entire population, whereas nonprobability methods do not support such robust generalisations.

The sample size of participants is determined by the scope of the evaluated use cases and the analytical methods applied during data analysis. According to Caine (2016), there are many ways to determine how many participants are needed for the study, including saturation, cost and feasibility analysis, guidelines, and prospective power analysis.

Saturation occurs in qualitative research when further data collection no longer reveals new relevant information (Glaser and Strauss, 2017), serving as a common indicator for concluding gathering of data from interviews and observations. *Cost analysis* can help researchers to determine the sample size based on the limit of funding. Beyond momentary costs, conducting a *feasibility analysis* is also crucial, as other constraints such as time available to complete the study and participant availability must be considered (Glaser and Strauss, 2017). *Guidelines* to determine sample size may come from recommendations by experts or local standard (Glaser and Strauss, 2017). *Prospective power analysis* is a rigorous and defensible method of determining sample size (Baxter et al., 2015; Glaser and Strauss, 2017; Sauro and Lewis, 2016).

3.9.2 Methods of Data Gathering

Collecting both quantitative and qualitative data by employing a mixed-methods approach enables a more comprehensive analysis. As Lakshman et al. (2000) state, combining both methods can enhance validity and reliability by providing a broader and deeper understanding of the research problem. Since quantitative research alone might not be able to answer all questions (Lakshman et al., 2000), qualitative methods often help contextualise the numerical data.

Questionnaires can be employed to gather both quantitative and qualitative data. As Wadsworth (2020) highlights, *closed-choice* questions are particularly useful for statistical analysis. Likert scale questions, as a simple to construct and reliable technique (Bertram, 2007). However, employing standardised and validated questionnaires is preferable to using self-constructed Likert scales. Additionally, *open-ended* questions provide greater flexibility in capturing deeper insights. As noted by Wadsworth (2020), *open-ended* questions can enhance validity and allow for a more nuanced understanding of participant experiences. The qualitative data obtained will be further analysed using methods such as thematic analysis (TA) to identify potential patterns and meanings.

Interviewing is a commonly used research method that encompasses various formats, including structured, semi-structured, and unstructured interviews. Robson and McCartan (2016) explain that directly asking individuals about their experiences is a straightforward way to obtain answers to research questions. As Wadsworth (2020) suggests, it is essential to refine interview questions before presenting them to all participants. Additionally, consider digital recording for interviews with participants.

As Wadsworth (2020) explains, digital recordings can serve as a useful reference for later note-taking and can provide highly detailed data for in-depth analysis and re-analysis. However, this needs to ensure that informed consent is obtained from each participant before recording.

Observations involve keeping records such as journal entries, short narratives, or measurements, which can be later expanded, analysed, and shared for further discussion on their meaning implications (Wadsworth, 2020). As Robson and McCartan (2016) highlight, a key advantage of observation is its directness, providing data that contrasts with and can effectively complement information gathering through other methods (e.g., questionnaires and interviews).

3.10 Data Analysis

After data have been collected in a research project, they are typically analysed and interpreted. The methods and outcomes of analysis provide the foundation for interpretation. As Robson and McCartan (2016) explain, most data can be broadly categorised as either qualitative (words or other non-numerical forms) or quantitative (numbers or data that can be transformed into numerical form). Qualitative data are often examined using interpretative methods, while quantitative data are analysed through statistical approaches.

As summarised by Robson and McCartan (2016), Onwuegbuzie and Teddlie (2003) propose a comprehensive framework for the process of data analysis in mixed-methods research, consisting of the following steps:

1. *Data reduction*: Involving summarising quantitative data, such as using descriptive statistics, and qualitative data through techniques such as TA.
2. *Data display*: Presenting quantitative data using tools such as tables and graphs, and qualitative data such as matrices, charts, and networks.
3. *Data transformation*: Referring to the process of 'qualitising' quantitative data and/or 'quantising' qualitative data.
4. *Data correlation*: Correlating quantitative data with transformed qualitative data.
5. *Data consolidation*: Combining both data types to generate new variables or datasets.
6. *Data comparison*: Focusing on comparing data from different data sources.
7. *Data integration*: Involving combining all data into a single coherent one, or constructing coherent quantitative and qualitative data separately.

3.10.1 Data Preprocessing and Validation

Data preprocessing is a critical step in data analysis, intended to enhance the overall quality and reliability of the dataset (Niu and Tang, 2024). It typically involves handling missing values, addressing outliers, and applying data normalisation techniques (Fu et al., 2023). Data validation is a decision-making process that determines whether data are accepted or rejected based on their adequacy and reliability, aiming to ensure the quality of the final dataset (Di Zio et al., 2016). As emphasised by Di Zio et al. (2016), data validation specifically addresses quality dimensions related to the ‘structure of the data’, including accuracy, comparability, and coherence.

3.10.2 Quantitative Analysis

Quantitative methods examine the effects of specified conditions (i.e., independent variable) on an outcome of interest (i.e., dependent variable) through approaches that can be expressed and analysed numerically (Lakshman et al., 2000). Quantitative analysis often relies on statistic (Kotronoulas et al., 2023). According to Mishra et al. (2019), two primary statistical methods are employed in data analysis: descriptive statistics, which summarise data using indices such as the mean, median, and standard deviation (SD); and inferential statistics, which draw conclusions from data using statistical tests such as the Student’s t-test and ANOVA.

Descriptive statistics, also referred to as summary statistics, represent key characteristics of a dataset using single numerical indicators (Robson and McCartan, 2016). They are commonly employed to describe individual variables and are typically conducted as univariate analyses (i.e., analysing one variable at a time). As noted by Patel (2009), researchers commonly perform descriptive statistics as a preliminary step before progressing to more advanced analytical methods.

Inferential Statistics are employed to test hypotheses by estimating the probability that a proposed effect, relationship, or difference is likely to be true (Kotronoulas et al., 2023). These statistical tests yield a probability value, commonly referred to as “p-value”. Inferential statistical methods are typically classified into two categories: parametric and non-parametric (Mishra et al., 2019).

3.10.3 Statistical Methods

According to Mishra et al. (2019), the selection of an appropriate statistical method is determined by three key factors: the aim and objectives of the study, the type and distribution of the dataset, and the nature of the observations (i.e., whether they are paired or unpaired).

Assessing normality is a key step in deciding between parametric and non-parametric approaches, as parametric methods assume normally distributed data while non-parametric methods do not (Kim and Park, 2019). Common methods for assessing normality include statistical tests such as the Shapiro-Wilk test, Kolmogorov-Smirnov test, and Anderson-Darling test, as well as visual techniques like quantile-quantile (Q-Q) plots. Notably, the central limit theorem states that when the sample size reaches 30 or more, the studentised sampling distribution tends to approximate the standard normal distribution, thereby reducing the importance of assumptions regarding the population distribution (Kwak and Kim, 2017). As a result, even when the sample mean is standardised using the sample variance for samples exceeding this threshold, the normal distribution can be appropriately employed to model the corresponding probability distribution.

Common parametric and non-parametric tests are presented in Table 3.1. As summarised by Nahm (2016), each approach has distinct advantages and limitations. Parametric methods assume specific characteristics of the population distribution and generally offer greater statistical power and efficiency when these assumptions are satisfied. In contrast, non-parametric methods are less sensitive to assumption violations, making them more suitable for small sample sizes or non-normally distributed data; however, they tend to have lower statistical power, provide less detailed information, and may be more difficult to interpret.

	Parametric tests	**Non-parametric tests**
One sample	Simple t-Test	One sample Wilcoxon signed rank test
Two dependent samples	Paired Sample t-Test	Wilcoxon signed-rank test
Two independent samples	Independent samples t-Test	Mann-Whitney U-Test
Three or more independent samples	One-way ANOVA	Kruskal-Wallis H Test
Three or more dependent samples	Repeated Measures ANOVA	Friedman Test
Correlation between two variables	Pearson Correlation	Spearman Correlation

Table 3.1: Comparison of parametric and non-parametric tests.

3.10.4 Qualitative Analysis

Qualitative analysis is applied when the relevant variables influencing an outcome are not clearly identifiable, or where the sample size or number of outcomes are too limited to support statistical analysis (Crabtree and Miller, 1992; Lakshman et al., 2000). Robson and McCartan (2016) summarised various approaches to qualitative analysis, including: *quasi-statistical approaches* that focus on the frequency and correlation of words of phrases; *thematic coding approach* that identifies and groups meaningful data segments into themes for further interpretation; and *grounded theory approach* that involves generating theory directly from the data through inductive, interpretive coding. Additionally, there are other specialised approaches, including *discourse and conversation analysis*, and the analysis of *narratives*.

Among these approaches, *thematic coding analysis*, also known as *thematic analysis* (TA), is a generic approach to qualitative data. It can be used as a realist approach to report participants' experiences and meanings. Alternatively, it can be applied as a constructionist approach to explore how social discourses shape those experiences and meanings (Robson and McCartan, 2016).

TA typically involves five key phases. First, researchers familiarise themselves with the data through repeated reading and initial note-taking. Second, initial codes are generated systematically across the dataset, either through a predefined framework or inductively from the data. Third, related codes are grouped into potential themes, which are then reviewed and refined to ensure coherence. Fourth, a thematic map or network is constructed to visually organise the themes. Lastly, the analysis is integrated and interpreted by comparing data using visual tools (e.g., tables and networks), then exploring, describing, summarising, and interpreting patterns while ensuring the quality of the analysis.

4

Process and Execution

This chapter outlines the thesis process, starting with the planning phase, in which the goals and direction of the study were defined. This is followed by the preparatory phase, where test equipment was installed, familiarised, and the methods were selected. Next, the execution of the user study is described in detail. Lastly, the chapter ends with an overview of the data processing and analysis procedures.

4.1 Planning Phase

This thesis was conducted in collaboration with multiple parties, including Chalmers University of Technology, VCC, RISE, and Smart Eye through the "SCREENS II" project. The project scope was defined after several meetings with the stakeholders to discuss the feasibility, potential, and overall objectives. Figure 4.1 below illustrates the planned timeframe for the project.

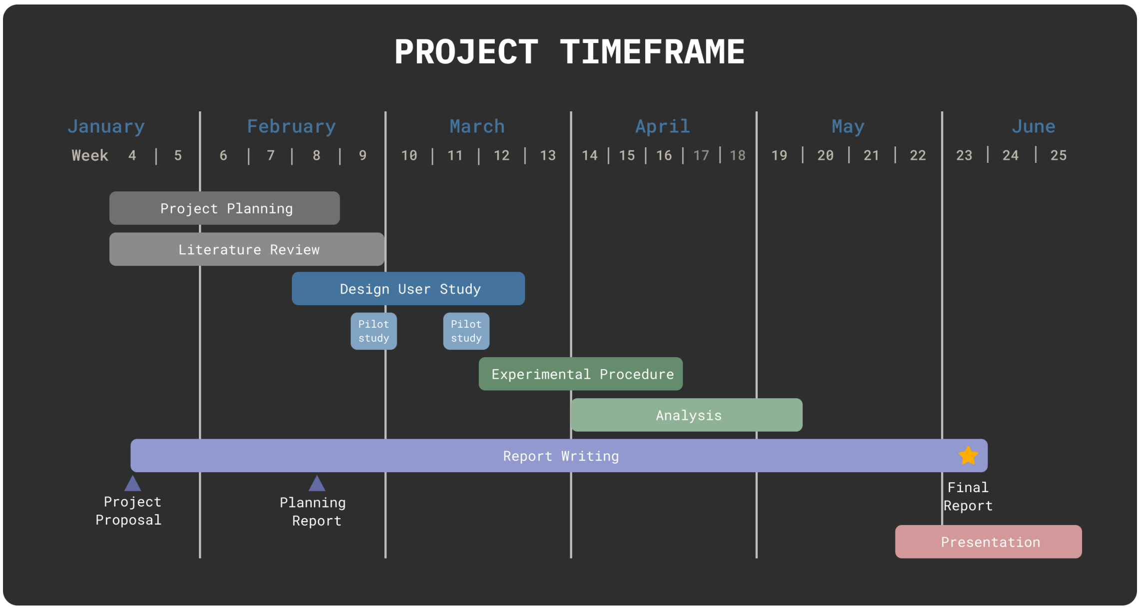


Figure 4.1: Project timeframe.

4.1.1 Behavioural Assessment Framework

The focus of the project was to assess the influence of in-vehicle displays by collecting data on driver behaviour, specifically attention, distraction, and cognitive load, and evaluating their effects on safety and usability. The framework can be seen below in Figure 4.2.

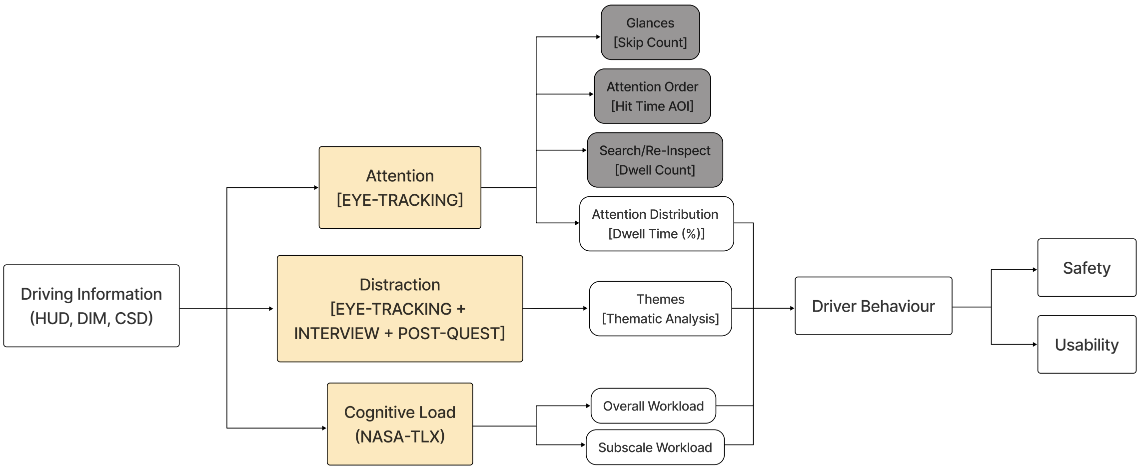


Figure 4.2: Framework for measuring driver behaviour influenced by in-vehicle displays.

In this project, a user study was conducted to gather empirical evidence essential for better understanding end users and creating improved designs (Wildemuth, 2003). The primary focus was on data and methodological triangulation to enhance the validity of the findings. Data triangulation involved gathering information from multiple sources, including test drive observations, post-drive interviews, and recorded behavioural metrics such as eye-tracking data. Methodological triangulation was achieved by integrating both qualitative and quantitative research methods. This included combining direct observations and interviews (qualitative) with questionnaires and objective eye-tracking data (quantitative), allowing for a comprehensive understanding of participants' behaviours and decision-making.

Quantitative data, such as gaze direction measurements from eye-tracking and responses from structured questionnaires, helped assess how drivers distributed their attention across different displays and the affected cognitive load under various conditions. This approach enabled the evaluation of aspects such as HUD effectiveness, attention shifts, and driver engagement. Meanwhile, qualitative data from observations and interviews provided deeper insight into UX, cognitive load, and participants' subjective interpretations of the quantitative results.

4.1.2 Tasks Definition

Following the definition outlined in the NHTSA (2012) guidelines, secondary tasks were defined as driver interactions with in-vehicle devices that were not directly related to the primary task of driving. As the focus of this thesis was to study the impact of a HUD on driving, the different features available in the Volvo XC60 HUD were first considered. The HUD in this vehicle primarily includes four functions: speed, cruise control, navigation, and road sign information (see Figure 2.3). After careful evaluation, the navigation feature was selected as the main focus of the study.

As manipulating the HUD content manually during the experiment was not feasible, the dynamic, continuous, and moderately demanding nature of the navigation feature made it the most suitable choice. In contrast, focusing solely on the speed display would have limited the range of tasks participants could perform in a naturalistic context. Cruise control was excluded because not all participants were familiar with it in their everyday driving, and the road sign display was considered too static and restricted in terms of tasks it could meaningfully support.

Although defining a task that combined multiple features was an option, the study ultimately concentrated on navigation to ensure the tasks remained relevant, safe, and representative of real-world driving. Additionally, handling incoming phone calls was included as a supplementary secondary task, as it could be manually triggered and reflects a common driver activity. Accordingly, the primary and secondary tasks were defined as follows:

(1) Primary task: Maintaining vehicle control by steering, accelerating, and braking, while following traffic regulations to ensure safe driving.

(2) Secondary task: Checking Global Positioning System (GPS) information on the displays and following navigation instructions; noticing incoming phone calls and responding by interacting with in-vehicle interfaces.

4.1.3 Distraction Identification

Following the definition of driver distraction by NHTSA (2012) in Section 2.4.2 and the task classifications specified in Section 4.1.2, the causes and manifestations of distraction examined in the study are summarised in Table 4.1.

	Visual distraction	**Manual distraction**	**Cognitive distraction**
Causes	Diversion of gaze from the road to attend to in-vehicle displays (CSD, DIM, and HUD)	Removal of one or both hands from the steering wheel to perform secondary tasks	Attention shift from driving to secondary tasks (e.g., processing navigation or phone call information)
Manifestations	Failure to monitor the road environment, mirrors, or potential hazards	Reduced steering stability and decreased precision in vehicle control	Slower processing, understanding, or reacting to road events
Possible influence on driving	Lane deviation, missed road signs or hazards, increasing hazard reaction time	Impaired vehicle handling, lane deviation	Miss navigation instructions, slower emergency response

Table 4.1: Summary of causes and manifestations of distraction in the study.

4.2 Preparatory Phase

Before conducting the user study, it was essential to prepare, mount and become familiar with the equipment. Starting from getting the eye-tracking system installed in the test vehicle, creating a world model of the in-vehicle components for the eye-tracking software, to choosing the test routes and conducting pilot studies. In addition, the eye-tracking system required a certain power output during operation. Therefore, power sockets providing household-level current were installed in the trunk of the vehicle, with the support from the Advanced Engineering Workshop at VCC. A list of the hardware and software equipments used in the study can be found in Appendix A.

4.2.1 Camera Installation

To capture drivers' behaviour in terms of visual attention, a stand-alone eye-tracking system (Smart Eye Pro) provided by Smart Eye was installed in a test vehicle (Volvo XC60) that was used during the study. This setup allowed for efficient and unobtrusive data collection in a natural driving environment, preserving the authenticity of the driving experience while recording gaze behaviour.

The installation process of the eye-tracking system was carried out with assistance from Smart Eye's technical team. This included deciding on the number of cameras to use, placing the eye-tracking cameras, and routing the cables to minimise the risk of accidental disconnection.

Smart Eye Pro is a multi-camera system that enables versatile setup configurations, including in-vehicle environments. For this study, three eye-tracking cameras were installed, including two IR light sources facing towards the driver's headrest, see Figure 4.3. By estimating the general position of the drivers head, the eye-tracking cameras were positioned to capture the *headbox*, which refers to the area within which eye-tracking is possible (Niehorster et al., 2018). The eye-tracking cameras were placed on top of the dashboard horizontally under the windshield to capture eye movements within the headbox. The surfaces of dashboards are typically smooth and slippery, making them unsuitable for attaching objects. Ideally, the cameras could have been securely mounted by attaching them to a metal pole fixed to the dashboard. However, such a setup would have been too intrusive by obstructing a large portion of the windshield. As the goal was to maintain a naturalistic driving environment, the setup needed to be minimally intrusive yet highly flexible, as suggested by Vehlen et al. (2021).

Therefore, double-sided acrylic tape was used instead. By compromising the stability of the eye-tracking camera placements, calibrations were done more frequently. Three eye-tracking cameras were placed across the dashboard in the leftmost, central, and rightmost positions (see yellow circles in Figure 4.3). In addition, two webcams were also mounted (see the red circles in Figure 4.3). One webcam was mounted behind the driver's headrest, using a flexible arm with a pipe clamp and a suction cup attached to the car's sunroof. This webcam served as a *Scene Camera*, recording the three displays: CSD, DIM, and HUD. The other webcam served as an *Environment Camera*, was mounted just below the windshield, in the middle of the dashboard, to record the traffic environment.

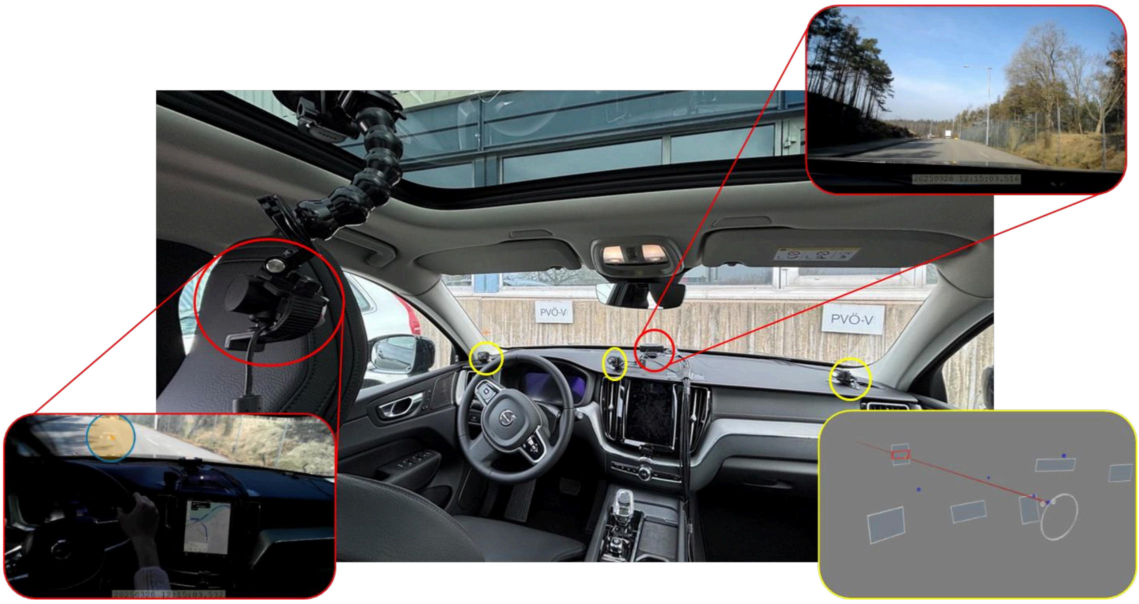


Figure 4.3: Settings in the test vehicle Volvo XC60 with three types of displays (HUD, DIM, CSD), three eye-trackers circled in yellow (Smart Eye Pro), and two webcams circled in red (Microsoft LifeCam as a Scene Camera, and Logitech C920 as an Environment Camera).

There was an option to mount a fourth eye-tracking camera between the CSD and the transmission lever to gain higher eye-tracking exposure and data quality. However, adding more cameras also increased the risk of accidental contact. The eye-tracking cameras are extremely sensitive to any changes in their relative positioning. If the cameras' positions deviate significantly from the created world model, it may be necessary to recreate the model (see Section 4.2.2 for details on model creation). Due to the high risk associated with this placement, the option was therefore discarded.

4.2.2 World Model Creation

To relate eye-tracking measurements to real-world physical space, defining a World Coordinate System (WCS) was essential (Smart Eye AB, 2002). This system acts as a common reference frame that all cameras and data can align to. In this study, the estimated headbox (around the headrest) of the driver was defined as the origin of the WCS. The origin was established using a specially designed chessboard positioned to be clearly visible to all the eye-tracking cameras. From the newly defined origin and the chessboard with an attached laser source, world model components were created based on existing elements inside the vehicle (see Figure 4.4).

When creating world model components, three to four reference markers were needed for each object in the vehicle (see Figure 4.5). The laser chessboard was then used to estimate the distance from the defined WCS origin to the object component by shooting the laser at the reference markers. With three reference points, x-, y-, and

Figure 4.4: The laser chessboard shoots lasers on the reference markers to get the coordinates in relation to the WCS.

z-axes of each real-world object were defined in the world model. It should be noted that with only three reference markers, the rectangular objects created in the world model are just estimations, and not an actual representation of the objects.

Figure 4.5: Reference markers (three yellow dots) on the left rear mirror.

One challenge during the creation of world model components was when creating the HUD object. The AOI of the HUD display is a virtual image around two metres in front of the driver's seat, so it does not exist in the physical space. This posed a challenge when using the laser chessboard to estimate the distances to each component. As a solution, a cardboard box was placed approximately at the virtual image plane to represent the estimated HUD position, which was then measured and marked as seen in Figure 4.6. Another factor to consider when creating a HUD object in the world model was its adjustable height, which varies based on the driver's preference and height. The HUD offers 30 different height settings. To account for the variation of HUD positions, three HUD objects stacked on top of each other were created in the world model (see Figure 4.7).

The initial world model consisted of 10 components which represented seven real-world objects: (1) left rear view mirror, (2) rear view mirror, (3) right rear view mirror, (4) left windshield, (5) right windshield, (6) CSD, (7) DIM, (8) HUD high,

Figure 4.6: Creating a HUD component in the world model.

(9) HUD mid, and (10) HUD low. As shown in Figure 4.7, the gaze vector is visualised as a red line extending from the driver to an object in the world model. A yellow dot indicates the point of intersection between the gaze and an object, while the first intersected object is highlighted with a red outline. As the HUD was the key object of interest in this study, having only the windshield lit up when the driver looks at the HUD was suboptimal during analysis. After having conducted a pilot study, it was concluded that removing the windshield component from the final world model was more beneficial. As a result, the remaining world model components were defined as AOIs for the study.

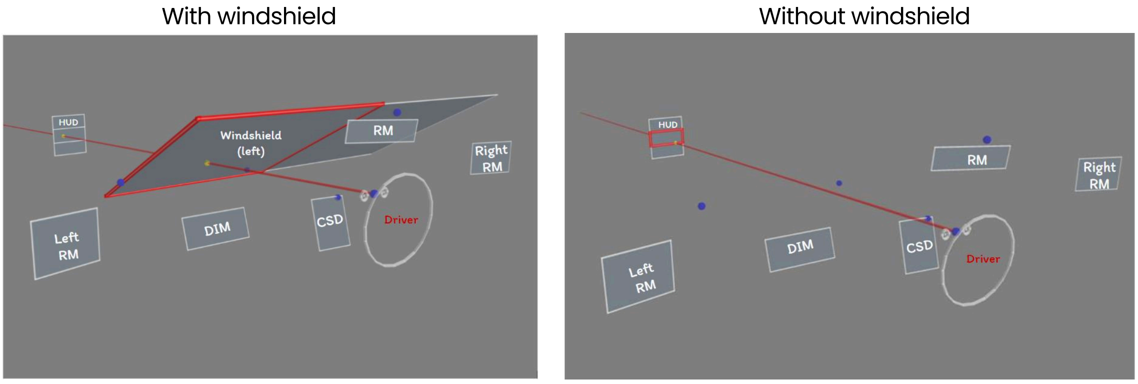


Figure 4.7: World model built in Smart Eye Pro Software.

From Figure 4.8, four calibration points were placed in the vehicle visible as yellow dots. These calibration points were placed in a way that covered a driver's varied gaze locations. One calibration point was placed at the lower-left edge of the windshield, another at the upper-right corner of the CSD, the third above the rear view mirror, and the fourth near the lower-middle edge of the windshield.

