



Designing for Perceived Safety in In-Vehicle Touch Interfaces

A Mixed-Methods Study

Master's thesis in Industrial Design Engineering

MARCUS MARTTILA

EDVIN NISBET

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

Master's Thesis

Designing for Perceived Safety in In-Vehicle Touch Interfaces

A Mixed-Methods Study

MARCUS MARTTILA • EDVIN NISBET

Supervisors from Volvo Cars – **Lisa Lindqvist & Maria Redström**
Academic Supervisor & Examiner from Chalmers – **Pontus Wallgren**

Department of Industrial and Material Science – Division of Design & Human Factors

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2025

Master's Thesis 2025 (IMSX30)

Designing for Perceived Safety in In-Vehicle Touch Interfaces

A Mixed-Methods Study

In collaboration with the Car User Research team within Volvo Car Group
(Volvo Cars).

© MARCUS MARTTILA & EDVIN NISBET, 2025

Chalmers University of Technology

SE-412 96 Gothenburg, Sweden

Cover

Volvo Cars' interior image, featuring one of the in-vehicle touch interface concepts developed as part of this thesis. Cover designed by Edvin Nisbet.

Abstract

As digital in-vehicle interfaces become increasingly central to the driving experience, ensuring that they are perceived as safe is essential. This thesis explores how perceived safety can be influenced through the design of human-machine interfaces (HMIs), with a particular focus on touch-based interaction in the centre stack display (CSD) of modern cars.

A mixed-methods approach was applied, starting with a literature review and expert interviews to define perceived safety, identify appropriate evaluation tools, and pinpoint key interface design factors. These insights informed the development of four interface concepts, varying in navigation structure (*flat* and *nested*) and button size. Nineteen participants took part in a user test, evaluating the concepts in a real vehicle on a closed test track. Both objective data (glance behaviour via eye-tracking) and subjective input (questionnaires on perceived safety, usability, and workload, as well as interviews) were collected.

The results formed a set of design and evaluation guidelines grounded in theory and empirical findings. Use flat layouts for quick tasks, particularly when stationary. Consider nested structures when the goal is to reduce the duration of individual glances. Increase button sizes to around 35×35 mm to support confidence and perceived safety.

From an evaluation perspective, use a combination of methods to build a more complete understanding of perceived safety. Interpret quantitative and subjective data alongside qualitative user narratives. While standardised tools such as UMUX-Lite and NASA-TLX offer helpful benchmarks, complement them with task-specific prompts and interviews to capture richer feedback. Furthermore, when comparing similar concepts, use comparative ranking over scalar ratings to reveal clearer preferences. Lastly, test under realistic conditions to ensure findings are grounded in real-world use.

Keywords: *perceived safety, in-vehicle touch interfaces, automotive HMI, eye-tracking, glance behaviour, usability, user experience.*

Acknowledgements

This master's thesis presents a project conducted between January and June 2025. The project corresponds to 30 ECTS credits and was carried out by two students at the Department of Industrial and Materials Science, Chalmers University of Technology, in collaboration with the Car User Research team at Volvo Cars.

We have been fortunate to receive support, feedback, and encouragement from many individuals throughout this project. We would like to extend our sincere thanks to our supervisors at Volvo Cars, Lisa Lindqvist and Maria Redström, for your continuous guidance and trust in our process. Your thoughtful input has been invaluable in shaping the direction of the work.

We are also deeply grateful to Martin Myhrman, Erick Hernández, Jon Sandström and the rest of the prototype team at UX Studio Engineering. Without your help, this project would not have been possible. To the Car User Research team, thank you for your insights and for welcoming us so warmly these past few months. We also appreciate the contributions from colleagues in the UX department and the Safety Centre, who generously shared their time and expertise. Your support throughout the process made a real difference.

We would also like to thank all participants who volunteered for the user test. Your involvement was essential to the success of this project.

Lastly, we want to express our heartfelt thanks to Pontus Wallgren, our supervisor and examiner at Chalmers, for your sharp insights, honest feedback and unwavering support.

Table of Contents

1	Introduction	1
1.1	Background.....	2
1.2	Aim	3
1.3	Delimitations	3
1.4	Thesis Outline	4
2	Theoretical Framework	5
2.1	The Automotive Context	6
2.2	Safety	8
2.3	Human Factors	12
2.4	Interface Design Variables	19
3	Methods & Process.....	27
3.1	Focus & Disclosure	29
3.2	Pre-Study	29
3.3	Research	31
3.4	Design	32
3.5	User Test.....	39
3.6	Synthesising & Design Guidelines	50
4	Results	51
4.1	Quantitative Results	52
4.2	Qualitative Results	67
5	Synthesis & Interpretation.....	76
5.1	Designing Quick Controls	77
5.2	Evaluation Methods	89

6	Design & Evaluation Guidelines	94
6.1	Design Guidelines	95
6.2	Evaluation Guidelines.....	98
7	Discussion	100
7.1	Aim & Objectives Revisited	101
7.2	Methodological Reflections	102
7.3	Ethical Considerations	105
7.4	Recommendations for Future Work.....	106
	References	108
	Appendices	121

1 Introduction

This chapter introduces the background and relevance of addressing perceived safety in in-vehicle touch interfaces. It further outlines the aim, research questions, and delimitations, providing a clear starting point for the thesis.

1.1 Background

In recent years, car interiors have increasingly shifted from physical controls to touch-based centre stack displays (CSDs). These displays have become a common feature in modern vehicles, and the market for them continues to grow (Global Growth Insights, 2024). Touch-based interfaces allow unified access to functions such as climate control, media, and navigation. However, they also increase both visual and cognitive demand, as drivers must divert their attention from the road to interact with on-screen elements (Burnett et al., 2004).

To mitigate distraction, the U.S. National Highway Traffic Safety Administration (NHTSA) recommends that secondary tasks