Figure 4.8: Four gaze calibration points marked in yellow.

4.2.3 Eye-tracking Recording in iMotions

The eye-tracking data was recorded using a separate software, iMotions. With iMotions, different sensors such as webcams, the eye-tracking system with a world model, and GPS were synchronised during the recording (see Figure 4.9). The Scene Camera mapped the gaze data from the world model onto the corresponding video recordings, allowing a visual representation of where the driver was looking. While the Environment Camera and GPS provided an overall understanding of the surrounding traffic environment. Together, these data sources formed a comprehensive system for monitoring and analysing driver behaviour within real-world environments.

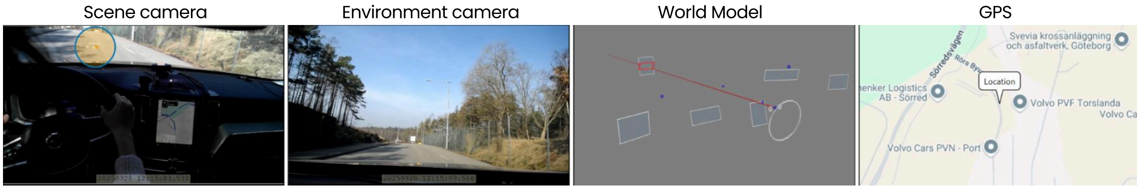


Figure 4.9: Synchronised scene camera, environment camera, world model, and GPS in iMotions.

4.2.4 Test Routes Selection

When selecting the test routes, the primary considerations were safety and consistency of traffic conditions, regardless of time or day. Since the study relied on the test vehicle's built-in navigation (Google Maps), it was not possible to freely customise the route through the software (e.g., specifying coordinates in between the trips). Thus, a route characterised by light traffic and scheduled outside of rush hours was preferred, in order to minimise the risk of automatic rerouting. With this in mind, several routes near the designated starting point of the study were tested and evaluated. This evaluation involved identifying each instance of navigation changes occurred on the HUD. These changes were detected by reviewing the HUD recordings captured from within the eye-box and manually placing event markers in the iMotions software. One of the evaluated routes had also been used in a previous thesis project by Hansols (2022).

In addition to traffic safety and consistency, the final test routes were selected based on environmental variety and complexity, driving duration, and a moderate number of HUD navigation changes. The out-and-back directions were assessed to have similar navigation patterns. The routes that met the specified requirements were selected: Route 1 from PVH (Volvo, Jakobs väg) to the destination at Gamla Flygplatsvägen (see Figure 4.10), and Route 2 which followed the same path in the opposite direction (see Figure 4.11). As shown in the figures, 16 changes on the HUD were marked on the map for both Route 1 and Route 2. Detailed information of the changes can be seen in Appendix B.8. The locations where test calls took place were marked in yellow. Each route had an approximate duration of 10 minutes, and the tests were conducted during morning to midday hours, between 08:00 and 15:00.

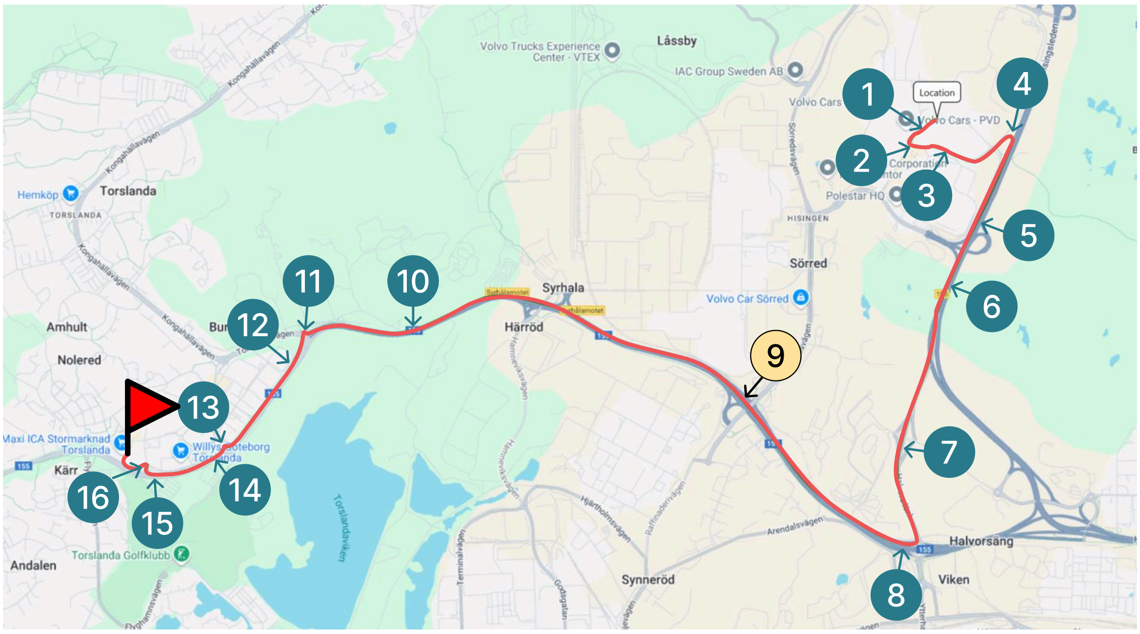


Figure 4.10: Route 1 with 16 marked navigation changes on the HUD starting from Volvo PVH to Gamla Flygplatsvägen.

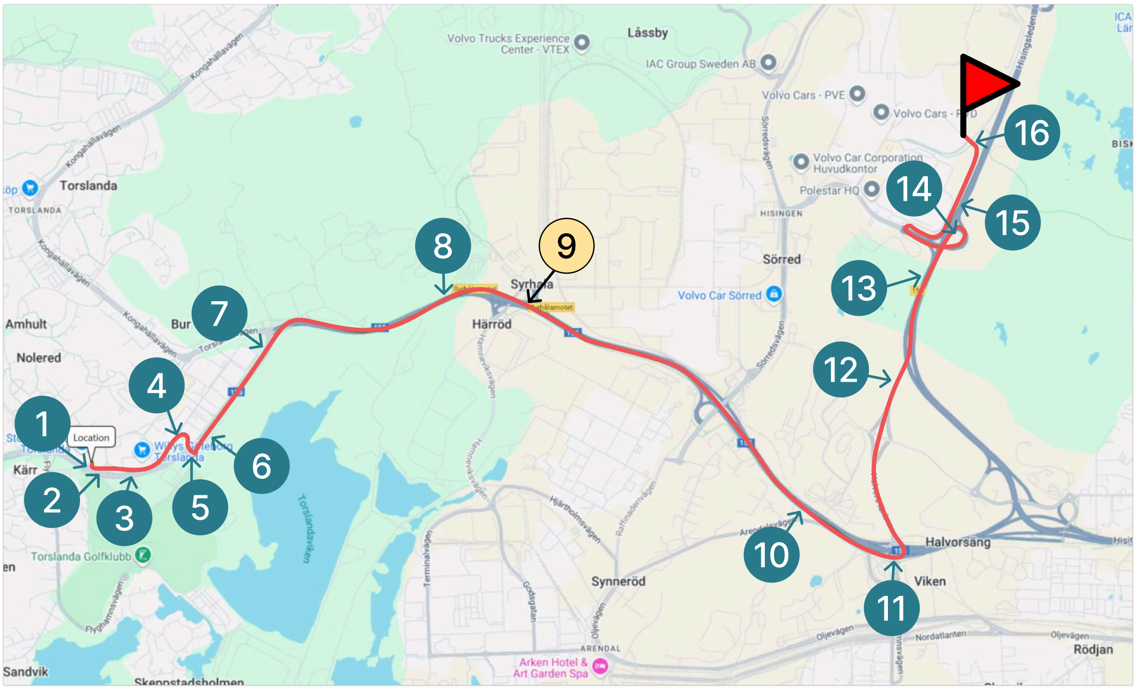


Figure 4.11: Route 2 with 16 marked navigation changes on the HUD starting from Gamla Flygplatsvägen to Volvo PVH.

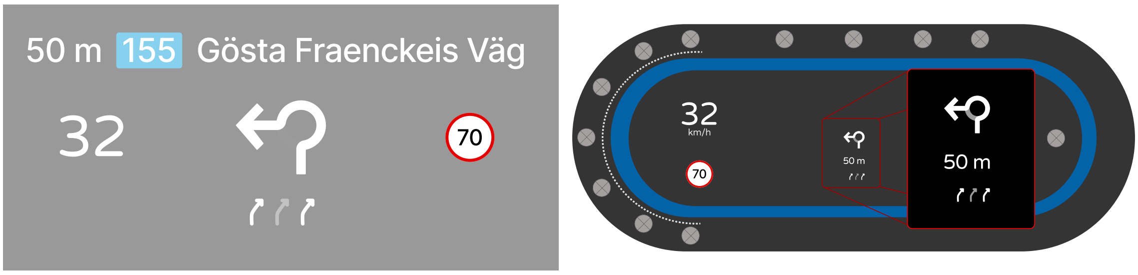
4.2.5 Display Configurations

In the study, navigation mode and UI settings in all displays were standardised across all participants. During the task description session, participants were instructed not to modify any configurations to maintain consistent experimental conditions.

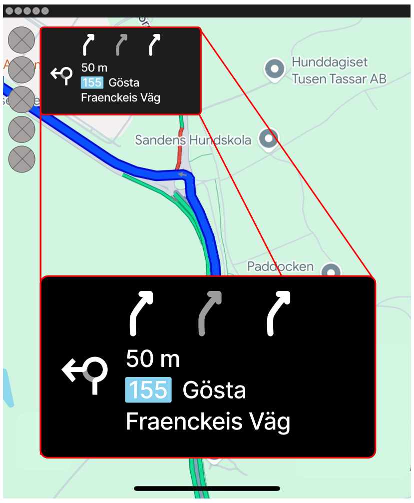
For the HUD, participants were able to adjust the position (e.g., height and rotation) and brightness (see Appendix B.7). However, the layout of the displayed information could not be changed due to system limitations; the HUD content is illustrated in Figure 4.12a. The DIM used in the study offered both a map view and turn-by-turn arrows; to ensure consistency between HUD and non-HUD conditions, the turn-by-turn arrows mode was selected (see Figure 4.12b). For the CSD, although alternative views such as satellite imagery were available, the default Google Maps view was used, based on insights from the pilot study (see Figure 4.12c).

4.2.6 Sample Size and Participant Recruitment

Before recruiting participants for the study, a power analysis was calculated using a formula recommended for cases where no prior estimate of variability was available (Sauro and Lewis, 2016). To achieve this, the typical definition of critical difference (d) was redefined in terms of effect size (e), expressed as a fraction of standard deviation (s).



(a) HUD. (b) DIM.



(c) CSD.

Figure 4.12: Configuration on displays.

$$\begin{aligned} e &= \tfrac{d}{s} \\ d &= e(s) \end{aligned} \tag{4.1}$$

As a rule-of-thumb, Cohen (1977) suggested using the values of 0.2, 0.5 and 0.8 for small, medium, and large effects. In this study, the aim was to detect small-to-medium effect sizes, defined as $e = 0.33$. A desired confidence of 95% was selected, corresponding to $z = 1.965$. These values were then applied to the power analysis formula proposed by Sauro and Lewis (2016), resulting in the following calculation:

$$n = \frac{z^2 s^2}{d^2} = \frac{z^2 s^2}{(e(s))^2} = \frac{1.96^2\,(s^2)}{0.33^2\,(s^2)} \approx \frac{3.8}{0.11} \approx 35 \tag{4.2}$$

The calculation indicated that a minimum of 35 participants would be required to achieve 95% confident in detecting effects equal to or greater than one-third of a standard deviation. This sample size was chosen to ensure moderate statistical power and to allow for appropriate assessments of normality.

To recruit participants, convenience sampling was employed by placing flyers around the VCC Torslanda office and distributing internal emails, inviting employees to volunteer for the study (see recruitment advertisement in Appendix B.1). The study focused on drivers who were legally eligible to drive personal vehicles in Sweden, as it involved real-road driving rather than a simulation-based environment. To minimise safety concerns and facilitate approval for the study, only employees from VCC were recruited. The recruitment procedure aligned with VCC's in-vehicle experiment guidelines as well as the ethical standards and approval processes for human-subject research outlined by Robson and McCartan (2016).

To ensure a diverse participant pool, no restrictions were placed on demographic characteristics, such as gender, age, driving experience, or prior HUDs familiarity. The only exclusion criterion was the use of polarised sunglasses[1]. Furthermore, to minimise potential bias related to participants' interest in HUDs, the recruitment advertisement did not explicitly mention HUD, thereby minimising the likelihood of self-selection bias. As a result, a total of 35 individuals were successfully recruited, but five dropped out before participating, yielding a final sample size of 30 participants.

4.2.7 Pilot Study

Pilot studies are valuable for identifying potential issues in research design, methods, or procedures before the main study, as mentioned in Section 3.9. In this study, pilot studies were conducted with four volunteers from VCC, including two females and two males, to ensure the feasibility and reliability of the study, as suggested by van Teijlingen and Hundley (2002).

The procedure closely followed the one illustrated in Figure 4.13, with a few minor adjustments. Results from the pilot studies revealed the need for compromises in camera calibration and adjustments in the procedure order, primarily due to time constraints and technical challenges. The consent form, participant instructions, questionnaires, and interview questions were refined during this phase to improve clarity and precision. The NASA-TLX questionnaire (see Appendix B.4) caused some confusion among the pilot study participants. As a result, an explanatory step was introduced when distributing the questionnaire. Additionally, a follow-up interview question was also included to gather participants opinions about the questionnaire itself.

4.3 Execution

The duration of the experimental procedure was approximately 60 minutes, with each step outlined in the flow chart in Figure 4.13. The flowchart represents a counterbalanced A/B test design, in which participants were randomly assigned to either the variant A or B. This design ensured that all participants experienced both

---

[1] Polarised sunglasses can interfere with HUD visibility, as the polarised lenses may block the display's light, making it difficult to see the HUD image.

conditions—HUD and non-HUD—while minimising potential biases due to the order of exposure.

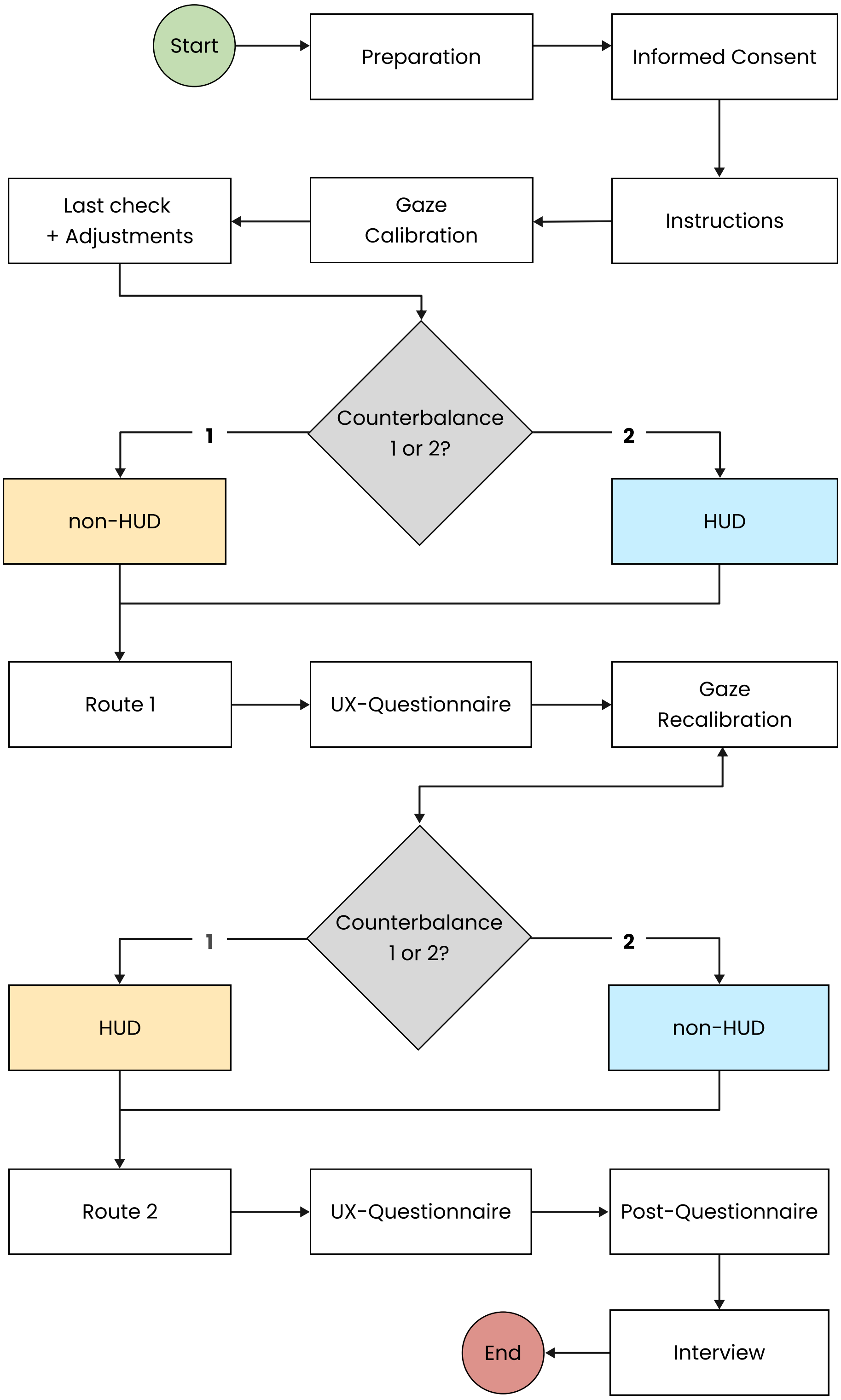


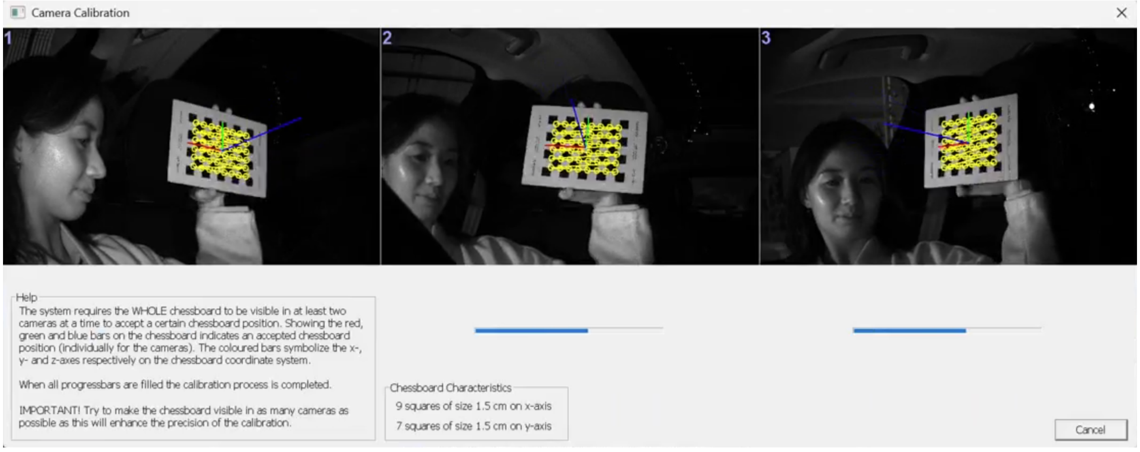
Figure 4.13: Flowchart of the experimental procedure.

Each test session involved a total of three individuals in the vehicle: the participant in the driver's seat, Researcher 1 in the front passenger seat, and Researcher 2 in the rear seat. Researcher 1 was responsible for leading the session by providing instructions, conducting the interview, responding to unexpected situations, and observing both the participant and traffic for any deviations from the planned procedure. Researcher 2 was in charge of the technical aspects, including device setup, participant gaze calibration, recording management, system monitoring, initiating test calls, and documenting interview responses.

The participants were randomly assigned to one of two counterbalanced groups, referred to as *CB1* and *CB2*, respectively. In Route 1, CB1 began the drive without the HUD, and then used the HUD in Route 2. Conversely, CB2 started with the HUD in Route 1 and completed Route 2 without it. After each route, participants completed a UX-Questionnaire (described in Section 4.3.6) to evaluate their experiences. The study concluded with a Post-Questionnaire (see Section 4.3.7) and a semi-structured interview (see Section 4.3.8), where participants provided qualitative insights into their experiences through the two display conditions. This counterbalanced design ensures that order effects (e.g. fatigue, nervousness) did not bias the results, allowing for a more accurate comparison between conditions.

4.3.1 User Study Preparation

Data collection began with in-vehicle preparations before the arrival of each participant. Firstly, the driver's seat was adjusted to a neutral position that accommodated most participants and allowed for effective eye-tracking camera calibration. Calibration was performed using a physical chessboard placed in front of the eye-tracking cameras. The process involved moving the chessboard until all progress bars were fully filled, as illustrated in Figure 4.14. A calibration result was considered acceptable only when all three cameras displayed green numerical indicators, indicating a successful calibration (see Figure 4.15).



Help

The system requires the WHOLE chessboard to be visible in at least two cameras at a time to accept a certain chessboard position. Showing the red, green and blue bars on the chessboard indicates an accepted chessboard position (individually for the cameras). The coloured bars symbolize the x-, y- and z-axes respectively on the chessboard coordinate system.

When all progressbars are filled the calibration process is completed.

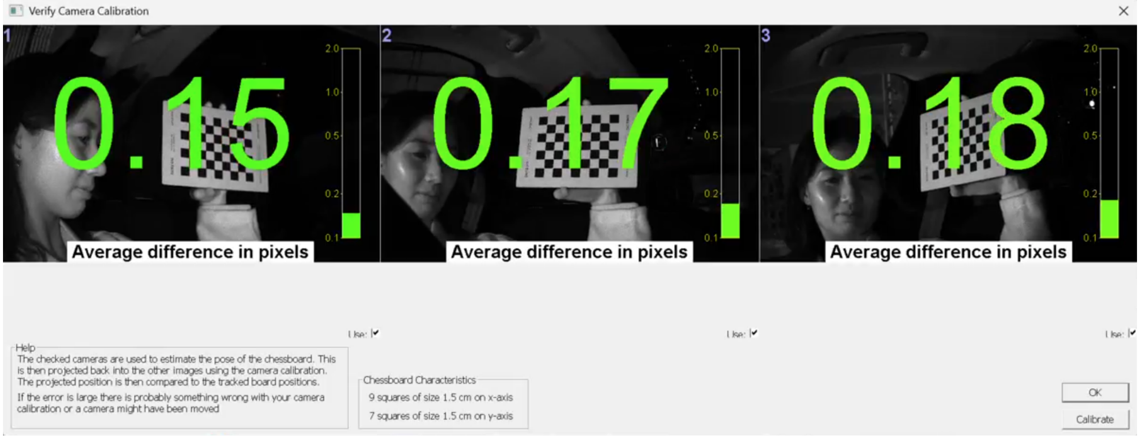
IMPORTANT! Try to make the chessboard visible in as many cameras as possible as this will enhance the precision of the calibration.

Chessboard Characteristics

9 squares of size 1.5 cm on x-axis

7 squares of size 1.5 cm on y-axis

Figure 4.14: Calibration process using a chessboard to define the spatial reference for the eye-tracking cameras.



Help
The checked cameras are used to estimate the pose of the chessboard. This is then projected back into the other images using the camera calibration. The projected position is then compared to the tracked board positions.

If the error is large there is probably something wrong with your camera calibration or a camera might have been moved

Chessboard Characteristics
9 squares of size 1.5 cm on x-axis
7 squares of size 1.5 cm on y-axis

Figure 4.15: Green numbers on all three cameras indicate successful calibration.

After the camera calibration, all three displays in the vehicle were checked to ensure they were configured according to the designated settings (see Section 4.2.5). Depending on each participant's assigned counterbalanced group, the HUD was either activated or deactivated at the beginning of the session. The DIM and CSD were consistently turned on, independent of counterbalance number. On the DIM, the turn-by-turn arrows mode was selected, and the CSD was set to the default Google Maps perspective view.

Lastly, all the necessary components were verified before starting the session. This included establishing a Bluetooth connection between the test vehicle and a test phone to enable phone calls during driving. The setup also involved confirming that the eye-tracking recording software, iMotions, was properly linked to all necessary input sources. These included Webcam 1 (Scene Camera), Webcam 2 (Environment Camera), eye-tracking system (Smart Eye Pro), and GPS. Additionally, weather conditions were documented throughout the study to support potential analysis in later stages.

4.3.2 Informed Consent

As participants entered the test vehicle, they were provided with a printed consent form (see Appendix B.2) to review and sign before proceeding further. Following the ethical guidelines of the Declaration of Helsinki (World Medical Association, 2024) and General Data Protection Regulation (GDPR) (European Union, 2016), the consent form briefly outlined the researchers' background, the study's objectives and procedures, data handling practices, participants' right to withdraw at any time, and relevant contact information.

To protect participant anonymity, each individual was assigned a participant ID, formatted as "P" followed by a number. Participants retained a copy of the consent form for their records, while the researchers kept the portion containing the participant's name, signature, date and assigned ID. The inclusion of the assigned ID ensures that individual data could be accurately identified and removed if a withdrawal occurred.

4.3.3 Study Instructions

After the participants signed the consent forms, they were given a printed sheet describing the task. As the goal was to capture natural driving behaviour, minimal instructions and prohibitions were given. To ensure all participants received the same information, each individual was guided through the task description before beginning the drive. The task description sheet given to participants included both a flowchart of the experimental procedure and a written explanation (see Appendix B.3). Participants were instructed to follow the navigation that appeared on the in-vehicle displays. They were also informed in advance about incoming phone calls during the drive, which they were expected to decline.

To ensure comfort, participants were encouraged to adjust the seat, mirrors, HUD's position and brightness (depending on the assigned counterbalanced group), air conditioning, and sunshades. However, they were asked to refrain from changing any settings related to the displays or cameras. For the HUD adjustment, an example of an appropriate positioning was shown to the participants, see Appendix B.7. During this phase, Researcher 2 adjusted the position and exposure of the Scene Camera based on factors such as weather conditions (lighting condition in particular), participant height, and individual HUD positioning preference.

Lastly, participants were reminded that safety was the first priority and they should not take any risky actions to complete the given tasks. To maintain a consistent testing environment and minimise disturbances, they were notified that the researchers would remain silent through the drive unless assistance was needed.

4.3.4 Gaze Calibration

After receiving the instructions, participants proceeded with gaze calibration. To ensure high eye-tracking data quality, participants were asked to secure any hair obstructing their eyes using hair ties or clips. Inside the test vehicle, there were four calibration points that participants were asked to look at for a duration of 3 seconds without blinking (see Figure 4.8). Calibration results were accepted only within a certain threshold (under five degrees); otherwise, the calibration was redone or invalidated. This threshold was established in consultation with project partners to ensure optimal eye-tracking quality for a naturalistic study.

4.3.5 Driving Sessions

Before participants began following the navigation for Route 1, a final check of all test equipments, settings, and seat adjustments was conducted during a short drive of approximately 300 meters. Once everything was confirmed in place, the eye-tracking recording started, and participants were once again reminded that silence from the researchers would be maintained throughout the drive, but they were encouraged to ask any questions related to the test.

When driving on Route 2, all participants were required to complete a recalibration of the eye-tracking system to maintain high data quality. The recalibration procedure was the same as the first gaze calibration. Afterwards, eye-tracking recording resumed and participants continued on Route 2, returning to the starting point of the test.

4.3.6 UX-Questionnnaire

After arriving at the designated destination and parking the vehicle, participants were given a UX-Questionnaire to complete. The UX-Questionnaire is essentially NASA-TLX, which was described in Section 3.6.2. NASA-TLX is a widely used tool for evaluating subjective workload across different contexts (Hart, 2006). For this study, minor modifications were made to the descriptions of each NASA-TLX subscale component to better align with the specific tasks involved (see Appendix B.4). These adjustments were carefully made considering preservation of the original NASA-TLX structure and wording as much as possible to maintain the validity, sensitivity, and reliability, as addressed by Hart (2006). A paper format of the NASA-TLX was used to reduce cognitive load during completion, following recommendations by Noyes and Bruneau (2007).

4.3.7 Post-Questionnaire

After arriving at the final destination, participants were asked to complete the UX-Questionnaire (NASA-TLX) along with a Post-Questionnaire on a tablet. The Post-Questionnaire was created in Microsoft Forms, consisted of 14–16 questions in total (see Appendix B.5). The first two questions, regarding participant ID and counterbalance number, were filled out by the researchers. Participants with no prior HUD experience answered questions 3 to 14, while those with prior HUD experience answered questions 3 to 16, which included how frequently they used a HUD and which models they had used. The Post-questionnaire covered questions about demographics, driving experience, perceived differences between the test routes, familiarity with the navigation service (Google Maps) used during the test, as well as experience and preferences related to HUD.

4.3.8 Interview

As suggested by Wadsworth (2020), the interview followed a consistent and standardised set of questions to enable comparability of the data during analysis (see Appendix B.6). Four questions from the Post-Questionnaire were selected: questions 10, 11, 12, and 16 (this would be question 14 if there was no prior HUD experience). This served as the basis for the interview by gathering further understanding of participants' experiences. The reason for using the same questions for both the Post-Questionnaire and the interview was to give participants time to thoughtfully consider their answers while enabling the collection of both quantitative and qualitative data. Reviewing the Post-Questionnaire answers gave participants an opportunity to reflect and elaborate on the reasoning behind their thoughts and choices. Additionally, participants were asked for their opinions on the UX-questionnaire as well as their general thoughts about the study.

4.4 Data Preprocessing and Validation

Before analysing the collected data, data preprocessing and validation were carried out to ensure the final dataset was valid and met an acceptable level of quality.

4.4.1 Normality Test

To determine the appropriate statistical tests for the analysis, Shapiro-Wilk tests were conducted to assess the normality of the data. Since the results indicated deviations from a normal distribution, non-parametric alternatives were selected in place of parametric tests such as the t-test.

4.4.2 Eye-tracking Data Quality

To validate the quality of the eye-tracking data, key metrics such as robustness, precision, and accuracy were reported. According to Vehlen et al. (2021), these indices are fundamental for assessing the suitability of eye-tracking data.

- **Robustness** was assessed using the *Exposure* metric provided by iMotions, which reflects the percentage of valid gaze samples relative to the expected number. A lower exposure rate indicates greater data loss, often due to eye closure or movement out of the headbox.
- **Precision** was measured by Standard Deviation (SD) of gaze points for both the left and right eyes.
- **Accuracy** was determined from the calibration results for each eye. As noted in Smart Eye AB (2002), “the accuracy will depend on various parameters such as distance from the cameras, distance and position of the object you look at and focal length of the lenses and individual differences”.

4.5 Data Analysis

This section outlines the objectives of the collected data, both quantitative and qualitative, and the methods used for the analysis. The quantitative data were analysed as a complete eye-tracking recording and overall workload, but also zoomed in into separate segments of the recording and dimensions of the workload. To examine whether there was any impact influenced by the HUD, statistical tests were applied. How the test were carried out, which tests, and why they were chosen are motivated and justified. The process of analysing the qualitative data from interviews was also described.

4.5.1 Aim and Objectives

The analysed quantitative data included eye-tracking metrics and NASA-TLX scores. AOI analyses were performed on the eye-tracking data. The primary eye-tracking metric used was gaze dwell time, which reflects the duration drivers spent gazing at different world model components, for instance, the displays. To account for varying exposure times across participants, the percentage of gaze dwell time was chosen instead of the absolute values, following the suggestion by B. Albert and Tullis (2013). Outliers were identified and removed to ensure the validity and reliability of the analysis.

For NASA-TLX analysis, as mentioned in Section 3.6.2, the sensitivity of the tool may vary depending on whether the Weighted TLX or Raw TLX is used, as determined by the specific context of each study (Hart, 2006). To address this, both workload scoring methods were compared in the study. Furthermore, as suggested by Galy et al. (2018), the subscales of NASA TLX should be taken into account and further analysed rather than relying solely on the overall workload score. This approach is particularly important because the performance of complex tasks such as driving can be understood in terms of three categories of mental workload: intrinsic, extraneous and germane load. Examining each subscale separately has a diagnostic value by allowing for identification of the main sources of workload (Hart, 2006), offering clearer insights than a single overall score.

To visualise and facilitate the interpretation of the data, descriptive statistics were used. Subsequently, non-parametric statistical tests were applied to derive valid inferences, given the characteristics of the dataset. This analytical approach led to the identification of significant differences in attention and gaze behaviour between the two display conditions.

To evaluate how system configuration and driving context influence drivers' attention, overall and event-based analyses were conducted on the eye-tracking data. The overall analysis captured overarching patterns of visual attention throughout the driving task, while the event-based analysis enabled a more fine-grained examination of participants' gaze behaviour in response to specific in-task events (e.g., following navigation, and reacting to phone calls). Event 0 (reacting to phone calls) was pre-defined during the study design phase due to its high potential to distract drivers as a typical secondary task. Events 1 to 6 were selected after data collection, based on noticeable changes in the HUD display and observed participant behaviours. These events were chosen to highlight moments likely to influence visual attention due to changes in task demands or interface interaction. A detailed summary of the selected events is provided in Table 4.2, with timing of the navigation change also illustrated on the map in Figure 4.16 and Figure 4.17.

Event	Route	Road Scenarios		HUD Information		
		No.	Content	Traffic Signs	Maneuvers	Other (e.g., text)
0	R1	9	Highway	-	-	Name + Phone numbers
	R2	9	Highway	-	-	Name + Phone numbers
1	R1	5	Regional road	Speed limit (80)	fork-left + LP	Highway No. + Road name
2	R1	10	Roundabout	Speed limit (80)	roundabout-left	Distance + Highway No. + Road name
		11	Roundabout	Speed limit (80)	roundabout-exit + LP	Distance + Highway No. + Road name
3	R1	12	Roundabout	Speed limit (70)	roundabout-straight + LP	Distance + Highway No. + Road name
		13	Roundabout	Speed limit (70)	roundabout-exit + LP	Distance + Highway No. + Road name
		14	Regional Road	Speed limit (70)	fork-right + LP	Distance + Road name
		15	Roundabout	Speed limit (50)	roundabout-left	Distance + Road name
		16	T-crossing	Speed limit (50)	turn-right	Distance + Instruction
4	R2	3	Roundabout	Speed limit (50)	roundabout-right	Distance + Road name
		4	Roundabout	Speed limit (50)	roundabout-left + LP	Distance + Highway No. + Road name
		5	Roundabout	Speed limit (70)	roundabout-exit + LP	Distance + Highway No. + Road name
		6	Highway	Speed limit (70)	fork-right + LP	Distance + Instruction
		7	Highway	Speed limit (70)	merge + LP	Distance + Highway No. + Road name
5	R2	10	Roundabout	Speed limit (80)	roundabout-left	Distance + Road name
6	R2	12	Regional road	Speed limit (80)	merge	Distance + Road name
		13	Regional road	Speed limit (80)	fork-right + LP	Distance + Road name

Table 4.2: Event selection based on road scenarios. LP = Lane Placements.

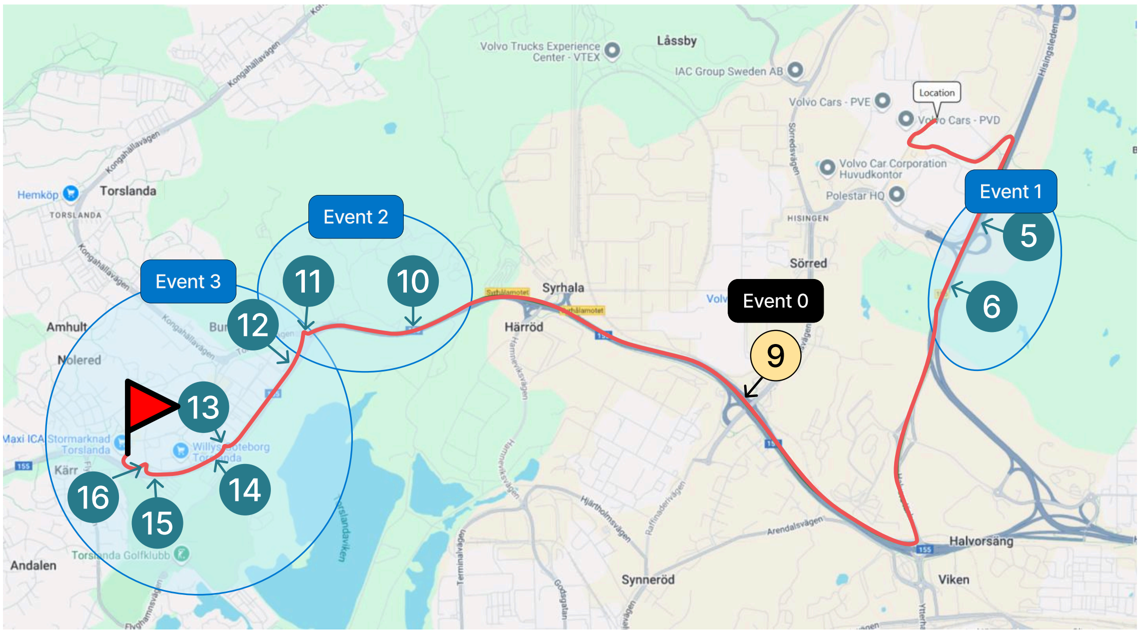


Figure 4.16: Route 1 with events 0, 1, 2, and 3 marked on the map.

During Route 1, three events were selected for analysis, see Figure 4.16. In Event 1, some participants missed the navigation cue to “fork left” from the regional road, resulting in a missed turn. This event was selected for further analysis to better understand why some participants missed the navigation cue. Event 2 involved a long, uninterrupted drive on the highway with minimal changes to the HUD, until participants encountered a roundabout where they had to take a left turn to exit the highway. In Event 3, participants received frequent navigation updates on the HUD as they navigated through three roundabouts.

For Route 2, another three events were selected, see Figure 4.17. Event 4 mirrored Event 3 but occurred in the opposite direction, navigating through three roundabouts. Event 5 required participants to exit the highway by changing lanes to the

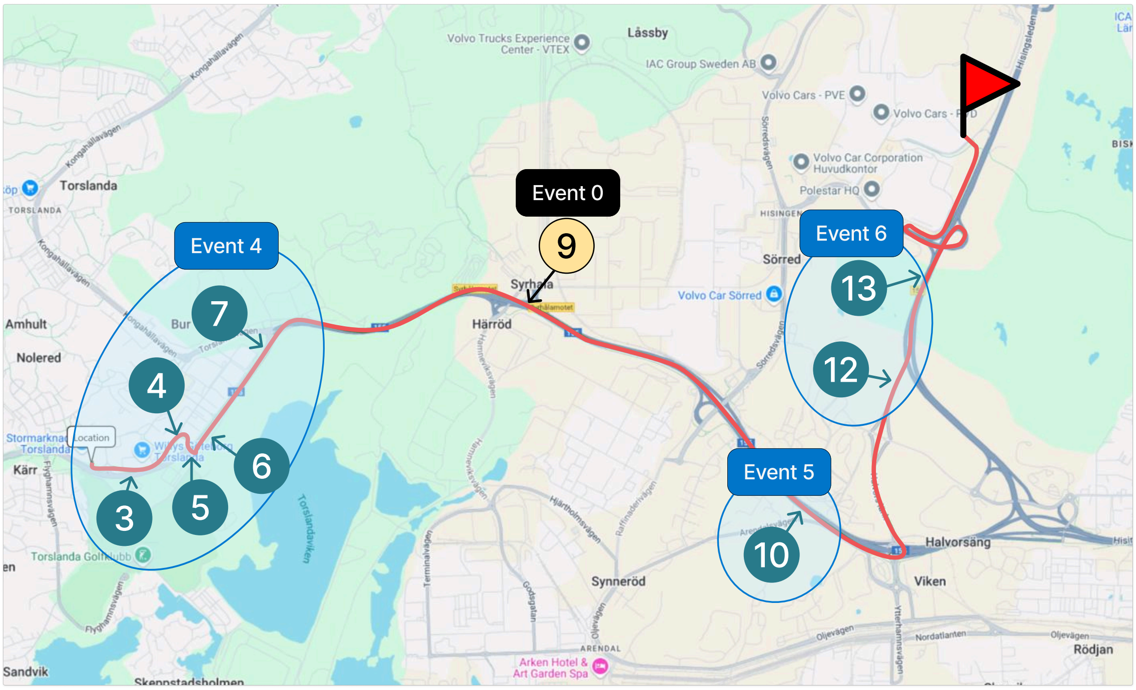


Figure 4.17: Route 2 with events 0, 4, 5, and 6 marked on the map.

right and crossing a bus lane, followed by a left turn at a roundabout. Event 6 was described by participants as stressful, involving a merge onto a high-speed regional road and unexpected right lane changes to exit it. This event was selected because some participants missed the navigation cues or appeared visibly stressed.

4.5.2 Selection of Statistical Methods

To examine the effects of HUD conditions on driving behaviours, including attention and cognitive load, statistical analyses were conducted using both Wilcoxon signed-rank test[2] for paired-sample (i.e., dependent sample) comparisons, and Mann-Whitney U-test[3] for independent sample comparisons. These tests were chosen based on the experimental design and the nature of the comparisons. The aim was to identify where there were significant differences in eye-tracking data quality, participants' visual attention to the road, and their cognitive load across the different experimental conditions.

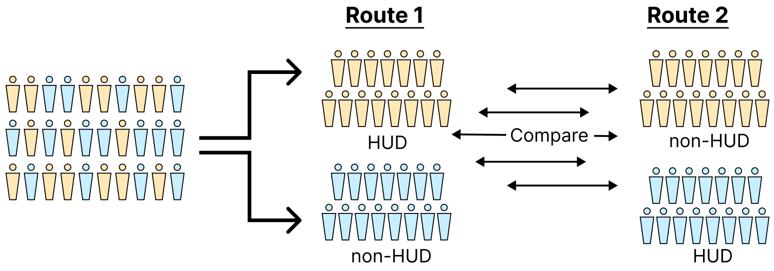
Since each participant experienced both experimental conditions (HUD vs non-HUD), comparisons were primarily focused on identifying differences between these settings. When comparing conditions where the same participants experienced both variables (HUD and non-HUD), the data were treated as dependent samples and analysed using Wilcoxon-tests. This within-subjects approach accounts for individual variability across conditions. For comparisons involving conditions where

---

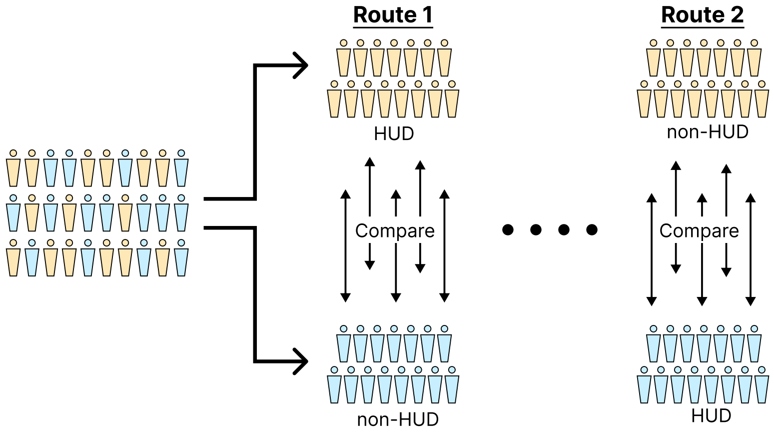
[2] Wilcoxon signed-rank test is a non-parametric test to determine whether two dependent groups differ significantly from each other.

[3] Mann-Whitney U-test is a non-parametric test to compare differences between two independent groups.

different participants experienced only one variable (HUD or non-HUD), the data were treated as independent samples. In such cases, Mann-Whitney U-tests were applied, as they are well-suited for handling differences in sample size and variance due to prior data validation. Based on these criteria, two types of comparisons were conducted (see Figure 4.18).



(a) Within-subject comparison.



(b) Between-subject comparison.

Figure 4.18: Illustrations of two types of comparisons across conditions (HUD vs non-HUD).

- **Within-subject comparison** (see Figure 4.18a): A within-participant comparison of HUD vs non-HUD conditions to assess the general effect of HUD presence. This approach was applied to eye-tracking data (including data quality and overall analysis) and NASA-TLX scores (both weighted and raw).
- **Between-subject Comparison** (see Figure 4.18b): A between-participant comparison focused on selected driving events, aimed at examining the specific effects introduced by the HUD in context. This was used exclusively for event-based analysis of eye-tracking data.

The dependent variables for each comparison are summarised in Table 4.3. Eye-tracking data quality was assessed using five metrics, as explained in Section 4.4.2. To evaluate participants' attention on the road, the percentage of *Dwell Time* was used as the primary gaze-based metric. *Dwell Time* refers to the average duration the participants gazed at the given AOI during the time it was active.

Data Source	**Analysis Object**	**Dependent Variables**	**Variable Units**
Eye-tracking	Data Quality	Left Eye Calibration Accuracy	Degree
		Right Eye Calibration Accuracy	Degree
		Left Eye Calibration Precision (SD)	Degree
		Right Eye Calibration Precision (SD)	Degree
		Exposure (valid data) ratio	Percentage
	Overall analysis	Road Dwell Time	Percentage
		CSD Dwell Time	Percentage
		DIM Dwell Time	Percentage
	Event-based Analysis	Road Dwell Time	Percentage
		CSD Dwell Time	Percentage
		DIM Dwell Time	Percentage
NASA-TLX	Weighted TLX	Weighted Overall Workload	Score
		Weighted Subscales (MD, P, T, P E, F)	Score
	Raw TLX	Raw Overall Workload	Score
		Raw Subscales (MD, P, T, P E, F)	Score

Table 4.3: Overview of data sources, analysis objects, and dependent variables using the independent variable of display conditions (HUD vs non-HUD). Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (P), Effort (E), and Frustration (F).

In this study, $\text{Road}_{\text{DwellTime}(\%)}$ was defined as the percentage of participants' gazes located outside the three display interfaces (CSD, DIM, and HUD). This definition was chosen to capture participants visual attention toward the broader driving environment, including essential components such as the road ahead, side mirrors, and rear-view mirror. These areas are critical for maintaining SA and completing the primary task of safe driving. By subtracting the gaze dwell time allocated to in-vehicle displays, the measure captured attention directed toward the external driving environment, which is considered indicative of real-world driving behaviour in the context of this study. Accordingly, $\text{Road}_{\text{DwellTime}(\%)}$ was calculated using the following formula:

$$\text{Road}_{\text{DwellTime}(\%)} = 100\% - (\text{CSD}_{\text{DwellTime}(\%)} + \text{DIM}_{\text{DwellTime}(\%)} + \text{HUD}_{\text{DwellTime}(\%)})$$

To further investigate how the presence or absence of the HUD affected participants attention to other displays, the *Dwell Time* metric was analysed for both the CSD and DIM. This analysis enabled the identification of potential shifts in attention patterns across different in-vehicle displays caused by the HUD.

Cognitive load was measured using both the Weighted TLX and Raw TLX scores. These provided an overall workload estimate along with six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration.

Each comparison was analysed using Python (version 3.13.3), with results reported in terms of means, Standard Deviations (SDs), test statistic (W or U), and p-values. A significance level of 0.05 (i.e., 5%) was used to determine statistical significance.

4.5.3 Thematic Analysis

For qualitative data collected through interviews, TA was conducted following the steps described in Section 3.10.4. The participants' quotes were extracted or paraphrased and sorted into different codes. These codes were then grouped into broader clusters referred to as themes, which captured the essence of the interviews. The initial codes and themes were generated using a customised GPT model for the initial coding of thematic analysis, as suggested by Turobov et al. (2024). The model generated codes with related quotations or paraphrases of the interview responses, which were subsequently synthesised into themes. The themes were then used to meaningfully interpret the quantitative findings and to understand how the presence of a HUD affected driver behaviour.

The use of the custom GPT model was a conscious choice, well aware of the compromised accuracy due to the nature of the GPT outputs as warned by Turobov et al. (2024). Thus, the use of GPT was limited to only the initial coding of questions 12 and 16 (see Appendix B.6). These generated outputs were manually refined by reviewing the interview responses, adding additional quotations or paraphrases to existing codes, modifying or adding new code and theme labels, and verifying the generated codes. The manual work was essential to avoid information taken out of context and keep the reliability of the analysis. After the extraction of the codes and themes was finalised, the relationships between the themes were visualised with a thematic map to more clearly conceptualise the findings. This process was carried out in Figma, a versatile software tool suitable for UI design, prototyping, and collaborative activities such as diagramming and brainstorming.

5

Results

This section presents the results of both quantitative and qualitative data collected and analysed in the study. Participants' demographic, background, and experience information were gathered through post-questionnaires. Eye-tracking data related to visual attention were analysed using both overall and event-based approaches. Cognitive load was assessed through the UX questionnaire, with analyses conducted both on the overall workload score and each individual subscale. Finally, the qualitative interview responses were thematically sorted and analysed.

5.1 Participant Background, Demographics and Experience

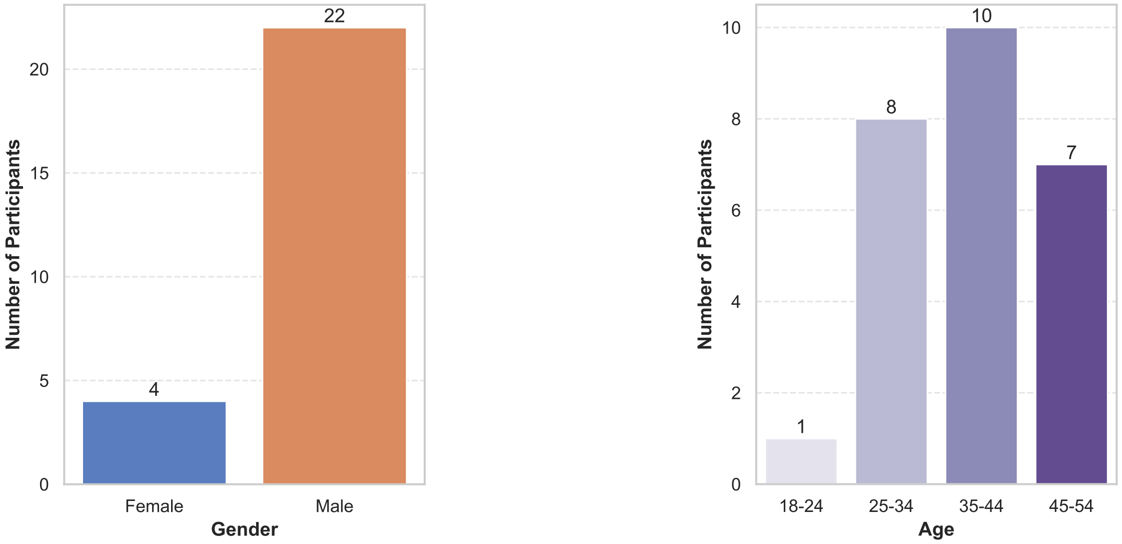
A total of 30 participants employed at VCC in Sweden were voluntarily recruited for this study. To ensure the validity of the analysis, data from 4 participants were excluded due to reasons outlined in Table 5.1. The excluded data include eye-tracking recordings, questionnaire responses, and interview transcripts, which were intended for the subsequent aggregated analysis.

Participant No.	CB No.	Data Removal Reason
P2	2	Route 1 was altered due to unclear bus lane markings, affecting route consistency.
P8	2	Both Route 1 and Route 2 deviated significantly as the participant failed to follow navigation instructions.
P9	1	An unexpected traffic accident caused by external factors occurred during Route 2, necessitating a route change.
P16	2	Substantial deviations occurred in both routes due to the participant missing navigation.

Table 5.1: Data removal reasons for the four excluded participants (CB = Counterbalance).

After data validation, 26 participants were included in the final aggregated analysis. Among them, 4 were identified as female and 22 as male, with age ranging from 18 to 54 (see Figure 5.1). Participants were also asked about their driving experience. This included how long they had held a driving licence and how many hours they drive per week. Most of the participants (20) had more than 10 years of driving experience. Regarding weekly driving, more than half of the participants (14) reported driving

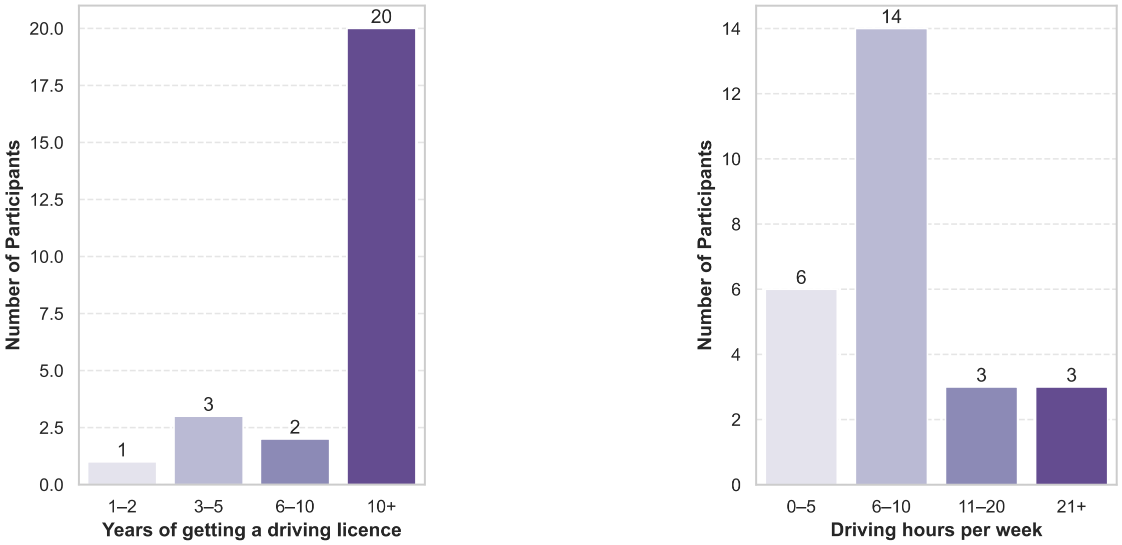
6–10 hours per week, followed by 6 participants driving 0–5 hours, and 3 each driving 11–20 hours or more than 21 hours per week (see Figure 5.2).



(a)

(b)

Figure 5.1: Participants' gender and age distribution.



(a)

(b)

Figure 5.2: Participants' driving experience.

Participants with normal or corrected-to-normal vision were recruited for the study. This included individuals who wore glasses, contact lenses, or had undergone LASIK[1] surgery. As shown in Figure 5.3, most participants reported normal or fully corrected vision. However, 3 out of 26 participants indicated that their vision was not fully

---

[1] LASIK: Laser-Assisted In Situ Keratomileusis, a form of laser eye surgery used to correct vision.

corrected to normal. Follow-up conversations confirmed that these cases did not compromise driving safety.

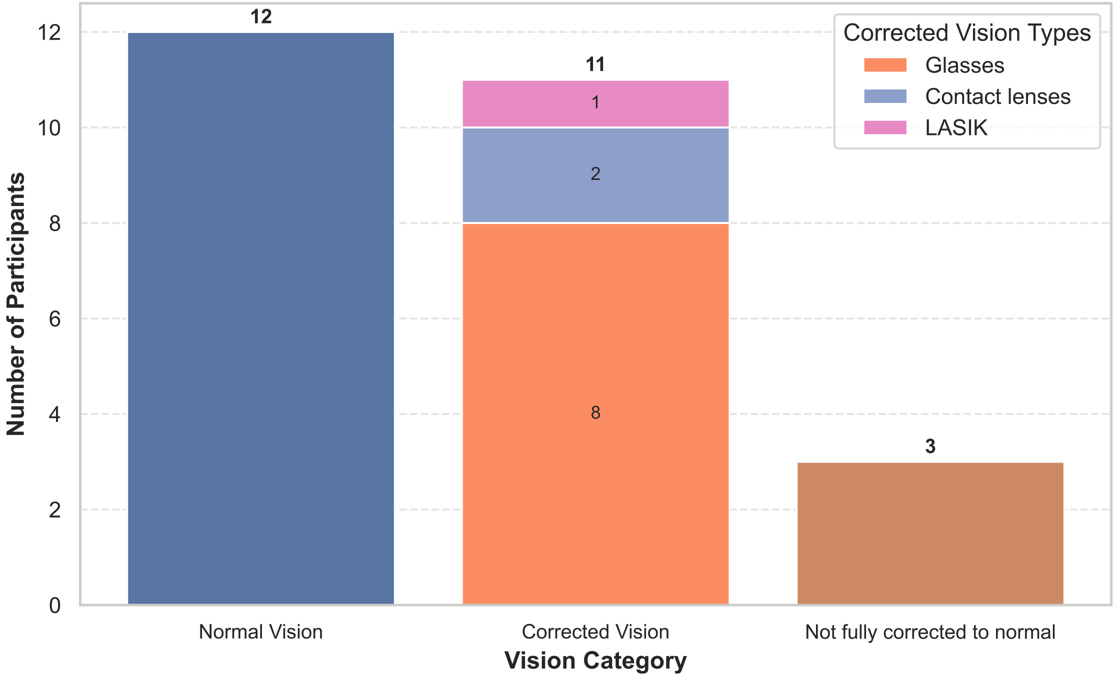
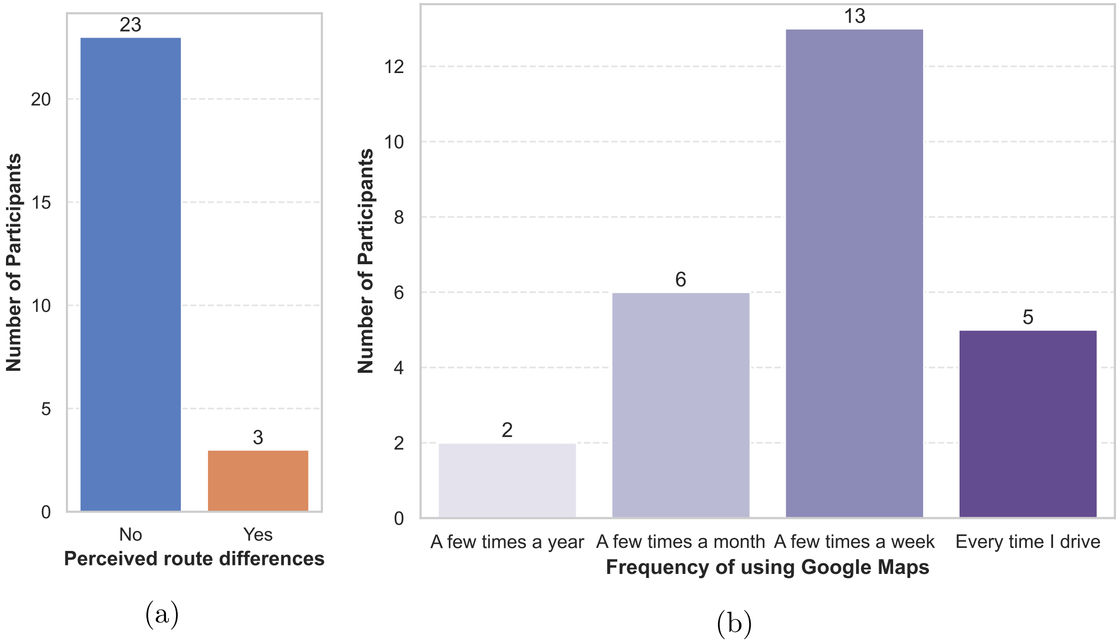


Figure 5.3: Participants' vision distribution.

To assess whether participants' familiarity with the selected test routes and navigation tools affected the results, they were asked if they noticed any differences between the two routes (e.g., traffic density, weather change), and how frequently they used Google Maps. As shown in Figure 5.4a, most participants (22) reported no perceived significant differences between the routes, while 3 participants indicated that some variations in route conditions were noticeable, and the reasons can be seen in Table 5.2. Regarding navigation habits, Figure 5.4b shows participants' frequency of Google Maps usage. During the interviews, participants in general stated that they were familiar with the symbols and information presented on Google Maps. Some participants also mentioned a personal preference for alternative navigation services, such as Waze, Baidu Maps, or Gaode Maps.

Participants were asked to rank the display locations where they typically checked navigation information while driving, based on their subjective experiences. The distribution of these reported behaviours is shown in Figure 5.5.

Participants were asked about their prior experience with and preferences toward HUDs. As illustrated in Figure 5.6a, 20 participants reported having some prior experience using HUDs, while 6 had none. However, as shown in Figure 5.6b, only a small number (3 participants) reported using HUDs regularly during everyday driving. Interview responses revealed that most of those with HUD experience had encountered the technology through work-related activities, such as vehicle testing, rather than through personal use—only a few had HUDs installed in their own vehicles.



(a) (b)

Figure 5.4: Perceived differences of test routes and familiarity of Google Maps navigation services.

Participant No.	**CB No.**	**Reason for Perceived Route Differences**
P10	2	P10 found Route 1 slightly easier, as they realised the destination and were more aware of the direction. In Route 2, a temporary construction site required a detour into the bus lane.
P24	2	P24 noted that a key difference was the absence of the HUD in one of the routes, resulting in fewer elements to attend to. They also described encountering unfamiliar traffic conditions in Route 2, particularly at a complex roundabout with three exits and two lanes.
P30	2	P30 observed more frequent lane changes on the highway during Route 2, and higher traffic density including the presence of more buses and trucks.

Table 5.2: Reported reasons from participants who perceived differences between the two test routes (CB = Counterbalance).

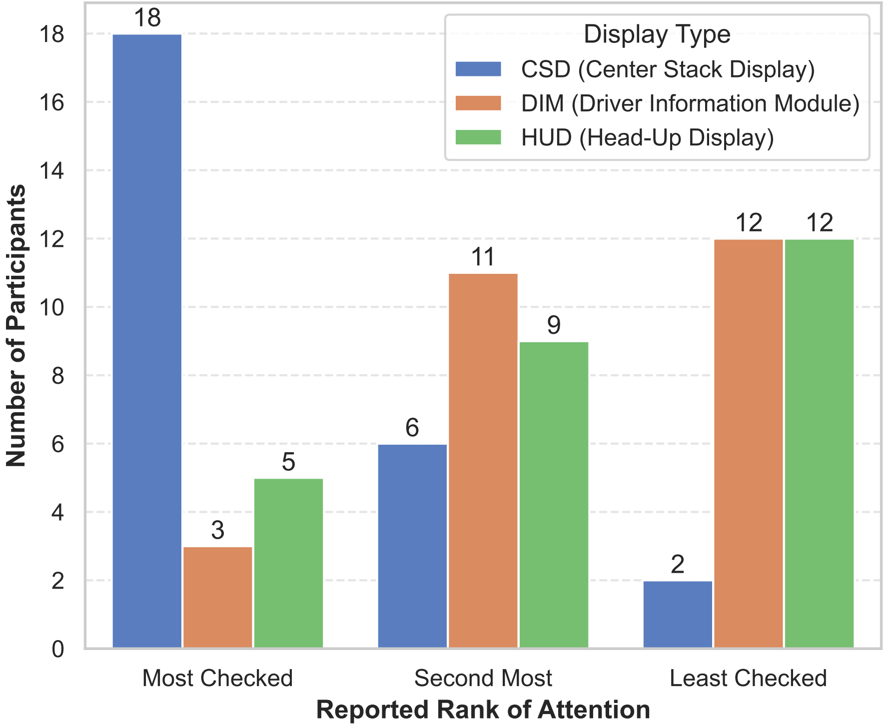
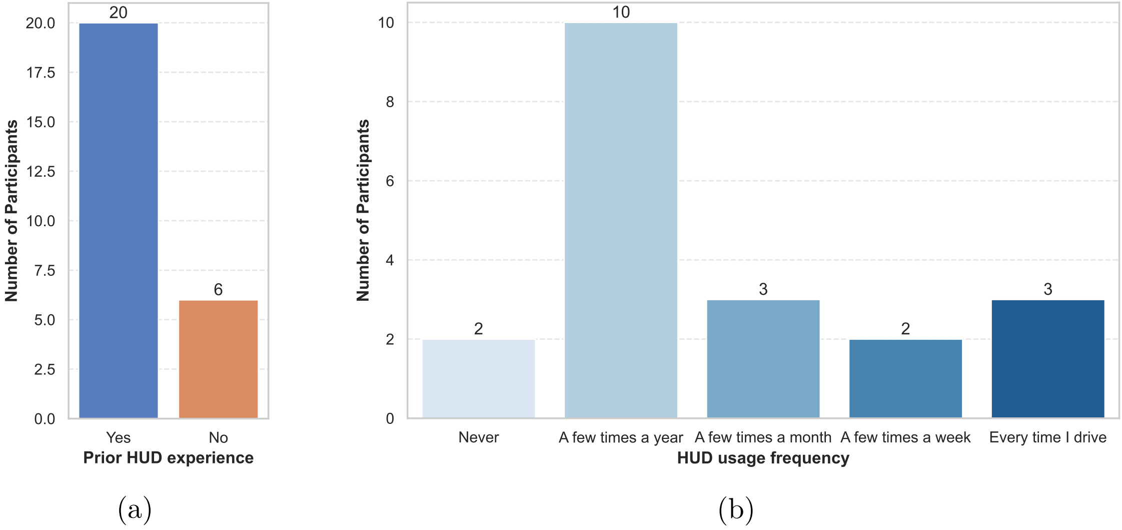


Figure 5.5: Distribution of participant-reported navigation display usage. Participants ranked three display types (CSD, DIM, HUD) according to how often they typically checked the displays for navigation information while driving.



(a) (b)

Figure 5.6: Participants' HUD experience and usage frequency.

Participants were also asked about their preference for using a HUD in their own vehicles. As illustrated in Figure 5.7, the majority (24 participants) expressed a positive attitude toward HUD usage, with 11 of them indicating they would always choose to use one if given the option.

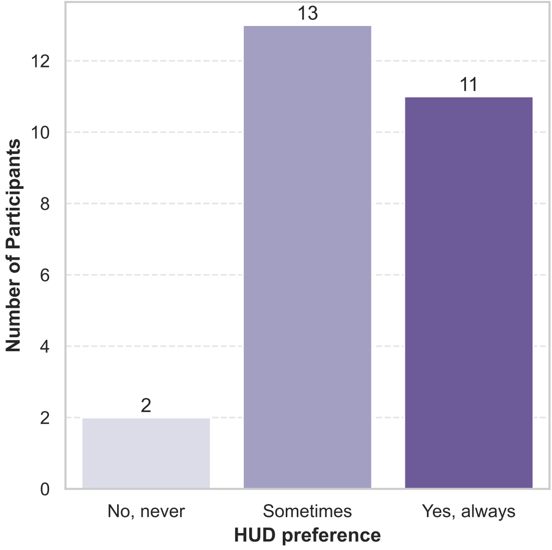


Figure 5.7: Participants' preference for using the HUD if given the option.

5.2 Eye-Tracking Data

This section presents the results of the eye-tracking data collected and the analysis. It includes an overview of data quality, an overall analysis of the full recording sessions, and an event-based analysis focused on selected driving events.

5.2.1 Data Quality

As described in Section 4.2.2, four calibration points were established in the eye-tracking world model. Each participant was required to complete gaze calibration before driving on each of the two test routes. As detailed in Section 4.5.2, statistical comparisons were conducted across five dependent variables under two predefined conditions (HUD vs non-HUD). Following the framework established by Vehlen et al. (2021), the eye-tracking data were evaluated in terms of *robustness*, *precision*, and *accuracy*.

After conducting Wilcoxon-test for within-subject comparison, significant differences were observed in Exposure ($W = 14.5$, $p = 0.001$), Left Eye Calibration Precision (SD) ($W = 55.0$, $p = 0.004$), and Left Eye Calibration Accuracy ($W = 87.5$, $p = 0.025$), as detailed in Table 5.3.

	HUD		non-HUD			
	Mean	**SD**	**Mean**	**SD**	$W(26)$	p
Robustness (valid data%)						
Exposure	94.04	2.86	93.19	3.26	14.5	0.001
Precision (SD) in degrees						
Left Eye Calibration Precision (SD)	1.22	0.87	1.70	1.24	55.0	0.004
Right Eye Calibration Precision (SD)	2.00	1.69	1.75	0.86	172.0	0.929
Accuracy in degrees						
Left Eye Calibration Accuracy	1.81	0.88	2.30	1.08	87.5	0.025
Right Eye Calibration Accuracy	2.70	1.84	2.45	1.02	87.5	0.696

Table 5.3: Eye-tracking data quality analysis in within-subject comparison: HUD vs non-HUD (Wilcoxon-test, $N = 26$).

5.2.2 Overall Analysis

Overall eye-tracking data analyses were performed across the entire recording sessions for each participant to assess their attention to the road, using the gaze-based eye-tracking metric of *Dwell Time*.

For within-subject comparison, Wilcoxon-tests were performed and statistically significant differences were found in *Dwell Time* on road ($W = 34.0$, $p = 0.0001$), CSD ($W = 34.0$, $p = 0.00001$), and DIM ($W = 34.0$, $p = 0.00001$), as shown in Table 5.4. Specifically, when the HUD was present, participants showed decreased time spent looking on the road, as well as reduced time looking at the CSD and the DIM, see Figure 5.8.

	HUD		non-HUD			
Metrics	**Mean**	**SD**	**Mean**	**SD**	$W(26)$	p
$\text{Road}_{\text{DwellTime}(\%)}$	79.82	7.77	85.28	4.35	34.0	0.0001
$\text{CSD}_{\text{DwellTime}(\%)}$	6.19	3.55	8.71	3.95	34.0	0.00001
$\text{DIM}_{\text{DwellTime}(\%)}$	2.67	2.39	6.01	3.31	34.0	0.00001
$\text{HUD}_{\text{DwellTime}(\%)}$	11.32	7.80				

Table 5.4: Overall analysis of eye-tracking data in a within-subject comparison: HUD vs non-HUD (Wilcoxon-test, $N = 26$).

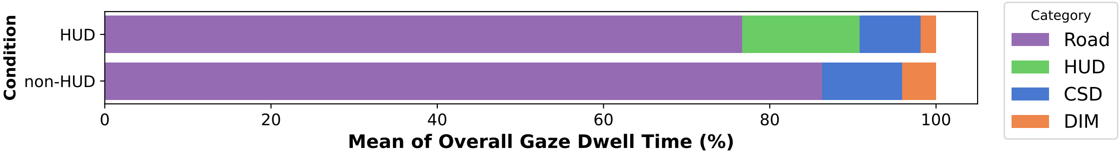


Figure 5.8: Distribution of gaze dwell time from the overall analysis of eye-tracking data in a within-subject comparison: HUD vs non-HUD (Wilcoxon-test, $N = 26$).

5.2.3 Event-based Analysis

As outlined in Section 4.5.1 and Section 4.5.2, Mann-Whitney U tests were used for between-subject comparisons. Table 5.5 presents the results, showing statistically significant differences in *Dwell Time*: attention to the road in Event 4 (U = 27.0, p = 0.004) and Event 6 (U = 44.0, p = 0.042); attention to the CSD in Event 2 (U = 32.0, p = 0.008); and attention to the DIM in Event 4 (U = 15.0, p = 0.0004), Event 5 (U = 46.0, p = 0.045), and Event 6 (U = 38.0, p = 0.019). In these events, the presence of the HUD was associated with decreased visual attention to the road, CSD, and DIM.

Route 1	**Metrics**	**HUD (N=12)**		**non-HUD (N=14)**			
Event	**Dwell Time(%)**	**Mean**	**SD**	**Mean**	**SD**	U	p
0	Road	72.83	12.09	73.10	12.26	82.0	0.939
	CSD	10.83	12.01	18.46	13.14	49.5	0.080
	DIM	5.25	7.52	8.44	11.36	69.0	0.449
	HUD	11.09	7.60				
1	Road	74.40	12.83	81.24	7.98	61.0	0.247
	CSD	8.95	6.42	11.33	4.81	63.0	0.292
	DIM	5.45	4.70	7.43	4.35	65.0	0.341
	HUD	11.20	10.21				
2	Road	84.07	5.65	83.79	14.81	65.0	0.341
	CSD	3.75	3.87	12.48	15.25	32.0	0.008
	DIM	3.59	3.94	3.74	2.75	73.0	0.589
	HUD	8.58	4.66				
3	Road	80.82	8.54	83.87	4.13	62.0	0.269
	CSD	9.82	8.60	11.49	4.38	68.0	0.425
	DIM	2.28	2.32	4.64	5.18	58.5	0.198
	HUD	7.08	6.75				
Route 2	**Metrics**	**HUD (N=14)**		**non-HUD (N=12)**			
Event	**Dwell Time(%)**	**Mean**	**SD**	**Mean**	**SD**	U	p
0	Road	64.02	14.47	72.50	8.13	49.0	0.076
	CSD	8.92	6.27	14.19	11.55	61.0	0.246
	DIM	4.86	8.05	13.31	13.88	49.0	0.065
	HUD	22.20	14.47				
4	Road	77.06	9.92	87.45	5.20	27.0	0.004
	CSD	5.24	2.93	7.14	4.84	66.0	0.368
	DIM	1.84	1.45	5.40	2.91	15.0	0.0004
	HUD	15.86	9.23				
5	Road	75.96	18.55	85.53	5.74	58.0	0.190
	CSD	7.73	7.65	10.15	7.72	66.0	0.368
	DIM	1.28	2.55	4.33	4.50	46.0	0.045
	HUD	15.03	16.90				
6	Road	76.70	13.41	86.34	7.95	44.0	0.042
	CSD	7.35	5.01	9.59	7.18	71.0	0.520
	DIM	1.81	2.41	4.07	3.12	38.0	0.019
	HUD	14.14	12.72				

Table 5.5: Event-based analysis for eye-tracking data in a between-subject comparison: HUD vs non-HUD (Mann-Whitney U-test).

5.3 Cognitive Load

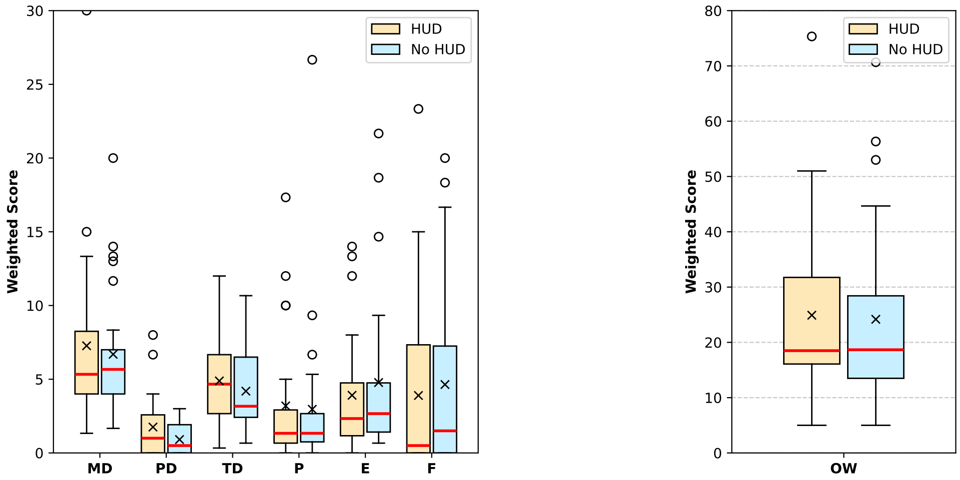
As described in Section 4.5.2, both the overall workload and individual subscales were examined using Weighted and Raw TLX scores in within-subject comparison. The results of conducting Wilcoxon-tests on the overall workload of the Weighted and Raw TLX scores can be seen in Table 5.6 and Table 5.7. A significant difference was found only in Weighted-TLX, specifically in the Physical Demand subscale ($W = 26.0$, $p = 0.017$). Figure 5.9 shows the box-and-whisker plot for weighted subscales and overall workload.

	HUD		**non-HUD**			
Metrics	**Mean**	**SD**	**Mean**	**SD**	$W(26)$	p
Mental Demand	7.27	5.83	6.69	4.42	112.5	0.284
Physical Demand	1.76	2.13	0.91	1.06	26.0	0.017
Temporal Demand	4.88	2.94	4.19	2.74	101.5	0.267
Performance	3.19	4.34	2.96	5.31	97.5	0.779
Effort	3.91	3.94	4.78	5.55	66.0	0.145
Frustration	3.90	6.03	4.64	6.38	45.0	0.234
Weighted Overall Workload	24.91	15.42	24.17	16.48	137.0	0.710

Table 5.6: Weighted TLX cognitive load analysis in within-subject comparison: HUD vs non-HUD (Wilcoxon-test, $N = 26$).

	HUD		**non-HUD**			
Metrics	**Mean**	**SD**	**Mean**	**SD**	$W(26)$	p
Mental Demand	25.96	17.72	26.73	17.03	84.0	0.656
Physical Demand	15.96	9.90	15.58	9.63	86.5	0.729
Temporal Demand	25.00	14.83	21.15	11.07	36.5	0.055
Performance	18.85	18.72	16.06	18.33	47.5	0.474
Effort	25.00	17.20	24.33	20.22	80.5	0.825
Frustration	22.88	19.24	23.46	20.68	52.0	0.975
Raw Overall Workload	22.28	13.34	21.22	13.67	132.0	0.607

Table 5.7: Raw TLX cognitive load analysis in within-subject comparison: HUD vs non-HUD (Wilcoxon-test, $N = 26$).



(a) Weighted Subscales.

(b) Weighted Overall Workload.

Figure 5.9: Weighted NASA-TLX in a within-subject comparison: HUD vs non-HUD ($N = 26$). MD=Mental Demand, PD=Physical Demand, TD= Temporal Demand, P=Performance, E=Effort, F=Frustration.

5.4 Thematic Analysis

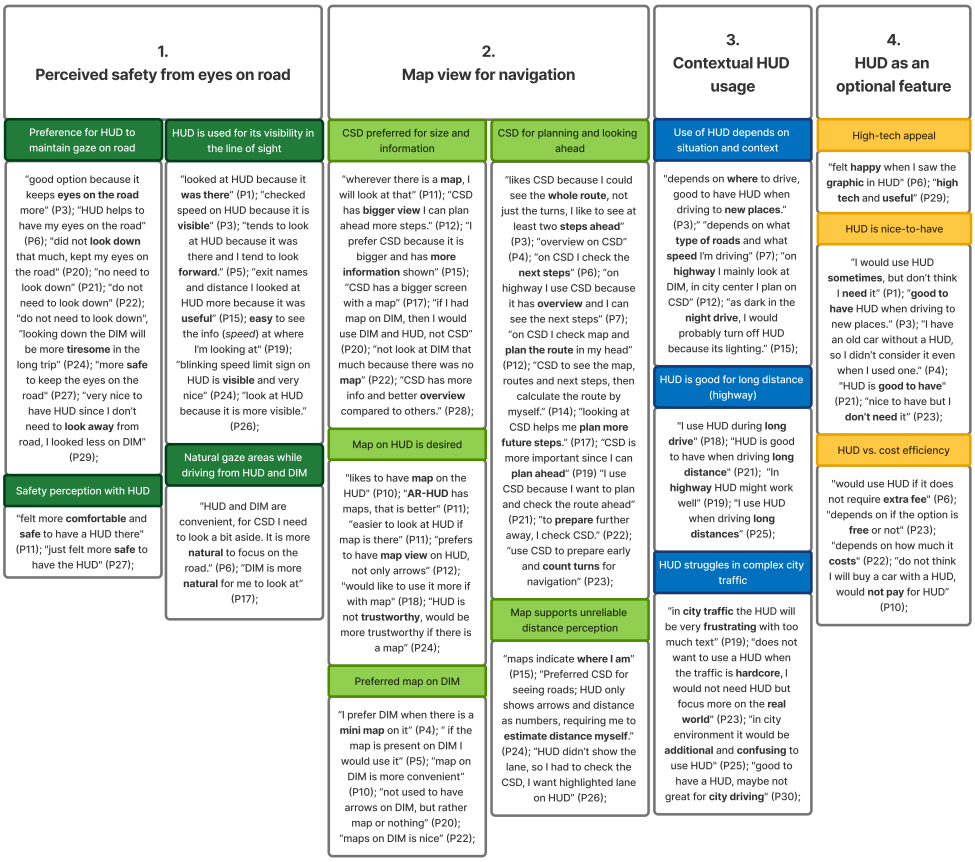
From the interview answers, the initial codes and themes were extracted using a custom GPT model[2] for the TA. The final codes and themes were drawn after additional manual manipulations, resulting in eight themes as shown in Figure 5.10 and Figure 5.11. Each theme consists of two to five codes, with a total of 28 codes. The colour of the codes indicates their belonging to a common theme. Each code contains a minimum of two quotations or paraphrases supporting the code.

The eight themes are as following:

1. Perceived safety from eyes on road
2. Map view for navigation
3. Contextual HUD usage
4. HUD as an optional feature
5. Usefulness of HUD information
6. Distraction and discomfort with HUD
7. Distrust and confusion from HUD
8. HUD capabilities and future potential

---

[2] GPT chat history:
Question 12: https://chatgpt.com/share/686abc52-6300-800d-8194-cf0c1cac0dc7 (Accessed 2025-07-06);
Question 16: https://chatgpt.com/share/68259c03-c1a8-8011-af70-d432e2d110a2 (Accessed 2025-07-06)



1\.
Perceived safety from eyes on road

Preference for HUD to maintain gaze on road

“good option because it keeps **eyes on the road** more” (P3); “HUD helps to have my eyes on the road” (P6); "did not **look down** that much, kept my eyes on the road" (P20); “no need to look down” (P21); "do not need to look down" (P22); "do not need to look down", “looking down the DIM will be more **tiresome** in the long trip” (P24); "more **safe** to keep the eyes on the road" (P27); “very nice to have HUD since I don’t need to **look away** from road, I looked less on DIM” (P29);

Safety perception with HUD

"felt more **comfortable** and **safe** to have a HUD there" (P11); "just felt more **safe** to have the HUD" (P27);

HUD is used for its visibility in the line of sight

“looked at HUD because it **was there**” (P1); “checked speed on HUD because it is **visible**” (P3); “tends to look at HUD because it was there and I tend to look **forward**.” (P5); “exit names and distance I looked at HUD more because it was **useful**” (P15); **easy** to see the info (*speed*) at where I’m looking at" (P19); “blinking speed limit sign on HUD is **visible** and very nice” (P24); “look at HUD because it is more visible.” (P26);

Natural gaze areas while driving from HUD and DIM

“HUD and DIM are convenient, for CSD I need to look a bit aside. It is more **natural** to focus on the road.” (P6); “DIM is more **natural** for me to look at” (P17);

2\.
Map view for navigation

CSD preferred for size and information

“wherever there is a **map**, I will look at that” (P11); “CSD has **bigger view** I can plan ahead more steps.” (P12); “I prefer CSD because it is bigger and has **more information** shown” (P15); “CSD has a bigger screen with a map” (P17); “if I had map on DIM, then I would use DIM and HUD, not CSD” (P20); “not look at DIM that much because there was no **map**” (P22); “CSD has more info and better **overview** compared to others.” (P28);

Map on HUD is desired

"likes to have **map** on the HUD" (P10); "**AR-HUD** has maps, that is better" (P11); “easier to look at HUD if map is there” (P11); “prefers to have **map view** on HUD, not only arrows” (P12); "would like to use it more if with map" (P18); “HUD is not **trustworthy**, would be more trustworthy if there is a map” (P24);

Preferred map on DIM

“I prefer DIM when there is a **mini map** on it” (P4); “ if the map is present on DIM I would use it” (P5); “map on DIM is more convenient” (P10); “not used to have arrows on DIM, but rather map or nothing” (P20); “maps on DIM is nice” (P22);

CSD for planning and looking ahead

“likes CSD because I could see the **whole route**, not just the turns, I like to see at least two **steps ahead**” (P3); “overview on CSD” (P4); “on CSD I check the **next steps**” (P6); “on highway I use CSD because it has **overview** and I can see the next steps” (P7); “on CSD I check map and **plan the route** in my head” (P12); “CSD to see the map, routes and next steps, then calculate the route by myself.” (P14); “looking at CSD helps me **plan more future steps**.” (P17); “CSD is more important since I can **plan ahead**” (P19) “I use CSD because I want to plan and check the route ahead” (P21); “to **prepare** further away, I check CSD.” (P22); “use CSD to prepare early and **count turns** for navigation” (P23);

Map supports unreliable distance perception

“maps indicate **where I am**” (P15); “Preferred CSD for seeing roads; HUD only shows arrows and distance as numbers, requiring me to **estimate distance myself**.” (P24); “HUD didn’t show the lane, so I had to check the CSD, I want highlighted lane on HUD” (P26);

3\.
Contextual HUD usage

Use of HUD depends on situation and context

"depends on **where** to drive, good to have HUD when driving to **new places**.” (P3);” “depends on what **type of roads** and what **speed** I’m driving” (P7); “on **highway** I mainly look at DIM, in city center I plan on CSD” (P12); “as dark in the **night drive**, I would probably turn off HUD because its lighting.” (P15);

HUD is good for long distance (highway)

"I use HUD during **long drive**" (P18); "HUD is good to have when driving **long distance**" (P21); “In **highway** HUD might work well” (P19); "I use HUD when driving **long distances**” (P25);

HUD struggles in complex city traffic

“in **city traffic** the HUD will be very **frustrating** with too much text” (P19); "does not want to use a HUD when the traffic is **hardcore,** I would not need HUD but focus more on the **real world**" (P23); “in city environment it would be **additional** and **confusing** to use HUD" (P25); "good to have a HUD, maybe not great for **city driving**" (P30);

4\.
HUD as an optional feature

High-tech appeal

"felt **happy** when I saw the **graphic** in HUD" (P6); “**high tech** and **useful**” (P29);

HUD is nice-to-have

“I would use HUD **sometimes**, but don’t think I **need** it” (P1); “**good to have** HUD when driving to new places.” (P3); “I have an old car without a HUD, so I didn’t consider it even when I used one.” (P4); "HUD is **good to have**" (P21); “nice to have but I **don’t need** it” (P23);

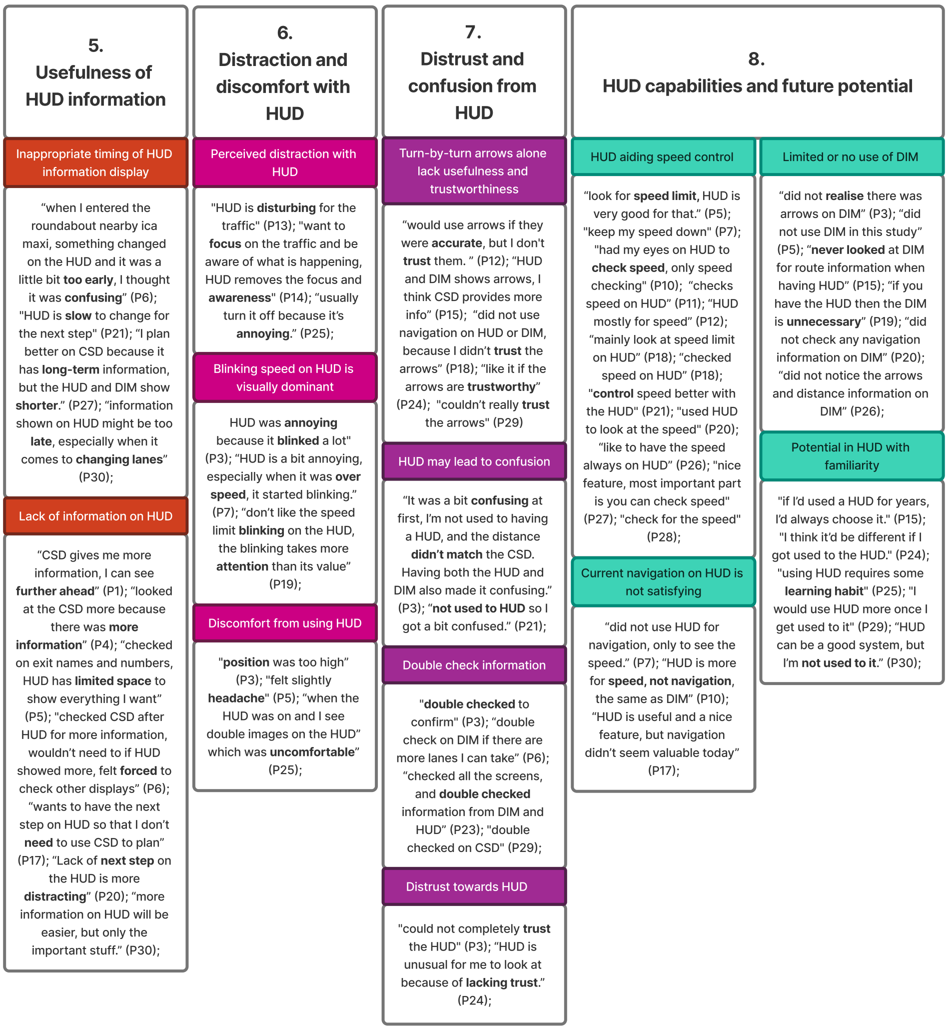
HUD vs. cost efficiency

"would use HUD if it does not require **extra fee**" (P6); "depends on if the option is **free** or not" (P23); "depends on how much it **costs**" (P22); “do not think I will buy a car with a HUD, would **not pay** for HUD” (P10);

Figure 5.10: Thematic analysis covering Themes 1 through 4.

First theme labelled *Perceived safety from eyes on road* consists of four codes: (1) Preference for HUD to maintain gaze on road, (2) Safety perception with HUD, (3) HUD is used for its visibility in the line of sight, and (4) Natural gaze areas while driving. These codes describes how people used the HUD because of its natural and convenient positioning as it was right in line of sight when looking on the road. Not having to look away from the road gave a feeling of perceived safety to many.

Second theme labelled as *Map view for navigation* consists of five codes: (1) CSD preferred for size and information, (2) Map on HUD is desired, (3) Preferred map on DIM, (4) CSD for planning and looking ahead, and (5) Map supports unreliable distance perception. The map view was generally preferred due to its provision of more information compared to turn-by-turn directions alone, as it gave an overview of the road leading to more perceived trust. During navigation in traffic, being able to plan a few steps ahead was considered essential. The map mode facilitated this by allowing drivers to create their own plan while also indicating their real-time location relative to upcoming decision points. Independent of the display type, the map view was preferred.



5. Usefulness of HUD information

Inappropriate timing of HUD information display

"when I entered the roundabout nearby ica maxi, something changed on the HUD and it was a little bit **too early**, I thought it was **confusing**" (P6); "HUD is **slow** to change for the next step" (P21); "I plan better on CSD because it has **long-term** information, but the HUD and DIM show **shorter**." (P27); "information shown on HUD might be too **late**, especially when it comes to **changing lanes**" (P30);

Lack of information on HUD

"CSD gives me more information, I can see **further ahead**" (P1); "looked at the CSD more because there was **more information**" (P4); "checked on exit names and numbers, HUD has **limited space** to show everything I want" (P5); "checked CSD after HUD for more information, wouldn't need to if HUD showed more, felt **forced** to check other displays" (P6); "wants to have the next step on HUD so that I don't **need** to use CSD to plan" (P17); "Lack of **next step** on the HUD is more **distracting**" (P20); "more information on HUD will be easier, but only the important stuff." (P30);

6. Distraction and discomfort with HUD

Perceived distraction with HUD

"HUD is **disturbing** for the traffic" (P13); "want to **focus** on the traffic and be aware of what is happening, HUD removes the focus and **awareness**" (P14); "usually turn it off because it's **annoying**." (P25);

Blinking speed on HUD is visually dominant

HUD was **annoying** because it **blinked** a lot" (P3); "HUD is a bit annoying, especially when it was **over speed**, it started blinking." (P7); "don't like the speed limit **blinking** on the HUD, the blinking takes more **attention** than its value" (P19);

Discomfort from using HUD

"**position** was too high" (P3); "felt slightly **headache**" (P5); "when the HUD was on and I see double images on the HUD" which was **uncomfortable**" (P25);

7. Distrust and confusion from HUD

Turn-by-turn arrows alone lack usefulness and trustworthiness

"would use arrows if they were **accurate**, but I don't **trust** them. " (P12); "HUD and DIM shows arrows, I think CSD provides more info" (P15); "did not use navigation on HUD or DIM, because I didn't **trust** the arrows" (P18); "like it if the arrows are **trustworthy**" (P24); "couldn't really **trust** the arrows" (P29)

HUD may lead to confusion

"It was a bit **confusing** at first, I'm not used to having a HUD, and the distance **didn't match** the CSD. Having both the HUD and DIM also made it confusing." (P3); "**not used to HUD** so I got a bit confused." (P21);

Double check information

"**double checked** to confirm" (P3); "double check on DIM if there are more lanes I can take" (P6); "checked all the screens, and **double checked** information from DIM and HUD" (P23); "double checked on CSD" (P29);

Distrust towards HUD

"could not completely **trust** the HUD" (P3); "HUD is unusual for me to look at because of **lacking trust**." (P24);

8. HUD capabilities and future potential

HUD aiding speed control

"look for **speed limit,** HUD is very good for that." (P5); "keep my speed down" (P7); "had my eyes on HUD to **check speed**, only speed checking" (P10); "checks speed on HUD" (P11); "HUD mostly for speed" (P12); "mainly look at speed limit on HUD" (P18); "checked speed on HUD" (P18); "**control** speed better with the HUD" (P21); "used HUD to look at the speed" (P20); "like to have the speed always on HUD" (P26); "nice feature, most important part is you can check speed" (P27); "check for the speed" (P28);

Current navigation on HUD is not satisfying

"did not use HUD for navigation, only to see the speed." (P7); "HUD is more for **speed, not navigation**, the same as DIM" (P10); "HUD is useful and a nice feature, but navigation didn't seem valuable today" (P17);

Limited or no use of DIM

"did not **realise** there was arrows on DIM" (P3); "did not use DIM in this study" (P5); "**never looked** at DIM for route information when having HUD" (P15); "if you have the HUD then the DIM is **unnecessary**" (P19); "did not check any navigation information on DIM" (P20); "did not notice the arrows and distance information on DIM" (P26);

Potential in HUD with familiarity

"if I'd used a HUD for years, I'd always choose it." (P15); "I think it'd be different if I got used to the HUD." (P24); "using HUD requires some **learning habit**" (P25); "I would use HUD more once I get used to it" (P29); "HUD can be a good system, but I'm **not used to it**." (P30);

Figure 5.11: Thematic analysis covering Themes 5 through 8.

Third theme is about *Contextual HUD usage*, which consists of three codes: (1) Use of HUD depends on situation and context, (2) HUD is good for long distance (e.g., highway), and (3) HUD struggles in complex city traffic. The perceived usefulness of HUD depended on the type of road and traffic situation. The use of the HUD was perceived as likeable during long-distance driving, on highways, or in unfamiliar areas. However, in denser traffic environments such as city traffic, HUD was perceived as "too much" or "additional", often leading to frustration and confusion.

Fourth theme, *HUD usefulness and perceived necessity*, consists of three codes: (1) High-tech appeal, (2) HUD is nice-to-have, and (3) HUD vs. cost efficiency. From the cluster, the HUD was perceived as having a high-tech appeal and seen as a nice-to-have feature rather than a necessity. Cost efficiency also emerged as a concern, with many participants stating they would not opt for the HUD if it came with an additional cost.

Fifth theme about *Usefulness of HUD information* consists of two codes: (1) Inappropriate timing of HUD information display, and (2) Lack of information on HUD. This theme captures the perceived suboptimal timing and limited navigation information presented on the HUD. Navigation information presented too early on the HUD was perceived as confusing, while late information disrupted forward planning. The limited amount of navigation information on HUD—due to its limited display size—led to confusion, distraction, and a desire for more information among participants.

Sixth theme highlights the *Discomfort and distraction with HUD* with three codes: (1) Perceived Distraction with HUD, (2) Blinking speed on HUD is visually dominant, and (3) Discomfort from using HUD. Participants reported experiencing ghosting, where double images appeared on the HUD, leading to discomfort and, in some cases, headaches. There was a comment about positioning of the HUD image being too high causing discomfort. The HUD was also perceived as 'disturbing' and 'annoying', among other things, caused by, e.g., the blinking speed sign. These experiences disrupted participants' focus and awareness on the road.

Seventh theme presents the *Distrust and confusion from HUD* with four codes: (1) Turn-by-turn arrows alone lack usefulness and trustworthiness, (2) HUD can lead to confusion, (3) Double check information, and (4) Distrust towards HUD. Participants reported that the turn-by-tun arrows felt inaccurate and untrustworthy. Furthermore, the mismatch of information on different displays, combined with inexperience with the HUD, contributed to confusion. As a result, participants frequently double-checked information on the displays for confirmation, reflecting a general lack of trust towards the HUD.

Eighth theme is a cluster about *HUD capabilities and future potential* consisting of four codes: (1) HUD aiding speed control, (2) Current navigation on HUD is not satisfying, (3) Limited or no use of DIM, and (4) Potential in HUD with familiarity. The use of the DIM became limited with the presence of the HUD, as speed was the primary reason for checking it. Navigation information on the DIM was either unnoticed, overlooked, or deemed unnecessary. While the HUD received positive feedback for speed monitoring and control, its use for navigation was reported as unsatisfying and currently of little value. Participants expressed the need to relearn certain habits and adapt to the new technology, while also recognising its potential benefits with increased familiarity and experience.

Figure 5.12 presents a *thematic map* that visually organises the themes by illustrating their interconnections. The sixth and seventh themes focus on various concerns regarding the use of HUD, including distraction, discomfort, distrust, and confusion. This in turn raises the questions about the usefulness of the HUD information, as discussed in the fifth theme. The first, third, and fourth themes each explore different dimensions of HUD usefulness. The first theme reflects a generally favourable view of the HUD, while also acknowledging its limitations, capabilities, and possibilities, as highlighted in the eighth and second theme. The third theme presents a

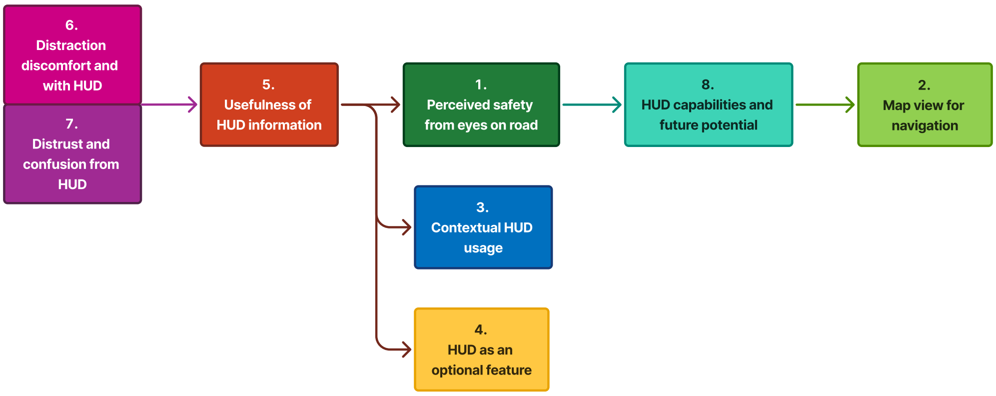


Figure 5.12: Thematic map illustrating the relationship between the themes.

more nuanced perspective, recognising both benefits and drawbacks depending on the context. The fourth theme offers a more neutral interpretation, viewing the HUD as optional.

5.5 General Feedback on the Study

As described in Section 4.3.8 and shown in Appendix B.6, participants were asked to share their views on both the NASA-TLX and the overall study. Of the 26 participants, 15 reported that certain aspects of the NASA-TLX were somewhat difficult to interpret, particularly the subscales related to Performance, Effort, Frustration, and Temporal Demand. In addition, several participants mentioned that the pairwise comparison sheet (the second page of the questionnaire) was more difficult to fill out than the initial rating section on the first page, requiring greater effort and time to complete.

With regard to the study itself, participants provided a range of feedback. Several expressed curiosity about the final results, the goals of the study, and their own eye-tracking behaviours, and some also commented positively on the experimental setup. However, remarks concerning task and route difficulty were also noted. A few participants had expected the driving tasks to be more complex but found them unchallenging, describing the experience as boring or too easy.

6

Discussion

This study investigated differences in driver behaviour between HUD and non-HUD conditions, focusing on attention, distraction, and cognitive load, and implications for usability and safety. Interpretations and key findings from the data analysis were drawn to address the research question: *What are some of the most significant differences in driver behaviours when it comes to HUD and non-HUD conditions?*

This chapter begins by discussing the quality of the collected eye-tracking data in terms of its validity and reliability. Then it explores the findings on attention and distraction based on the analysed eye-tracking data, as well as the cognitive load derived from subjective assessments and TA of interview responses. Insights from participant demographics were integrated to enable triangulation and enhance the validity and depth of the findings. Furthermore, by relating these results to prior research, the analysis critically evaluates both the potential and the limitations of HUDs in traffic. Finally, this section addresses the studys limitations and the ethical considerations that were taken into account.

6.1 Eye-Tacking Data Reliability and Validity

Significant differences were observed in Exposure, Left Eye Calibration Precision (SD), and Left Eye Calibration Accuracy, indicating variations in eye-tracking data quality, see Table 5.3. These metrics demonstrated greater robustness, precision, and accuracy under the HUD condition. In the case of Exposure, it is speculated that the presence of a HUD may have helped direct participants attention forward, reducing glances toward peripheral objects that could cause data loss due to their position. This may have contributed to a higher proportion of valid eye-tracking data. As for the Left Eye Calibration Precision (SD) and Accuracy, these measures may have been influenced by lighting conditions and participants' vision conditions during gaze calibration; however, further analysis is required to identify the key contributing factors.

According to iMotions[1], an accuracy threshold of 3.0 degrees and a minimum of 90% valid eye-tracking samples (i.e., Exposure) are recommended. Although the study aimed to meet these standards for all participants, some recordings—particularly

---

[1] Note on data quality from iMotions' help centre: https://help.imotions.com/v2/docs/en/analysis (*Customer account required for login.)

those affected by fluctuating lighting conditions—did not consistently achieve the desired quality. However, after reviewing the eye-tracking data quality results with project stakeholders, it was concluded that the data from the 26 participants included in the study exhibited sufficiently accurate calibration for a naturalistic study to support further analysis.

6.2 Attention and Distraction

The results in Section 5.4 present the eight themes of the TA representing the participants' perceptions of the HUD in a navigational context within the setup of the study. The findings from the TA suggested that the current usability of the HUD may still be unsatisfactory for several participants. Issues such as insufficient information and mistimed prompts raised questions about the usefulness of the HUD information. This finding is consistent with the results of AblaSSmeier et al. (2007), where the participants expressed a desire for more information and features on the HUD, while also noting that its visualisation and user-friendliness in the navigation map were still underdeveloped. In this study, many participants perceived safety from the use of the HUD, emphasising its capabilities to aid speed and its potential, while also addressing its limitations for navigational purposes. Many participants expressed a preference for having an overview, such as a map, when navigating, allowing forward planning and reducing reliance on their own sense of location. This was in contrast to the turn-by-turn arrows on the HUD. Some saw the usefulness of the current capabilities of HUD as contextual, primarily on highways when driving at high speed and long distances. Others questioned its usefulness, describing it as merely an optional feature.

The findings from the eye-tracking data indicated a significant shift in visual attention with the presence of the HUD. Whether this shift resulted in safer driving behaviours remains unclear. Although HUDs are designed to reduce glance duration of driver's sight deviations from the road (Maroto et al., 2018), the findings in this study indicate that some participants felt confused or uncertain about the information shown, as highlighted in the seventh theme. Without measuring whether visual attention on the HUD improves driving performance, such as reaction time or decision accuracy, it is challenging to determine whether this shift in gaze supports or impairs driving safety. Although Liu and Wen (2004) found that HUDs improved driver response time to speed limits and urgent events in a simulated environment, applying such evaluations in real traffic is more complex. It requires greater safety considerations and controlled methods.

6.2.1 Overall Analysis

The overall analysis of the eye-tracking data in Section 5.2.2 revealed significant reductions in participants' gaze dwell time on the road, CSD, and DIM with the presence of the HUD. The time dwelt on the HDDs was reduced by more than one-fourth on the CSD (from around 9% to 6%) and by more than half the time on the DIM (from about 6% to 3%), see Table 5.4. Out of all the displays, HUD was dwelt on the most (around 11%), more than CSD and DIM combined. The time spent gazing at the road was reduced by 5 percentage points (from around 80% to 85%). The road dwell time was previously defined in Section 4.5.2 as $\text{Road}_{\text{DwellTime}(\%)}$, which represents the time spent looking at the road. This was defined as the total eye-tracking recording time subtracted by the time spent looking at the displays (CSD, DIM, and HUD). Based on this definition, one could argue that the reduced time spent looking at the road, and more time on the displays, raises safety concerns regarding the use of HUDs.

There have been concerns about safety and usability with HUDs from previous studies as mentioned previously in Sections 2.4.3 and 2.4.4 about cognitive capture and superimposition. The sixth theme of the TA raised concerns about distraction and discomfort regarding the use of HUDs. Participants expressed “disturbance”, “annoyance”, and reduced “focus” and “awareness” when using the HUD, indicating possible superimposition experienced by the participants. The relatively high portion of time dwelt on the HUD may be attributed to cognitive capture, where the HUD diverts the participants’ attention from the primary task of driving to focus on the display itself. It may also be a result from superimposition, where the virtual image obscures critical road information. In either case, the HUD could arguably pose a significant safety concern.

Another concern with the HUD is regarding its usability. The fifth and seventh themes were about the usefulness of the navigation information (turn-by-turn arrows) on both the DIM and the HUD. The arrows by themselves relied heavily on the participants’ sense of distance, which participants expressed as challenging to judge. The lack of information, in addition to the inappropriate timing of the navigation information on the HUD, often led to confusion and divided attention. Moreover, participants’ feedback on the seventh theme revealed that the navigation information was one of the sources of confusion and distrust. The turn-by-turn instructions were described as “inaccurate”, appearing “too early”, “too late”, or for “too short” time. The inconsistency of HUD navigation was most pronounced in fast-changing traffic environments, such as city driving. Although one of the factors contributing to the inconsistency of the HUD information was the inappropriate timing of the HUD navigation, the navigation system itself (Google Maps) was also suboptimal. The suggested lane placements and the speed sign could often be inaccurate, which was addressed by the participants. This reflects what Brown and Laurier (2012) refer to as “normal troubles”, which are common GPS issues that may confuse or disorient drivers. These problems often arise not just from the technology itself, but from broader IxD challenges. Thus, the overwhelming dwell time on HUD with this usability concern may negatively affect driving safety.

This highlights the importance of a map overview provided by the CSD, which was perceived as the primary source of navigation information by participants. This finding aligns with the previous study indicating that participants believe the HUD cannot fully replace customary display types, such as CSD and DIM (AblaSSmeier et al., 2007). The second theme from the TA supports this, highlighting the participants’ preference for map-based guidance to plan ahead and reduce uncertainty during driving. While the CSD was preferred, many participants expressed a desire for the map to be shown on the DIM instead, as it is closer to the line of sight and reflects the configuration in their personal vehicles. Some participants even showed openness to the idea of integrating a map into the HUD. Suggestions to improve navigation displays included showing the next upcoming turn and reducing the amount of text to help focus on essential information.

Conversely, adhering to the definition that the HUD is within the driver's line of sight and displaying virtual, transparent images, then looking at the HUD could be considered as "eyes-on-road". In this case, only time spent on the HDDs would be counted as looking away from the road. This notable reduction in down-glances may indicate increased visual attention toward the road, potentially enhancing traffic safety. Road dwell time could in this case be defined as the total eye-tracking duration subtracted by the time spent on the HDDs, counting the HUD as a part of the road. Following this reasoning, the HUD may contribute to more consistent forward glances and increased time spent looking in the direction of travel. This shift in gaze behaviour aligns with prior research comparing HDDs and HUDs (AblaSSmeier et al., 2007; Maroto et al., 2018), which suggests that HUDs can effectively reduce drivers' need to divert their gaze down to look at the HDDs. The first theme of the TA highlights participants' perceived safety when using the HUD as they do not need to look down and keep their eyes on the road more. Whether this perceived safety translates into objectively safer driving remains uncertain.

Overall, the HUD was positively received, particularly for its positioning within the participants' line of sight, as highlighted in the first theme. Its convenient placement, along with the perceived usefulness of displaying the speedometer, contributed to a sense of increased safety, as reflected in the first and eighth theme. This aligned well with the claims made by Rutley (1975) and Sojourner and Antin (1990), who found that HUDs enhance drivers' awareness of their vehicle and help maintain more consistent speed control, enabling drivers to follow speed limits more accurately and thereby contribute to safe driving.

Out of all the displays, the HUD was dwelt on the most, and the CSD was the second in line, and lastly the DIM. This aligns with the eighth theme in the TA where participants expressed that they had minimal to no use of the DIM when the HUD was present. However, the CSD placing second in dwell time after the HUD does not align with participants' self-reported rankings of the displays according to their own perception of navigation usage (see Figure 5.5). Here, CSD was ranked as the most frequently checked display by the majority of participants (18 out of 26), while the DIM and HUD were less frequently prioritised and selected with a similar frequency. Several factors may have contributed to this information mismatch. As reported in the first and eighth themes of the TA, the speedometer on the HUD was well received by participants, while the current navigation on the HUD was perceived as lacking. Thus, the increased in HUD dwell time was likely due to more participants monitoring the speedometer than to their use of the navigation feature.

6.2.2 Event-Based Analysis

For the event-based analysis shown in Section 5.2.3, events 2, 4, 5, and 6 showed significant decreases in dwell time on the road, CSD, or DIM with the presence of a HUD. In contrast, it was observed that when the HUD was present, dwell time did not increase significantly on any of the displays or on the road. This suggests that participants directed most of their visual attention to the HUD rather than the road or other displays. Furthermore, when dwell time on the road decreased, dwell time on the DIM also dropped, but not the other way around. There were cases where the DIM or the CSD showed significant decreases in dwell time without affecting attention to the other AOIs.

During Event 2, the dwell time on the CSD was found to be significantly lower when having the HUD, from around 12% to 4%. The DIM and the road dwell time remained about the same through both display conditions, around 4% and 84%, respectively. This indicated that the participants shifted most of their gaze from the CSD to the HUD during the event. The significant dwell time on the HUD could possibly be due to the absence of navigation changes for a long period of time. When driving on the highway, the navigation information on the HUD was perceived as sufficient as highlighted in the third theme of the TA. In the theme, participants expressed a preference for using the HUD on highways primarily to check their speed, as little planning was required due to the monotony of the road.

In Event 4, two significant differences were observed: a decrease in the dwell time of the road and the DIM with the presence of the HUD. The traffic situation during Event 4 included multiple navigation changes in the HUD due to three consecutive roundabouts. This event occurred early in Route 2 and may have sparked curiosity about the HUD among participants in counterbalanced group 1, who had previously driven Route 1 without it. Thus, this could be one of the affecting factors in more time spent looking at the HUD and less on the DIM and the road during the event. Additionally, as discussed in the third theme on the contextual use of the HUD, participants reported difficulties with HUD usage during complex traffic situations, mentioning city traffic as an example. A contributing factor to the preference for using the HUD on highways rather than in city traffic may be the usability issues discussed earlier, such as cognitive capture, problems with superimposition, and the drawbacks of the provided HUD information. The frequent navigation changes on the HUD during the event were possibly perceived as distracting, as participants described the HUD as “intrusive”, “annoying”, and “disturbance of traffic awareness”, particularly when visual elements were blinking or overly dominant. This may have led them to spend additional time focusing on the HUD in an effort to interpret the interface, raising concerns about the safety and usability of HUDs during complex navigation tasks. While a previous study suggested that a HUD has high potential for efficient information capturing in complex scenarios (AblaSSmeier et al., 2007), participants in this study reported contradictory experiences. Since gaze data, including participants’ attention to the HUD, was not examined across varying levels of traffic complexity, and cognitive load was not measured, further research is needed to explore this potential divergence.

During Event 5, there was a significant decrease in DIM dwell time when the HUD was active. This event was relatively short compared to the others, involving a lane change for a highway exit. The eighth theme also supports the claim about the limited use of DIM in the presence of the HUD. The participants reported that the information shown on the HUD and DIM was similar. However, many preferred to look at the HUD more due to its convenient placement, as discussed in the first theme. Furthermore, during this event, navigation information appeared on the HUD approximately 2 km before the manoeuvring instruction. This prolonged visibility may have increased dwell time on the HUD, as participants relied more on it and less on the DIM, for both navigation and speed information.

For Event 6, a significant decrease in the road and the DIM dwell time was found with the presence of the HUD. In other words, more time is spent on the HUD and the CSD. Several participants pointed out that this event was challenging with a short time to prepare, several unexpected lane changes, and merging into an 80 km/h speed lane. In this fast-paced traffic, participants may have felt an increased need to check navigation information for an accurate lane change as they close to the merging lane. However, as highlighted in the fifth and seventh themes, participants expressed their opinions about the lack of sufficient and appropriate information on HUD, "forcing" them to "double-check" the information on other displays.

6.3 Cognitive Load

Following the model proposed by Galy et al. (2018) as discussed in Section 2.6, the impact on drivers' cognitive load could be explained. The model illustrates how cognitive load factors affect cognitive resources during driving. The usability concern with the HUD likely imposed a high extraneous cognitive load on the participants due to the perceived inefficient UI. Parts of the load could have been caused by the fact that several participants were relatively inexperienced with the HUD, which contributed to an increased intrinsic load. Furthermore, as mentioned in Section 1.1, the presence of an additional display may have increased cognitive load by contributing to more frequent attention shifts (Cheng et al., 2023; Maroto et al., 2018). During the shift, or saccadic transition, the visual perception is momentarily suppressed (B. Albert & Tullis, 2013). Although most of the participants had experience with Google Maps (see Figure 5.4b), how the information was shown on the HUD still differed and could have affected an increase in intrinsic load.

In contrast, the subjective evaluation of cognitive load, both the Weighted and Raw TLX results, did not indicate significant differences in workload between the HUD and non-HUD conditions (see Section 5.3). The only notable exception was the weighted Physical Demand score, which was significantly higher in the HUD condition. As mentioned in Section 1.1, Liu and Wen (2004) found that first-time HUD users often experienced discomfort while familiarising themselves with the tool and adapting their driving routines to the new setup. In this study, the majority of the participants were not regular HUD users, even if they had some experience using a HUD, see Figure 5.6. This corresponds to participants' positive views on the potential of HUDs in the eighth theme of the TA, while noting that the HUD required a degree of familiarity (see Section 5.4). Thus, the significantly higher Physical Demand could be due to the limited HUD experience in the collected sample.

Reported discomfort and the difference in Physical Demand may also be linked to the fact that some participants experienced seeing ghost images or even headaches while using the HUD. This may explain why a significant difference was found, as many participants, particularly those with over 10 years of driving experience, reported relatively low physical demand. Some attributed the low demand to their familiarity with the routes and the relatively low difficulty of the driving tasks. Since no workload subscale other than Physical Demand indicated a significant difference between HUD and non-HUD conditions, it could be concluded that participants did not perceive the overall workload as substantially different.

In terms of understanding of NASA-TLX, as observed and pointed out by participants, many experienced challenge with assessing the weighing part of NASA-TLX, aligning with what was reported by Hart (2006) as described in Section 3.6.2. In this part, participants were asked to compare each subscale with each other, forming 15 pair comparisons. The questionnaire instructions were followed to ensure that standardised information was provided, emphasising that the questionnaire reflected the participants' own subjective experience of workload. However, many expressed their confusion on how to compare the subscales when they are overlapping in dimension and not fully separable, especially comparisons regarding Performance, Effort and Frustration. Many were also confused about Temporal Demand and its definition. This confusion may have slightly compromised the trustworthiness of the subjective evaluation; a limitation inherent to the nature of the assessment tool, as noted by Hart (2006).

6.4 Limitations and Future Work

Several contextual and methodological factors should be considered when interpreting the findings, which can be addressed in four main aspects: participant sampling, real-road experimentation constraints, time limitations, and scope of available resources. These limitations also highlight directions for future research.

6.4.1 Participant Sampling

One limitation of the study is that all participants were recruited exclusively from VCC, introducing potential bias due to prior familiarity with vehicle interfaces or brand-specific design expectations. Although the study aimed to recruit participants with different backgrounds (e.g., gender, age, driving experience, and occupations, vision profiles), and correlation analysis was considered to help mitigate such biases, the sample lacked sufficient stratification to support this approach fully.

Gender: As shown in Figure 5.1a, the participant pool was predominantly male, with 22 out of 26 participants. As noted in Section 4.2.6, gender was not specified during recruitment and participation was voluntary. However, given that all participants were recruited from an automotive company, which is often considered a male-dominated industry (Bullock, 2019; Lloyd and Mey, 2007), this gender imbalance was difficult to avoid and likely contributed to the skewed sample.

Previous research has highlighted gender-based differences in the perception and use of automotive HUDs. Y. Li et al. (2024) have shown significant differences on driving interface usage between men and women, with male participants focusing more on functional information and female participants focusing more on the aesthetic design of the HUD. This may have influenced the results of this study, as most participants expressed opinions on the HUD functions (e.g., speedometer), but seldom commented on its visual or graphic design. Visual design elements such as layout can significantly affect cognitive performance and driving workload (J. Li et al., 2024). The limited feedback on these aspects may be due to the participant gender imbalance and could have introduced bias to the findings.

In terms of gender differences in driving workload, Luo et al. (2022) demonstrated that female experienced significantly higher workload than male drivers, as evidenced by a greater heart rate growth rate (HRGR) under distracted driving conditions. While the current study relied on subjective measures of workload and did not reveal significant differences in overall workload scores between genders, the under-representation of female participants may have affected these findings.

The gender imbalance in this study may also have influenced participants' attention to the HUD. According to Campisi et al. (2021), men tend to adopt new technologies and incorporate them into their habits more quickly than women. Consequently, a male-dominated participant sample may have led to increased attention directed toward the HUD during the study. This, in turn, could have contributed to the observed significant decreases in attention to other components, such as the DIM and CSD, in the presence of the HUD.

Age: The age distribution of participants in this study was concentrated within the 25–54 age range, with the largest subgroup aged 35–44 (10 participants), as shown in Figure 5.1b. Only one participant was aged 18–24, resulting in minimal representation from young drivers. Similarly, no drivers older than 55 were included in the study. This concentration of mid-career adults suggests that the findings in this study may be more reflective of the attitudes and behaviours typical of this demographic, potentially limiting the applicability of the results to younger or older drivers, who may have different ergonomic or cognitive needs. For example, research by Lockhart and Shi (2010) has demonstrated that age significantly affects dynamic visual accommodation due to changes in biomechanical and neural processing, using a simulated dashboard reading task while driving. This suggests that older adults may experience reduced visual performance in typical driving conditions.

If the study had included a more age-diverse sample, particularly younger drivers (aged 18–24), and old drivers (aged 55 and above), the results might have shown different patterns. Younger drivers tend to demonstrate distinct attentional behaviours, potentially shaped by their greater familiarity with digital technologies. Previous studies (Amini et al., 2023; Gauld et al., 2017; Prat et al., 2015) have shown that drivers in the young age group (typically aged 17–25) are more likely to engage in secondary tasks, such as responding to phone calls or interacting with in-vehicle displays. These tendencies may influence how they attend to HUDs during driving, potentially leading to increased attention toward the HUD.

On the other hand, older drivers often face age-related cognitive and perceptual changes that can increase distraction. As discussed in Section 3.4.3, distraction is closely related to age and perceptual load. Lavie (2005) addressed that older individuals experience more distraction under low perceptual load, a finding supported by Maylor and Lavie (1998), whose experiment showed significantly greater distractor effects in older participants (aged 65–79) under minimal load conditions. Including such participants in the current study could have led to higher overall workload scores. This is particularly relevant given the feedback from participants discussed in Section 5.5, where some described the driving tasks as “easy” or “boring”. The presence of older drivers, who may find even simple tasks more demanding, might have yielded higher subjective workload assessments.

Driving Experience and Route Familiarity: The final dataset included 26 participants with considerable driving experience—most having held a licence for over a decade and driving regularly each week (see Figure 5.2). Familiarity with the test routes may also have influenced the outcomes. Several participants mentioned during the interviews that they were familiar with the driving routes, mainly because the selected roads were located near VCC. This is a crucial point to address, as familiarity with the route strongly influences the perceived difficulty of the navigation task. This familiarity could have reduced cognitive demands during the task, potentially diminishing the observable effects of the HUD. Some participants described the study as “boring” or “easy,” which may be attributed not only to their extensive driving experience, but also to their familiarity with the environment. According to Luo et al. (2022), novice drivers tend to experience higher levels of cognitive burden and nervousness during distracted driving scenarios compared to more experienced drivers. If the study had included more novice drivers with limited exposure to the test routes, it is possible that different patterns of distraction and cognitive load would have emerged.

Vision: Vision characteristics were also a relevant factor for the reliability of the study, particularly given the use of eye-tracking as a primary research tool. As shown in Figure 5.3, the participant pool exhibited a diverse range of visual profiles. Such variation is critical to consider in eye-tracking research, as glasses or contact lenses can introduce calibration noise, interfere with IR signal detection, or reduce tracking stability. The significant differences observed in Left Eye Calibration Precision (SD) and Accuracy between the HUD and non-HUD conditions may partially reflect these individual visual differences. Although this study did not conduct a detailed analysis of visual profile correlations with eye-tracking reliability, these individual differences should be acknowledged when interpreting gaze behaviour and cognitive responses in the context of HUD usability and driver safety.

In summary, future research should aim to recruit a more diverse and representative participant pool to enhance the generalisability of the findings. This includes involving participants from outside of VCC and the broader automotive industry, as well as ensuring a balanced distribution of gender and age to capture perspectives across a wider range of demographic groups. Furthermore, given the potential impact of vision characteristics on eye-tracking data quality and HUD interaction, future studies should consider explicitly analysing the relationship between visual profiles and gaze metrics, such as incorporating correlation analyses (e.g., Pearson or Partial Correlation). This would offer deeper insights into how individual differences in vision affect perception, attention, and cognitive load in HUD-supported driving environments.

6.4.2 Real-Road Experimentation Constraints

Real-road studies provide valuable ecological validity but also introduce a range of practical and methodological constraints. This section outlines the key challenges encountered during the study, focusing on uncontrolled conditions, naturalistic driving and safety considerations, eye-tracking calibration variability, and limitations in cognitive load measurement. Each of these factors may have influenced data quality and the consistency of the participant experiences.

Uncontrolled Real-World Conditions: Conducting the study in real traffic environments introduced several uncontrollable variables, such as changing weather, fluctuating traffic conditions, and occasional disruptions. While this approach increased ecological validity by reflecting realistic driving scenarios, it also limited the ability to maintain strict experimental control. Efforts were made to mitigate the impact of these variables, such as excluding certain data as detailed in Table 5.1. However, some unforeseen events still occurred within the included dataset. For example, participant P10 (see Table 5.2) was exposed to conditions that could not be fully controlled.

Naturalistic Driving and Safety Considerations: Another factor that may have affected the results was the study's attempt to balance naturalistic driving behaviours while ensuring safety. Participants were instructed to drive as naturally as possible to maintain a realistic experience. However, some drivers chose to use cruise control to keep a steady speed, which may have reduced their need to monitor in-vehicle displays such as the speedometer. While this behaviour may have remained consistent across both test routes, some participants reported that driving without cruise control on Route 1 felt unnatural and therefore requested to use it on Route 2. Furthermore, to ensure safety, participants were encouraged to ask the researchers questions if they were uncertain about any aspect of the study. In some cases, this led to off-topic conversations, which may have influenced gaze behaviour and introduced additional variability in the eye-tracking data.

Calibration Variability in Eye-Tracking: Variability in eye-tracking calibration posed a notable limitation in this study, including both the eye-tracking cameras and participant gaze calibration. Although the cameras were calibrated before each participant session (see Section 4.3.1), the resulting accuracy and validity differed across participants. During the camera calibration, researchers pre-adjusted the driver's seat to suit an average-height individual, as participants had not yet arrived during the setup phase. This required estimating the headbox position, which may not have aligned with participants' actual seating preferences. Ideally, this calibration should occur after participants have adjusted the seat to their preferred position. Eye-tracking cameras are highly sensitive to their relative positioning, and even small shifts can affect accuracy. Such shifts may result from seat movements, vibrations during driving, or softening of the acrylic tape used to mount the cameras. This pre-calibration approach represented a compromise between procedural efficiency and data quality.

As for participant gaze calibration, this process sometimes had to be repeated due to challenging lighting conditions, especially in bright and sunny weather. In contrast to screen-based eye-tracking studies, where the distance and angle between the participant and the cameras remain fixed (Vehlen et al., 2021), real-world driving environments introduce more variability. Eye-tracking in a dynamic, world-referenced setup requires a more standardised and robust calibration procedure to ensure consistent data quality.

Future research should focus on improving the standardisation of calibration procedures, potentially through adaptive protocols that account for environmental variables such as lighting conditions and seat position. This could include conducting gaze calibration in shaded environments or involving participants directly in the camera calibration process to ensure alignment with their actual seating preferences.

Cognitive Load Measurement: Pupil dilation is commonly used in eye-tracking studies to assess cognitive load (Bitkina et al., 2021; Palinko et al., 2010; Recarte & Nunes, 2003; J. T.-y. Wang, 2011). This method was considered in this study but ultimately excluded due to the uncontrolled nature of the real-world driving environment. Pupil size can be influenced by various external factors such as ambient lighting, fatigue, and cognitive idleness, which makes the identification of the cognitive component challenging (Vilotijevi & Mathôt, 2024). As noted by Farnsworth (2022), accurate interpretation of pupil dilation requires controlling the participant's light exposure. The brightness of the visual stimuli must be kept consistent to ensure changes in pupil size reflect cognitive or emotional processes rather than a natural reaction to light intensity. However, such standardisation is difficult to achieve in a naturalistic driving context. Future studies may also explore non-visual indicators of cognitive load, such as integrating biometric data (e.g., EEG, EMG) for a more comprehensive analysis.

6.4.3 Time Limitations

As this study was conducted as a master's thesis project, time constraints significantly shaped both the design and scope of the research. The limited time affected the study design, participant recruitment, and comprehensive data analysis.

Limited Focus on Driving Scenarios and HUD Functions: As mentioned in Section 4.2.4 and 4.3, each participant session was designed to last approximately 60 minutes, with around 10 minutes allocated to the actual driving task for each test route. This time limitation restricted the diversity of driving scenarios that could be included, such as long-distance driving, HUD functions (e.g., map view, speed limit, or cruise control), or more complex navigation tasks (e.g., city-centre driving, dynamic lane guidance, or construction detours). The short-term nature of the sessions made it difficult to assess long-term adaptation to HUD systems. Some participants reported initial discomfort or uncertainty when using the HUD, suggesting that a learning curve may influence perceived usability. Longitudinal studies would offer valuable insight into whether such discomfort decreases over time with continued exposure, and whether sustained use leads to improved driving performance, user confidence, and trust in HUDs.

Additionally, the study was conducted exclusively during daytime, which further limited the ability to evaluate HUD performance under different lighting conditions. Expanding the study to include both day and night sessions would provide a more comprehensive understanding of display legibility, user comfort, and visual distraction across varying ambient light levels.

Sample Size and Statistical Power: Due to time limitations inherent in the scope of a master's thesis project conducted by two students, the number of participants was restricted, which directly influenced the choice of statistical methods. Due to the modest sample size, non-parametric tests were employed, as these methods are more appropriate when data distribution assumptions cannot be met. However, non-parametric methods generally offer lower statistical power compared to parametric tests. As noted by Nahm (2016), non-parametric analyses help reduce the risk of incorrect conclusions because they do not rely on assumptions about the population, but they are also less sensitive in detecting significant effects compared to parametric tests. To increase statistical power and enable the use of parametric analyses, future studies should aim to recruit a larger and more diverse sample. Specifically, reaching a sample size of at least 30 participants is often recommended to approximate a normal distribution and satisfy the assumptions underlying parametric tests such as t-tests or ANOVAs (Daniel and Wood, 1980; Devore, 2011; Kim and Park, 2019; Kwak and Kim, 2017).

Underutilised Contextual Data: Lastly, although contextual data such as weather conditions, time of day, and the frequency and reasons for missed navigation were documented, they were not incorporated into the final analysis. The limited time available for both data analysis and thesis completion restricted the scope of variables that could be explored. As a result, potentially valuable insights into how environmental conditions influence driver behaviour and HUD usability were not fully realised. Future studies with extended timelines could integrate these factors to provide a more nuanced and holistic understanding of the influence of in-vehicle displays with deeper contextual insights.

6.4.4 Scope of Available Resources

The findings of this study are closely tied to the specific tools and technologies employed. These include the conventional HUD in the Volvo XC60, Google Maps for navigation, the Smart Eye Pro eye-tracking system, and the iMotions eye-tracking analysis software. As a result, generalising the outcomes to other HUD models, navigation systems, or eye-tracking data collection and analysis tools requires caution, since hardware and software variations may influence tracking accuracy, precision, and tracking robustness (Vehlen et al., 2021).

In addition, the eye-tracking analysis primarily relied on *Gaze Dwell Time* (%) within the defined AOIs. The gaze-based metric captured the total gaze duration which was automatically calculated for each component in the world model. Although *Gaze Dwell Time* reflects visual engagement, it is not common to use raw gaze points in eye-tracking research (Blascheck et al., 2014). As discussed in Section 3.3.1, gaze includes both fixations and saccades, which is aggregated gaze points. Fixations and saccades are more common eye-tracking metrics (Ke et al., 2024), which could have provided more detailed insight into visual attention and information processing flow. However, the current version of iMotions software used in this study did not support the calculation of fixation- or saccade-based metrics for world model AOIs, limiting the analytical depth. If given more time, post processing of the raw eye-tracking data could have been done to extract fixation and saccade data to more accurately infer visual processing and create visual attention patterns. For example, tracking how a participant's gaze moves from one component to another. Future studies could benefit from alternative software or enhanced analysis tools that allow access to a broader range of gaze data.

6.5 Ethical Considerations and Implications

Ethical considerations in the study included ensuring a safe test environment by minimising potential external factors that could have distracted the participants while driving. This included careful task selection, excluding non-naturalistic activities such as introducing additional tasks during driving, for instance, requiring participants to activate cruise control. Consequently, activities like conversation or playing music while driving were deliberately avoided. During eye-tracking installation, the idea of mounting the cameras on a metal pole for greater stability was discarded as it was considered too intrusive for the driver. Instead, the cameras were attached to the dashboard surface, compromising the quality of the data to prioritise traffic safety. In addition, VCC's in-vehicle experiment guidelines were strictly adhered to throughout the study. This included obtaining driving permits for public road testing, complying with all permit regulations, and ensuring the test vehicle was properly maintained for safe driving.

Throughout the design of the user study, ethical considerations were made in compliance with established ethical guidelines. The Declaration of Helsinki (World Medical Association, 2024), mainly ensured respect for participants' rights and well-being, while the GDPR (European Union, 2016) safeguarded the privacy and confidentiality of participants' data. To comply with the ethical guidelines, informed consent was provided to participants at the beginning of each test. The consent form primarily focused on transparency regarding the researchers background, the study's objectives, data handling, as well as informing participants of their right to withdraw at any point during the study, or even afterwards.

According to the GDPR (European Union, 2016), biometric data is classified as sensitive personal data under special categories that require a higher level of protection in data handling. The use of biometric data should be limited to contexts such as healthcare, research, or activities that lawfully serve the public interest. GDPR *Article 4(14)* provides the following definition of biometric data:

> "*Biometric data means personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data;*"

Although in some contexts, eye-tracking data may constitute biometric data under the named article, no authentication or identification of individuals was performed in this study. Thus, within the scope of this study, eye-tracking data is not considered biometric data based on the GDPR's definition. The eye-tracking data was collected solely for research purposes to study gaze behaviour while driving. The data was pseudonymised and stored securely, preventing the unique identification of participants. Additionally, recordings from the Scene Camera include some facial and body images of participants; however, these were not processed or published for unique identification either.

In light of the studies discussed in Section 3.3.4, although eye-tracking data in this study was performed responsibly with informed consent, future research should continue to consider its potential sensitivity. Researchers must ensure transparent consent processes, address the possibility of implicit data collection, and apply robust privacy safeguards to protect participant rights in line with evolving ethical standards.

7

Conclusion

This study explored the influence of a conventional HUD on driver behaviour in terms of attention, distraction, and cognitive load in a navigational context, to evaluate its impact on safety and usability in a real-road traffic environment. The TA suggested usability concerns related to the HUD's UI and UX, particularly regarding the navigation feature, which was seen as a potential source of distraction. Participants criticised the feature as insufficient, poorly timed, and inaccurate, or expressed indifference toward it. However, the idea of a HUD positioned within the driver's line of sight was well received and perceived as enhancing safety. Participants were particularly positive about the speedometer feature. Consequently, the shift in attention across the displays was more likely due to participants checking the speedometer or distracted by the navigation information.

As a result, opinions were divided about the usefulness of the HUD navigation. The statistical analysis proved that drivers' visual attention on the CSD, DIM, and road, decreased significantly with the presence of a HUD over the entire recording. Significant dwell time differences also emerged in Events 2, 4, 5, and 6. These events involved various traffic scenarios, including highway exits, roundabouts, merging lanes, and lane changes, and resulted in significant decreases across different AOIs.

These findings suggest that while the HUD may redirect drivers' attention from the road, this shift does not necessarily imply reduced safety depending on context and activated HUD feature. However, the significant reduction in attention to the road during critical and complex driving scenarios highlights the need for further investigation to fully understand the implications for driver safety and HUD usability.

When looking into the significant differences in cognitive load between HUD and non-HUD conditions, no significant differences were found in the overall Weighted or Raw TLX. The only significant variation occurred in the Weighted Physical Demand subscale. This difference appears unrelated to the HUD's content or functionality, but rather from visual discomfort reported by some participants. While visual discomfort could compromise the usability of the HUD and potentially cause safety concerns, it did not significantly affect the overall cognitive load. This suggests that the physical demand was not overwhelmingly high to elevate the overall cognitive load. Thus, it can be concluded that the impact of the HUD on participants' overall perceived cognitive load was minimal. Consequently, the HUD does not present notable safety or usability concerns for most users, while a subset of users may experience discomfort affecting both aspects.

However, as discussed in the previous chapter, these results must be interpreted within the context of several limitations. The participant sample lacked sufficient diversity in gender, age, and driving experience, and all participants were recruited from a single automotive company, which may have introduced biases. Additionally, real-world constraints, such as variable traffic conditions, safety concerns, and eye-tracking calibration inconsistencies, complicated the standardisation of data collection and analysis. Time and resource limitations also constrained the range of test scenarios and depth of evaluation. Despite these challenges, the study contributes valuable insights into the complex dynamics of HUD-supported driving. Future research should prioritise more diverse sampling, improved balance between realistic driving conditions and experimental control, and multimodal HUD functions (e.g., map view, speed limit, or cruise control) and objective cognitive load assessment (e.g., through biometric data integration) to build a more generalisable and robust understanding of HUD design and its implications for HMI usability and driver safety.

It is also important to acknowledge that the findings in this study may not extend to all types of automotive HUDs, as variations in functions, placement, display size, and projection techniques can influence usability and driver interaction. Similarly, different eye-tracking systems may vary in accuracy, sampling rate, and calibration reliability, which can affect the quality and interpretation of gaze data.

References

AblaSSmeier, M., Poitschke, T., Wallhoff, F., Bengler, K., & Rigoll, G. (2007). Eye gaze studies comparing head-up and head-down displays in vehicles. *2007 IEEE International Conference on Multimedia and Expo*, 2250–2252.

Albert, B., & Tullis, T. (2013). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics*. Newnes.

Albert, W., & Tedesco, D. (2010). Reliability of self-reported awareness measures based on eye tracking. *Journal of Usability Studies*, *5*(2), 50–64.

Amini, R. E., Al Haddad, C., Batabyal, D., Gkena, I., De Vos, B., Cuenen, A., Brijs, T., & Antoniou, C. (2023). Driver distraction and in-vehicle interventions: A driving simulator study on visual attention and driving performance. *Accident Analysis & Prevention*, *191*, 107195.

Badger, D., Nursten, J., Williams, P., & Woodward, M. (2000). Should all literature reviews be systematic? *Evaluation & Research in Education*, *14*(3-4), 220–230.

Baron, R. S. (1986). Distraction-conflict theory: Progress and problems. *Advances in experimental social psychology*, *19*, 1–40.

Barry, E. S., Merkebu, J., & Varpio, L. (2022). State-of-the-art literature review methodology: A six-step approach for knowledge synthesis. *Perspectives on Medical Education*, *11*(5), 281–288.

Baxter, K., Courage, C., & Caine, K. (2015). *Understanding your users: A practical guide to user research methods*. Morgan Kaufmann.

Beanland, V., Fitzharris, M., Young, K. L., & Lenné, M. G. (2013). Driver inattention and driver distraction in serious casualty crashes: Data from the australian national crash in-depth study. *Accident Analysis & Prevention*, *54*, 99–107.

Bell, C. (1823). Xv. on the motions of the eye, in illustration of the uses of the muscles and nerves of the orbit. *Philosophical Transactions of the Royal Society of London*, (113), 166–186.

Bertram, D. (2007). Likert scales. *Retrieved November*, *2*(10), 1–10.

Betancur, J. A., Villa-Espinal, J., Osorio-Gómez, G., Cuéllar, S., & Suárez, D. (2018). Research topics and implementation trends on automotive head-up display systems. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *12*, 199–214.

Bitkina, O. V., Park, J., & Kim, H. K. (2021). The ability of eye-tracking metrics to classify and predict the perceived driving workload. *International Journal of Industrial Ergonomics*, *86*, 103193.

Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., & Ertl, T. (2014). State-of-the-art of visualization for eye tracking data. *Eurovis (stars)*, 29.

Broadbent, D. E. (1958). *Perception and communication.* Pergamon Press.

Brown, B., & Laurier, E. (2012). The normal natural troubles of driving with gps. *Proceedings of the SIGCHI conference on human factors in computing systems*, 1621–1630.

Bullock, C. L. (2019). Driving new narratives: Women-leader identities in the automotive industry. *Gender in Management: An International Journal*, *34*(3), 200–216.

Caine, K. (2016). Local standards for sample size at chi. *Proceedings of the 2016 CHI conference on human factors in computing systems*, 981–992.

Cameron, A. (2015). In the blink of an eye: Head mounted displays development within bae systems. *Display Technologies and Applications for Defense, Security, and Avionics IX; and Head-and Helmet-Mounted Displays XX*, *9470*, 181–192.

Campisi, T., Skoufas, A., Kaltsidis, A., & Basbas, S. (2021). Gender equality and e-scooters: Mind the gap! a statistical analysis of the sicily region, italy. *Social Sciences*, *10*(10). https://doi.org/10.3390/socsci10100403

Carroll, J. M. (2009). Human computer interaction (hci). *Encyclopedia of Human-Computer Interaction. Aarhus: The Interaction Design Foundation.*

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49–62.

Carter, C. R., Maki, J., Ackermann, N., & Waters, E. A. (2023). Inclusive recruitment strategies to maximize sociodemographic diversity among participants: A st. louis case study. *MDM Policy & Practice*, *8*(1), 23814683231183646.

Cheng, Y.-n., Zhong, X., & Tian, L.-w. (2023). Does the ar-hud system affect driving behaviour? an eye-tracking experiment study. *Transportation research interdisciplinary perspectives*, *18*, 100767.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the acoustical society of America*, *25*, 975–979.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* Elsevier Science & Technology.

Crabtree, B. F., & Miller, W. L. (1992). Doing qualitative research. *Annual North American Primary Care Research Group Meeting, 19th, May, 1989, Quebec, PQ, Canada.*

Daniel, C., & Wood, F. S. (1980). *Fitting equations to data: Computer analysis of multifactor data.* John Wiley & Sons, Inc.

Demberg, V., Sayeed, A., Mahr, A., & Müller, C. (2013). Measuring linguistically-induced cognitive load during driving using the contre task. *Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular applications*, 176–183.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological review*, *70*(1), 80.

Devore, J. L. (2011, January). *Probability and statistics for engineering and the sciences* (8th) [ISBN-13: 978-0-538-73352-6]. Brooks/Cole.

Di Zio, M., Fursova, N., Gelsema, T., GieSSing, S., Guarnera, U., Petrauskien, J., Quensel-von Kalben, L., Scanu, M., ten Bosch, K., van der Loo, M., et al. (2016). Methodology for data validation 1.0. *Essnet Validat Foundation.*

Dombrovskis, S. (2010). Implementation and real-traffic assessment of a new infotainment interface concept.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society annual meeting*, *32*(2), 97–101.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human factors*, *37*(1), 32–64.

Endsley, M. R., & Connors, E. S. (2008). Situation awareness: State of the art. *2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, 1–4.

Endsley, M. (2020). Situation awareness in driving. *Handbook of human factors for automated, connected and intelligent vehicles. London: Taylor and Francis.*

Eugensson, A., & Ivarsson, J. (2022). What the car industry can do, volvo cars. In *The vision zero handbook: Theory, technology and management for a zero casualty policy* (pp. 1–39). Springer.

European Union. (2016). Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (general data protection regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

Evans, K. K., Horowitz, T. S., Howe, P., Pedersini, R., Reijnen, E., Pinto, Y., Kuzmova, Y., & Wolfe, J. M. (2011). Visual attention. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 503–514.

Farnsworth, B. (2022). *Pupillometry 101: Everything you need to know* [Accessed: 2025-05-26]. https://imotions.com/blog/learning/best-practice/pupillometry-101/

Fracker, M. L. (1988). A theory of situation assessment: Implications for measuring situation awareness. *Proceedings of the Human Factors Society Annual Meeting*, *32*(2), 102–106.

Fu, Z., Liu, C., Peng, J., Peng, L., & Qin, S. (2023). Prediction of automobile aerodynamic drag coefficient for suv cars based on a novel xgboost model. *Iranian Journal of Science and Technology, Transactions of Mechanical Engineering*, *47*(4), 1349–1364.

Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International journal of psychophysiology*, *83*(3), 269–275.

Galy, E., Paxion, J., & Berthelon, C. (2018). Measuring mental workload with the nasa-tlx needs to examine each dimension rather than relying on the global score: An example with driving. *Ergonomics*, *61*(4), 517–527.

Gauld, C. S., Lewis, I., White, K. M., Fleiter, J. J., & Watson, B. (2017). Smartphone use while driving: What factors predict young drivers' intentions to initiate, read, and respond to social interactive technology? *Computers in Human Behavior*, *76*, 174–183.

Gish, K. W., & Staplin, L. (1995). Human factors aspects of using head up displays in automobiles: A review of the literature.

Glaser, B., & Strauss, A. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

Gong, Z. (2025). Challenges and opportunities of automotive hmi. *International Conference on Human-Computer Interaction*, 23–35.

Gulino, M.-S., Vichi, G., Cecchetto, F., Di Lillo, L., & Vangi, D. (2025). A combined comfort and safety-based approach to assess the performance of advanced driver assistance functions. *European Transport Research Review*, *17*(1), 1–16.

Guo, W., Yan, D., Liu, T., & Zhang, Z. (2020). Technical challenge and solution for vehicle-mounted ar-hud mass commercial application. *International Conference on Optoelectronic and Microelectronic Technology and Application*, *11617*, 964–974.

Gupta, S., Kohavi, R., Tang, D., Xu, Y., Andersen, R., Bakshy, E., Cardin, N., Chandran, S., Chen, N., Coey, D., et al. (2019). Top challenges from the first practical online controlled experiments summit. *ACM SIGKDD Explorations Newsletter*, *21*(1), 20–35.

Hansols, J. (2022). Interaction design for augmented reality head-up displays: Developing graphics design and evaluating perceptual and safety aspects for navigation use cases.

Harms, L., & Patten, C. (2003). Peripheral detection as a measure of driver distraction. a study of memory-based versus system-based navigation in a built-up area. *Transportation Research Part F: Traffic Psychology and Behaviour*, *6*(1), 23–36.

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the human factors and ergonomics society annual meeting*, *50*(9), 904–908.

Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology* (pp. 139–183, Vol. 52). Elsevier.

Hartson, R., & Pyla, P. S. (2012, January 1). *The UX book : Process and guidelines for ensuring a quality user experience*. Morgan Kaufmann. Retrieved February 21, 2025, from https://research.ebsco.com/linkprocessor/plink?id=c45ae858-a567-39b8-b3cd-cca1d70ba21a

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

Hofmaenner, D. A., Herling, A., Klinzing, S., Wegner, S., Lohmeyer, Q., Schuepbach, R. A., & Buehler, P. K. (2021). Use of eye tracking in analyzing distribution of visual attention among critical care nurses in daily professional life: An observational study. *Journal of Clinical Monitoring and Computing*, *35*, 1511–1518.

Hurtado, S., & Chiasson, S. (2016). An eye-tracking evaluation of driver distraction and unfamiliar road signs. *Proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications*, 153–160.

Iatropoulos, J., Panzer, A., & Henze, R. (2024). *Analysis of human driving behavior with focus on vehicle lateral control* (tech. rep.). SAE Technical Paper.

iMotions. (2024). Eye tracking the complete guide [Accessed: 2025-02-28]. https://go.imotions.com/eye-tracking-guide

iMotions. (2025). Eye tracking infographic [Accessed: 2025-03-04]. https://imotions.com/wp-content/uploads/2025/01/eye-tracking-infographics_print.pdf

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye* (pp. 573–605). Elsevier.

Jakus, G., Dicke, C., & Sodnik, J. (2015). A user study of auditory, head-up and multi-modal displays in vehicles. *Applied ergonomics*, *46*, 184–192.

Jin, C.-x., You, F., & Wang, J.-m. (2016). Automotive hud interaction design based on lane changing scenario. *Human-Computer Interaction. Novel User Experiences: 18th International Conference, HCI International 2016, Toronto, ON, Canada, July 17-22, 2016. Proceedings, Part III 18*, 277–284.

Jupp, V. (2006). The sage dictionary of social research methods.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Citeseer.

Kang, H.-B. (2013). Various approaches for driver and driving behavior monitoring: A review. *2013 IEEE International Conference on Computer Vision Workshops*, 616–623. https://doi.org/10.1109/ICCVW.2013.85

Ke, F., Liu, R., Sokolikj, Z., Dahlstrom-Hakki, I., & Israel, M. (2024). Using eye-tracking in education: Review of empirical research and technology. *Educational technology research and development*, *72*(3), 1383–1418.

Khedkar, S., Kasav, S., & Mahajan, S. (2015). Head up display techniques in cars. *International Journal of Engineering Science and Innovative Technology*, *4*, 119–124.

Kiefer, R. J., & Gellatly, A. W. (1996). Quantifying the consequences of the" eyes-on-road" benefit attributed to head-up displays. *SAE transactions*, 1219–1240.

Kim, T. K., & Park, J. H. (2019). More about the basic assumptions of t-test: Normality and sample size. *Korean journal of anesthesiology*, *72*(4), 331–335.

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy online controlled experiments: A practical guide to a/b testing*. Cambridge University Press.

Korkmaz, A., & Gulsecen, S. (2023, August). The intersection of ethics and biometric data in eye-tracking systems. https://doi.org/10.59287/irshas.402

Kosch, T., Karolus, J., Zagermann, J., Reiterer, H., Schmidt, A., & Woniak, P. W. (2023). A survey on measuring cognitive workload in human-computer interaction. *ACM Computing Surveys*, *55*(13s), 1–39.

Kosslyn, S. M. (1985). Graphics and human information processing: A review of five books. *Journal of the American Statistical Association*, *80*(391), 499–512.

Kotronoulas, G., Miguel, S., Dowling, M., Fernández-Ortega, P., Colomer-Lahiguera, S., Baçivan, G., Pape, E., Drury, A., Semple, C., Dieperink, K. B., et al. (2023). An overview of the fundamentals of data management, analysis, and interpretation in quantitative research. *Seminars in oncology nursing*, *39*(2), 151398.

Krassanakis, V., Filippakopoulou, V., & Nakos, B. (2014). Eyemmv toolbox: An eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification. *Journal of Eye Movement Research*, *7*(1).

Kröger, J. L., Lutz, O. H.-M., & Müller, F. (2020). What does your gaze reveal about you? on the privacy implications of eye tracking. *IFIP International Summer School on Privacy and Identity Management*, 226–241.

Kwak, S. G., & Kim, J. H. (2017). Central limit theorem: The cornerstone of modern statistics. *Korean journal of anesthesiology*, *70*(2), 144.

Lakshman, M., Sinha, L., Biswas, M., Charles, M., & Arora, N. K. (2000). Quantitative vs qualitative research methods. *The Indian Journal of Pediatrics*, *67*, 369–377.

Langlois, S. (2013). Adas hmi using peripheral vision. *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 74–81.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in cognitive sciences*, *9*(2), 75–82.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current directions in psychological science*, *19*(3), 143–148.

Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of experimental psychology: General*, *133*(3), 339.

Li, J., Zhang, W., Feng, Z., Wei, L., Tang, T., & Gu, T. (2024). Effects of head-up display information layout design on driver performance: Driving simulator studies. *International Journal of Human–Computer Interaction*, 1–17.

Li, Y., Wang, Y., Song, F., & Liu, Y. (2024). Assessing gender perception differences in color combinations in digital visual interfaces using eye tracking–the case of hud. *International Journal of Human–Computer Interaction*, *40*(20), 6591–6607.

Liang, Y. (2009). *Detecting driver distraction* [Doctoral dissertation] [Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2023-03-02]. http://proxy.lib.chalmers.se/login?url=https://www.proquest.com/dissertations-theses/detecting-driver-distraction/docview/304900929/se-2

Lin, Y., Leng, H., Yang, G., & Cai, H. (2007). An intelligent noninvasive sensor for driver pulse wave measurement. *IEEE Sensors Journal*, *7*(5), 790–799.

Liu, Y.-C., & Wen, M.-H. (2004). Comparison of head-up display (hud) vs. head-down display (hdd): Driving performance of commercial vehicle operators in taiwan. *International Journal of Human-Computer Studies*, *61*(5), 679–697. https://doi.org/https://doi.org/10.1016/j.ijhcs.2004.06.002

Lloyd, H., & Mey, M. (2007). Gender differences in perceptions of workplace progression: An automotive industry case study. *Southern African Business Review*, *11*(3), 95–120.

Lockhart, T. E., & Shi, W. (2010). Effects of age on dynamic accommodation. *Ergonomics*, *53*(7), 892–903.

Luo, S., Yi, X., Shao, Y., & Xu, J. (2022). Effects of distracting behaviors on driving workload and driving performance in a city scenario. *International journal of environmental research and public health*, *19*(22), 15191.

Ma, J., & Gong, Z. (2024). *Automotive human-machine interaction (hmi) evaluation method*. Springer.

Ma, Y., Hu, B., Chan, C.-Y., Qi, S., & Fan, L. (2018). Distractions intervention strategies for in-vehicle secondary tasks: An on-road test assessment of driving task demand based on real-time traffic environment. *Transportation Research Part D: Transport and Environment*, *63*, 747–754.

Mahlke, S. (2005). Understanding users' experience of interaction. *Proceedings of the 2005 annual conference on European association of cognitive ergonomics*, 251–254.

Maroto, M., Caño, E., González, P., & Villegas, D. (2018). Head-up displays (hud) in driving. *arXiv preprint arXiv:1803.08383*.

Marr, D. (1982). Representing the image: Zero-crossings and the raw primal sketch. *Vision. San Francisco, CA: WH Freeman*, 54–68.

Maylor, E. A., & Lavie, N. (1998). The influence of perceptual load on age differences in selective attention. *Psychology and aging*, *13*(4), 563.

Mishra, P., Pandey, C. M., Singh, U., Keshri, A., & Sabaretnam, M. (2019). Selection of appropriate statistical methods for data analysis. *Annals of cardiac anaesthesia*, *22*(3), 297–301.

Nahm, F. S. (2016). Nonparametric statistical tests for the continuous data: The basic concept and the practical use. *Korean journal of anesthesiology*, *69*(1), 8–14.

National Highway Traffic Safety Administration (NHTSA). (2010). *Overview of the national highway traffic safety administrations driver distraction program* (tech. rep. No. DOT HS 811 299). National Highway Traffic Safety Administration. Washington, DC. https://www.nhtsa.gov/sites/nhtsa.gov/files/811299.pdf

National Highway Traffic Safety Administration (NHTSA). (2012). *Visual-manual nhtsa driver distraction guidelines for in-vehicle electronic devices* (tech. rep. No. Docket No. NHTSA-2010-0053). U.S. Department of Transportation, National Highway Traffic Safety Administration. Washington, DC. https://www.nhtsa.gov/sites/nhtsa.gov/files/distraction_npfg-02162012.pdf

Neumann, T. (2024). Analysis of advanced driver-assistance systems for safe and comfortable driving of motor vehicles. *Sensors*, *24*(19), 6223.

Nidamanuri, J., Nibhanupudi, C., Assfalg, R., & Venkataraman, H. (2021). A progressive review: Emerging technologies for adas driven solutions. *IEEE Transactions on Intelligent Vehicles*, *7*(2), 326–341.

Niehorster, D. C., Cornelissen, T. H., Holmqvist, K., Hooge, I. T., & Hessels, R. S. (2018). What to expect from your remote eye-tracker when participants are unrestrained. *Behavior research methods*, *50*, 213–227.

Nightingale, A. (2009). A guide to systematic literature reviews. *Surgery (Oxford)*, *27*(9), 381–384.

Niu, X., & Tang, T. (2024). A study on automotive hmi design evaluation method based on usability test metrics and xgboost algorithm. *International Conference on Human-Computer Interaction*, 225–235.

Noyes, J. M., & Bruneau, D. P. (2007). A self-analysis of the nasa-tlx workload measure. *Ergonomics*, *50*(4), 514–519.

Oh, H. J., Ko, S. M., & Ji, Y. G. (2016). Effects of superimposition of a head-up display on driving performance and glance behavior in the elderly. *International Journal of Human-Computer Interaction*, *32*(2), 143–154.

Okabayashi, S., Sakata, M., & Hatada, T. (1991). Drivers ability to recognize objects in the forward view with superimposition of head-up display images. *Proceedings of the Society for Information Display*, *32*, 465–468.

Okabayashi, S., Sakata, M., Fukano, J., Daidoji, S., Hashimoto, C., & Ishikawa, T. (1989). Development of practical heads-up display for production vehicle application. *SAE Transactions*, 638–647.

Olsen, W. (2004). Triangulation in social research: Qualitative and quantitative methods can really be mixed. *Developments in sociology*, *20*, 103–118.

Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. *Handbook of mixed methods in social and behavioral research*, *2*(1), 397–430.

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior research methods*, *50*, 1645–1656.

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. *Proceedings of the 2010 symposium on eye-tracking research & applications*, 141–144.

Patel, P. (2009). Introduction to quantitative methods. *Empirical Law Seminar*, *14*(1), 1–14.

Posner, M. I., Cohen, Y., et al. (1984). Components of visual orienting. *Attention and performance X: Control of language processes*, *32*, 531–556.

Prat, F., Planes, M., Gras, M., & Sullman, M. (2015). An observational study of driving distractions on urban roads in spain. *Accident Analysis & Prevention*, *74*, 8–16.

Proctor, R. W., & Vu, K.-P. L. (2007). Human information processing: An overview for human-computer interaction. *The human-computer interaction handbook*, 69–88.

Punde, P. A., Jadhav, M. E., & Manza, R. R. (2017). A study of eye tracking technology and its applications. *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, 86–90.

Quin, F., Weyns, D., Galster, M., & Silva, C. C. (2024). A/b testing: A systematic literature review. *Journal of Systems and Software*, *211*, 112011. https://doi.org/https://doi.org/10.1016/j.jss.2024.112011

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of experimental psychology: Applied*, *9*(2), 119.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In *Advances in psychology* (pp. 185–218, Vol. 52). Elsevier.

Reingold, E. M. (2014). Eye tracking research and technology: Towards objective measurement of data quality. *Visual cognition*, *22*(3-4), 635–652.

RISE. (2023). Safe chauffeurs in safe and healthy multimodal driver information environments [Accessed: 2025-01-28]. https://www.ri.se/en/expertise-areas/projects/safe-chauffeurs-in-safe-and-healthy-multimodal-driver-information-env

Robson, C., & McCartan, K. (2016). Real world research, 4th edn. hokoben.

Russell, S. M., Radlbeck, J., Atwood, J., Schaudt, W. A., McLaughlin, S. B., et al. (2023). *Head-up displays and distraction potential* (tech. rep.). United States. Department of Transportation. National Highway Traffic Safety

Rutley, K. (1975). Control of drivers' speed by means other than enforcement. *Ergonomics*, *18*(1), 89–100.

Sauro, J., & Lewis, J. R. (2016). *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann.

Sharp, H., Preece, J., & Rogers, Y. (2019). *Interaction design : Beyond human-computer interaction* (Fifth edition, Vol. null). Wiley. https://research.ebsco.com/linkprocessor/plink?id=dffd37b0-ef0d-3e80-975f-59360548946c

Singh, H., & Kathuria, A. (2021). Analyzing driver behavior under naturalistic driving conditions: A review. *Accident Analysis & Prevention*, *150*, 105908.

Siroker, D., & Koomen, P. (2015). *A/b testing: The most powerful way to turn clicks into customers*. John Wiley & Sons.

Smart Eye AB. (2002, August). *Smart eye pro 5.0 user manual* (Revision 249) [Accessed: 2025-07-05]. https://reshare.ukdataservice.ac.uk/851661/13/SmartEyePror-UserManual.pdf

Smith, M., Gabbard, J. L., & Conley, C. (2016). Head-up vs. head-down displays: Examining traditional methods of display assessment while driving. *Proceedings of the 8th international conference on Automotive User Interfaces and Interactive Vehicular Applications*, 185–192.

Sojourner, R. J., & Antin, J. F. (1990). The effects of a simulated head-up display speedometer on perceptual task performance. *Human factors*, *32*(3), 329–339.

Stojmenova, K., & Sodnik, J. (2015). Methods for assessment of cognitive workload in driving tasks. *ICIST 2015 5th International Conference on Information Society and Technology*, 229–234.

Stutts, J. C., Reinfurt, D. W., Staplin, L., Rodgman, E., et al. (2001). The role of driver distraction in traffic crashes.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*, 251–296.

Sweller, J., Van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational psychology review*, *31*, 261–292.

Tangmanee, K., & Teeravarunyou, S. (2012). Effects of guided arrows on head-up display towards the vehicle windshield. *2012 Southeast Asian network of ergonomics societies conference (SEANES)*, 1–6.

Thalen, J. (2006, August). Adas for the car of the future. http://essay.utwente.nl/58373/

Treisman, A. M. (1964). Selective attention in man. *British medical bulletin.*

Turobov, A., Coyle, D., & Harding, V. (2024). Using chatgpt for thematic analysis. *arXiv preprint arXiv:2405.08828.*

van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies. *Nursing standard (Royal College of Nursing (Great Britain) : 1987)*, *16*, 33–6. https://doi.org/10.7748/ns2002.06.16.40.33.c3214

Vehlen, A., Spenthof, I., Tönsing, D., Heinrichs, M., & Domes, G. (2021). Evaluation of an eye tracking setup for studying visual attention in face-to-face conversations. *Scientific reports*, *11*(1), 2661.

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

Vilotijevi, A., & Mathôt, S. (2024). Functional benefits of cognitively driven pupil-size changes. *Wiley Interdisciplinary Reviews: Cognitive Science*, *15*(3), e1672.

von Janczewski, N., Kraus, J., Engeln, A., & Baumann, M. (2022). A subjective one-item measure based on nasa-tlx to assess cognitive workload in driver-vehicle interaction. *Transportation research part F: traffic psychology and behaviour*, *86*, 210–225.

Wadsworth, Y. (2020). *Do it yourself social research: The bestselling practical guide to doing social research projects*. Routledge.

Wang, J. T.-y. (2011). Pupil dilation and eye tracking. *A handbook of process tracing methods for decision research: A critical review and users guide*, 185–204.

Wang, X., & Qin, Q. (2014). Vehicle head-up display technology. *Electro-Optics and Control*, *47*(1), 55–58.

Ward, N. J., & Parkes, A. (1994). Head-up displays and their automotive application: An overview of human factors issues affecting safety. *Accident Analysis & Prevention*, *26*(6), 703–717.

Wege, C., Will, S., & Victor, T. (2013). Eye movement and brake reactions to real world brake-capacity forward collision warningsa naturalistic driving study. *Accident Analysis & Prevention*, *58*, 259–270.

Weihrauch, M., Meloeny, G. G., & Goesch, T. C. (1989). The first head up display introduced by general motors. *SAE International Congress and Exposition.* https://doi.org/https://doi.org/10.4271/890288

Wickens, C. (1984). Processing resources and attention, varieties of attention. *R. Parasuraman and D. Davis, Eds. Academic Press.*

Wildemuth, B. M. (2003). Why conduct user studies? the role of empirical evidence in improving the practice of librarianship. *Inforum*, *2003*, 9th.

World Medical Association. (2024). Wma declaration of helsinki – ethical principles for medical research involving human subjects [Last amended 2024. Accessed: 2025-07-04]. https://www.wma.net/policies-post/wma-declaration-of-helsinki/

Xia, T., Lin, X., Sun, Y., & Liu, T. (2023). An empirical study of the factors influencing users intention to use automotive ar-hud. *Sustainability*, *15*(6), 5028.

Yang, G., Lin, Y., & Bhattacharya, P. (2010). A driver fatigue recognition model based on information fusion and dynamic bayesian network. *Information Sciences*, *180*(10), 1942–1954.

Yantis, S., et al. (2000). Goal-directed and stimulus-driven determinants of attentional control. *Attention and performance*, *18*(Chapter 3), 73–103.

Yardm, S., & Pedgley, O. (2023). Targeting a luxury driver experience: Design considerations for automotive hmi and interiors. *International Journal of Design*, *17*(2), 45–66.

You, F., Li, Y., Schroeter, R., Friedrich, J., & Wang, J. (2017). Using eye-tracking to help design hud-based safety indicators for lane changes. *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*, 217–221.

Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring cognitive load using eye tracking technology in visual computing. *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 78–85. https://doi.org/10.1145/2993901.2993908

Zhang, X., Liao, X.-P., & Tu, J.-C. (2022). A study of bibliometric trends in automotive humanmachine interfaces. *Sustainability*, *14*(15). https://doi.org/10.3390/su14159262

Zhang, Y., Huang, Y., Chen, H., Na, X., Chen, Z., & Liu, Y. (2021). Driving behavior oriented torque demand regulation for electric vehicles with single pedal driving. *Energy*, *228*, 120568.

Zhou, C., Qiao, W., Hua, J., & Chen, L. (2024). Automotive augmented reality head-up displays. *Micromachines*, *15*(4). https://doi.org/10.3390/mi15040442

A

Experimental Equipment Toolbox

A.1 Hardware Components

- **Test Vehicle**: *Volvo XC60*[1]—Equipped with a conventional HUD to provide a realistic test environment for in-vehicle display study.
- **Eye-tracking Cameras**: *Smart Eye Pro*[2]–Three eye-tracking cameras (two with IR lights, one without) were used to collect eye-tracking data.
- **Environment Camera**: *Logitech C920*[3]—A webcam was mounted on the dashboard, facing outward to capture traffic conditions.
- **Scene Camera**: *Microsoft LifeCam Studio 1080p HD*—A webcam was mounted next to the drivers headrest to capture the displays.
- **GPS**: *Navisys AG3335*[4]—A GPS device was used to track the real-time driving route, recording both coordinates and speed.
- **Lab Computer**: *HP ZBook Powder 15.6 inch G10 Mobile Workstation PC*[5]—A high-performance computer running Windows 11 and equipped with iMotions software was connected to the Environment Camera, Scene Camera, and GPS to collect eye-tracking data and conduct analysis. It also served to remotely control the Smart Eye computer.
- **Smart Eye Computer**: A computer provided by Smart Eye was installed with Smart Eye Pro software for creating the World Model and calibrating the eye-tracking cameras.
- **Smart Eye Frame Grabber**: Acquire images from the three eye-tracking cameras.

---

[1] Volvo XC60:
https://www.volvocars.com/az/support/car/xc60/article/9422b84288b9ecccc0a801510e8c4bed/

[2] Smart Eye Pro: https://www.smarteye.se/smart-eye-pro/

[3] Logitech C920:
https://www.logitech.com/sv-se/products/webcams/c920-pro-hd-webcam.960-001055.html?srsltid=AfmBOoogsFvG5d4a83zEGdmUZA9DIztZ_aK1HbbgQj5xZovL6Bxqe797

[4] Navisys AG3335 GPS:
https://www.navisys.com.tw/productdetail?name=GRM02&class=GPS

[5] HP ZBook Powder 15.6 inch G10 Mobile Workstation PC: https://www.hp.com/us-en/shop/pdp/hp-zbook-power-156-inch-g10-mobile-workstation-pc-wolf-pro-security-edition

- **Power Source**: Power was supplied through direct access to the battery of the test vehicle.
- **Tablet**: An 11-inch iPad Pro (2nd-generation) was used for participants to complete the post-questionnaire.

A.2 Software Components

- **Eye-Tracking Software**: *iMotions Lab*[6] (Version 10.1)—A platform for collecting, processing, and analysing eye-tracking data.
- **Eye-Tracker Settings**: *Smart Eye Pro* (Version 13.0) was used for tracking and measurements as well as controlling the system configuration.
- **Quantitative Data Analysis Platform**: Quantitative data analysis was conducted using *Python* (version 3.13.3) within the *JupyterLab* environment (version 4.4.1).
- **Qualitative Data Analysis and Visualisation:** Qualitative data analysis and visualisations were conducted using *Figma* (version 125.1.5).

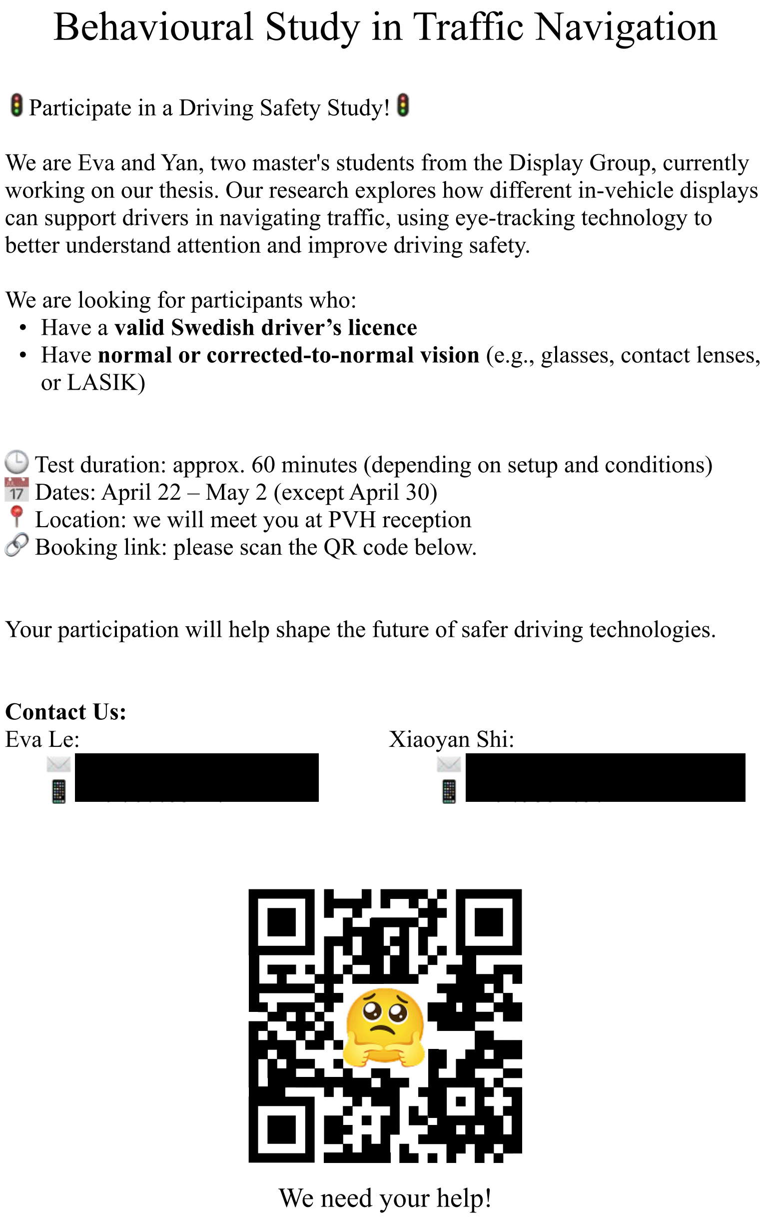
---

[6] iMotions Lab: https://imotions.com/products/imotions-lab/

B

User Study Documents

B.1 Recruitment Advertisement



Behavioural Study in Traffic Navigation

🚦Participate in a Driving Safety Study!🚦

We are Eva and Yan, two master's students from the Display Group, currently working on our thesis. Our research explores how different in-vehicle displays can support drivers in navigating traffic, using eye-tracking technology to better understand attention and improve driving safety.

We are looking for participants who:
- Have a **valid Swedish driver's licence**
- Have **normal or corrected-to-normal vision** (e.g., glasses, contact lenses, or LASIK)

🕒 Test duration: approx. 60 minutes (depending on setup and conditions)
📅 Dates: April 22 – May 2 (except April 30)
📍 Location: we will meet you at PVH reception
🔗 Booking link: please scan the QR code below.

Your participation will help shape the future of safer driving technologies.

Contact Us:

Eva Le:
✉️ 📱 [redacted]

Xiaoyan Shi:
✉️ 📱 [redacted]

We need your help!

B.2 Consent Form

Behavioural Study in Traffic Navigation

As a part of this master's thesis project at Chalmers University of Technology and Volvo Cars, we use eye-tracking technology to study how various in-vehicle displays can help drivers navigate traffic, with the goal of improving driving safety. In this experiment, we are interested to see your natural behaviours while following Google Maps through different displays setups.

During the study, we will record data on your visual attention (e.g., eye movements, head rotation, and pupil dilation) as well as the content displayed on the in-vehicle screens. Three eye-tracking cameras positioned facing toward you will collect numerical eye-tracking metrics. Additionally, an environment camera mounted under the windshield will record traffic conditions, and a scene camera positioned behind your headrest will capture display changes. Please note that due to the nature of the setup, your face, the back of your head, as well as your hands and arms may be recorded during the study. However, your facial data will not be used for analysis, and no facial footage will be published or shared publicly. All recordings will be handled confidentially and in accordance with data protection regulations.

You will also be asked to complete questionnaires and answer interview questions after the drive. There are no right or wrong answers here, we are only interested in your honest opinions. The experiment will last approximately 60 minutes, depending on the setup and test conditions. After the test, please avoid sharing study details (e.g., driving routes, test conditions) with others, as it may affect data validity.

This study is conducted in accordance with the ethical guidelines of Chalmers University of Technology, Volvo Cars, the European Union, and the Helsinki Declaration. You may withdraw from the study at any point: if you no longer wish to take part in the experiment, any data collected up to that point will be removed from the study and deleted. All data collected within the study will be entirely confidential and your data will be referenced only by a participant number. If you would like further information, or would like to know the outcome of the study, please contact the following researchers:

Thesis students: Eva Le (leev@student.chalmers.se), Xiaoyan Shi (shixi@student.chalmers.se)
Supervisors: Ilaria Torre (ilariat@chalmers.se), David Hermann (david.s.hermann@volvocars.com)

If you are willing to take part in the Behavioural Study in Traffic Navigation, please sign your name and give the date below (participant's copy).

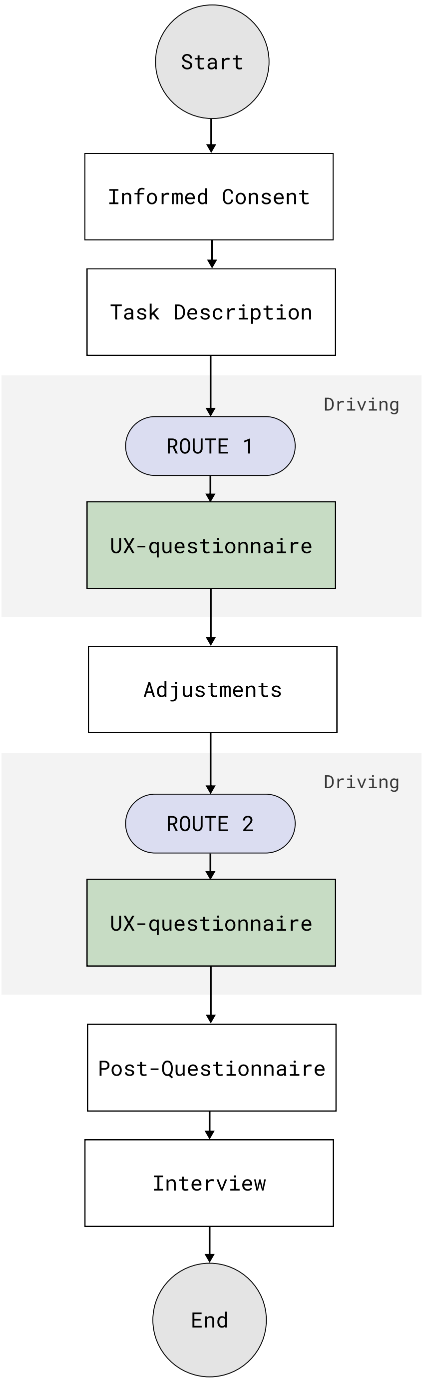
Name:	Date: ....
Signature:	

If you are willing to take part in the Behavioural Study in Traffic Navigation, please sign your name and give the date below (researcher's copy).

Name:	Date: ....
Signature:	

B.3 Task Description

Task Description



On the left side of this paper you see the procedure of the test, starting with an informed consent that you just signed. Now we will go through the task.

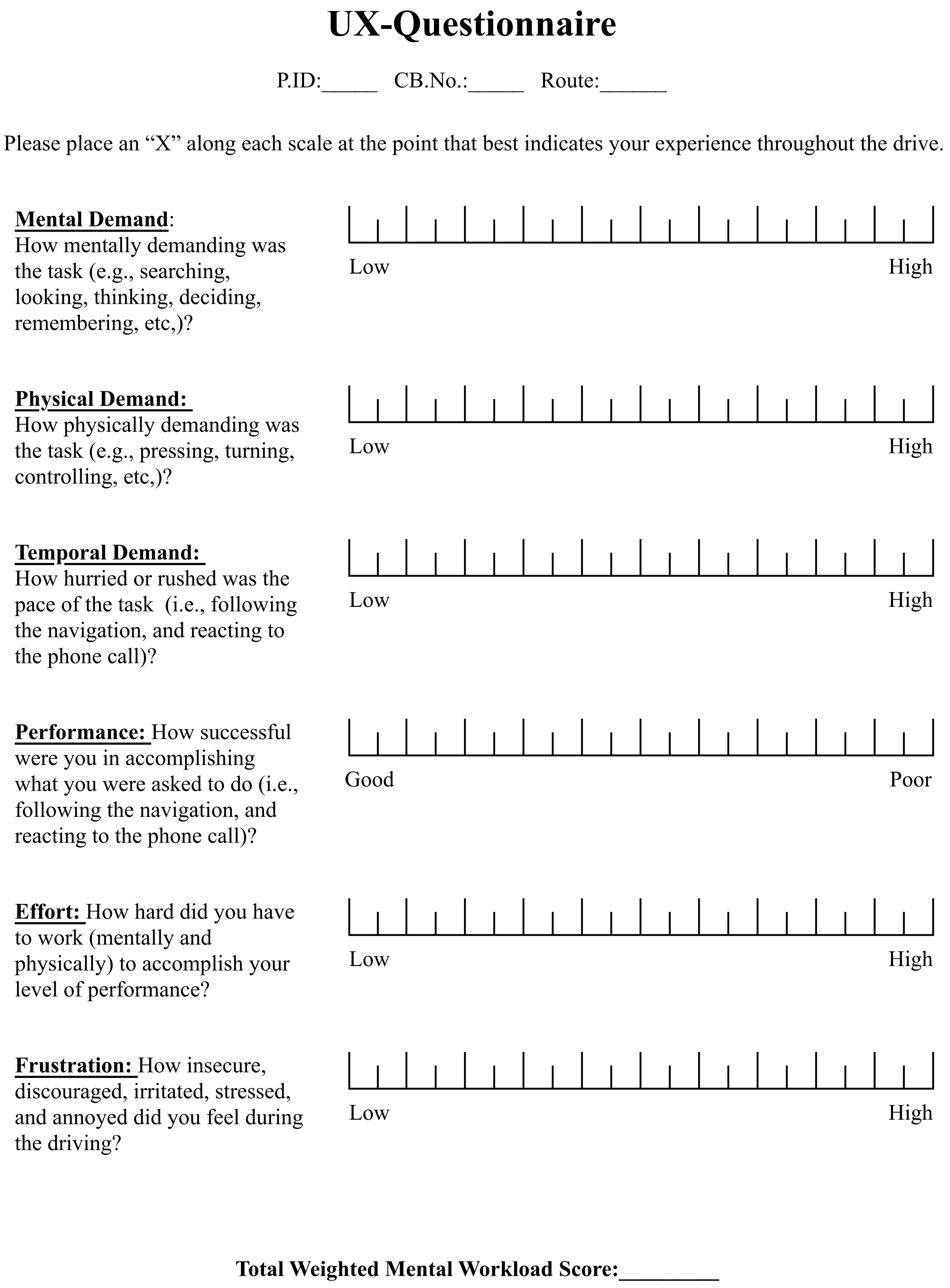
Task introduction
Your task is to follow the navigation that will appear on the displays in the car. You might also need to decline a random phone call while driving. Please do not panic and act as you would normally do. You will drive on two routes in this test, and make a stop in between. You will be in informed when that happens. The instructors (we) will remain silent throughout the drive, but if you are unsure of anything we will assist you.

You are free to make yourself comfortable (e.g. adjusting AC), but please refrain from changing anything on the displays or the cameras. While following and completing the task is important for us, safety is the first priority, no risky actions should be taken in order to follow the task.

Do you have any questions?

Before driving:
1. First, please adjust the seat and mirrors that fit you.
2. For a better recording of the displays, we need to ask your help of adjusting your hair. Feel free to use the hairclips and hairband we prepared.
3. We need your help to calibrate the eye-tracking system. Please look at the points we tell you and try not to blink when we are counting down (3 seconds for each point). Let us know when you are ready, then we will start counting down.

B.4 UX-Questionnaire (NASA-TLX)



UX-Questionnaire

P.ID:_____ CB.No.:_____ Route:______

Please place an "X" along each scale at the point that best indicates your experience throughout the drive.

Mental Demand: How mentally demanding was the task (e.g., searching, looking, thinking, deciding, remembering, etc,)?

Low High

Physical Demand: How physically demanding was the task (e.g., pressing, turning, controlling, etc,)?

Low High

Temporal Demand: How hurried or rushed was the pace of the task (i.e., following the navigation, and reacting to the phone call)?

Low High

Performance: How successful were you in accomplishing what you were asked to do (i.e., following the navigation, and reacting to the phone call)?

Good Poor

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low High

Frustration: How insecure, discouraged, irritated, stressed, and annoyed did you feel during the driving?

Low High

Total Weighted Mental Workload Score:_________

UX-Questionnaire

P.ID:_____ CB.No.:_____ Route:______

For each of the pairs listed below, circle the scale title that represents the more important contributor to workload throughout the drive.

Mental Demand	or	Physical Demand
Mental Demand	or	Temporal Demand
Mental Demand	or	Performance
Mental Demand	or	Effort
Mental Demand	or	Frustration
Physical Demand	or	Temporal Demand
Physical Demand	or	Performance
Physical Demand	or	Effort
Physical Demand	or	Frustration
Temporal Demand	or	Performance
Temporal Demand	or	Frustration
Temporal Demand	or	Effort
Performance	or	Frustration
Performance	or	Effort
Frustration	or	Effort

B.5 Post-Questionnaire

ID

(Filled by Researchers)

1. **Participant ID**
2. **Counterbalance Number**

Basic information

(Filled by Participant)

3. **Age**
    - 18–24
    - 25–34
    - 35–44
    - 45–54
    - 55–64
    - 65+
4. **Gender**
    - Female
    - Male
    - Non-binary
    - Prefer not to say
5. **What is your department at Volvo Cars?**
6. **What is your profession? (e.g. Mechanical Engineer, UX Designer, Product Manager...)**
7. **For how many years have you had your driving license?**
    - less than a year
    - 1–2 years
    - 3–5 years
    - 6–10 years
    - 10+ years
8. **How many hours per week do you drive?**

- 0–5 hours
- 6–10 hours
- 11-20 hours
- 21+ hours

9. **What best describes your current vision?**
    - I have normal (uncorrected) vision
    - I use glasses to correct my vision to normal
    - I use contact lenses to correct my vision to normal
    - I have had vision correction surgery (e.g., LASIK)
    - My vision is not fully corrected to normal
    - Prefer not to say

Familiarity & Preference & Usability

(Please fill in this part after driving two test routes)

10. **Between the two routes, did you experience any significant differences that could affect your driving? (Aside from one having the HUD and the other not, e.g. weather change, traffic density, fatigue...)**
    - Yes
    - No
11. **How often do you use Google Maps for navigation while driving?**
    - Never
    - A few times a year
    - A few times a month
    - A few times a week
    - Every time I drive
12. **Where do you typically check navigation information while driving? (Rank them based on the frequency you use them)(*A demonstration for this question is provided in Figure 1.2)**
    - CSD (Center Stack Display)
    - DIM (Driver Information Module)
    - HUD (Head-Up Display)

13. **Do you have any experience in driving a car with a HUD before this study?**
    - Yes (Go to Q.14)
    - No (Go to Q.16)
14. **How frequently do you use HUD (Head-Up-Display)?**
    - Never
    - A few times a year
    - A few times a month
    - A few times a week
    - Every time I drive
15. **What brand and model of the car did you drive with a HUD (Head-Up-Display)?**
16. **If given the option, would you use a HUD (Head-Up-Display) in your car?**
    - Yes, always
    - Sometimes
    - No, never

B.6 Interview Questions

Interview

Route

10. “Between the two routes, did you experience any significant differences that could affect your driving? (Aside from one having the HUD and the other not, e.g. weather change, traffic density, fatigue...)”
    - If yes: Could you describe the differences you noticed?
    - If no: why? Describe the similarity between two routes
    - Familiarity with the road?

Navigation

11. “How often do you use navigation services while driving?”
    - In our study we use Google Maps, what do you think of the navigation using google maps? (easy to understand?)

12. “Where do you typically check navigation information while driving?”
    - Can you explain the reason for your ranking?
        - Differences between Route 1 and Route 2?
    - What kind of information do you check for each display?

HUD, distraction, safety

16. “If given the option, would you use a HUD in your car?”
    - Why? (amount of information, HUD position, information content)
    - Do you behave any different ways when using a HUD vs not using a HUD?

NASA TLX

- Any comments on the UX-questionnaire? (NASA TLX)

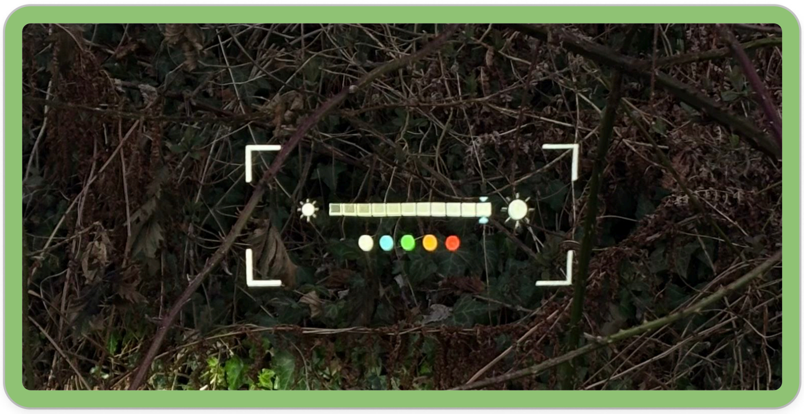
General question:

- Do you have any comments on the study?

B.7 HUD Adjustment

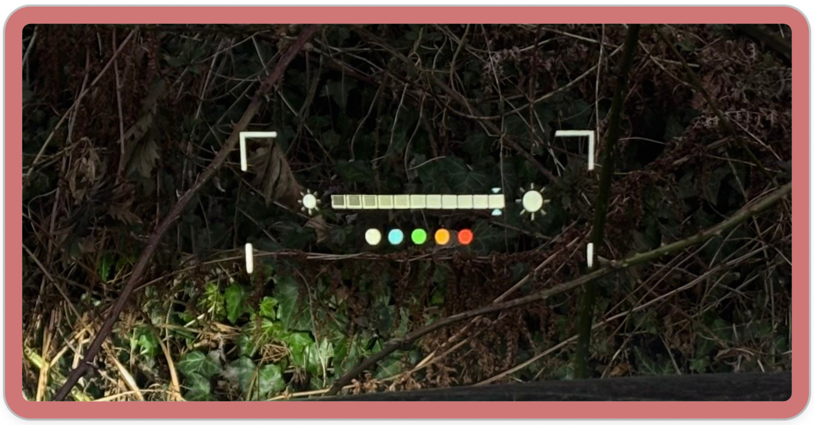
HUD adjustment

Appropriate position



Too high

Too low



B.8 Changes in the HUD on Test Routes

No.	Road Scenarios	Traffic Signs	Manoeuvrers	Other Information
1	T-crossing	Speed limit (50)	turn-left	Distance + Road name
2	Crossroad	Speed limit (50)	turn-left	Distance + Road name
3	Regional road	Speed limit (50)	turn-slight-right	Instruction + Highway No. + Road name
4	Regional road	Speed limit (80)	merge	Road name
5	Regional road	Speed limit (80)	fork-left + LP	Highway No. + Road name
6	Regional road	Speed limit (80)	straight	Road name
7	Roundabout	Speed limit (80)	roundabout-right + LP	Highway No. + Road name
8	Highway	Speed limit (80)	merge	Distance + Highway No. + Road name
9	Highway	-	-	Name + Phone numbers
10	Roundabout	Speed limit (80)	roundabout-left	Distance + Highway No. + Road name
11	Roundabout	Speed limit (80)	roundabout-exit + LP	Distance + Highway No. + Road name
12	Roundabout	Speed limit (70)	roundabout-straight + LP	Distance + Highway No. + Road name
13	Roundabout	Speed limit (70)	roundabout-exit + LP	Distance + Highway No. + Road name
14	Regional road	Speed limit (70)	fork-right + LP	Distance + Road name
15	Roundabout	Speed limit (50)	roundabout-left	Distance + Road name
16	T-crossing	Speed limit (50)	turn-right	Distance + Instruction

Table B.1: HUD information changes on Route 1: Volvo Torslanda PVH to Gamla Flygplatsvägen. LP = Lane Placements.

No.	Road Scenarios	Traffic Signs	Manoeuvrers	Other Information
1	T-crossing	Speed limit (50)	turn-right	Distance + Road name
2	Roundabout	Speed limit (50)	roundabout-left	Distance + Road name
3	Roundabout	Speed limit (50)	roundabout-right	Distance + Road name
4	Roundabout	Speed limit (50)	roundabout-left + LP	Distance + Highway No. + Road name
5	Roundabout	Speed limit (50)	roundabout-exit + LP	Distance + Highway No. + Road name
6	Highway	Speed limit (70)	fork-right + LP	Distance + Instruction
7	Highway	Speed limit (70)	merge + LP	Distance + Highway No. + Road name
8	Highway	Speed limit (80)	fork-right + LP	Distance + Road name
9	Highway	-	-	Name + Phone numbers
10	Roundabout	Speed limit (80)	roundabout-left	Distance + Road name
11	Roundabout	Speed limit (50)	roundabout-exit + LP	Distance + Road name
12	Regional road	Speed limit (80)	merge	Distance + Road name
13	Regional road	Speed limit (80)	fork-right + LP	Distance + Road name
14	Roundabout	Speed limit (80)	roundabout-right + LP	Distance
15	Urban driving	Speed limit (50)	straight	Distance + Road name
16	T-crossing	Speed limit (50)	turn-right	Distance + Instruction

Table B.2: HUD information changes on Route 2: Gamla Flygplatsvägen to Volvo Torslanda PVH. LP = Lane Placements.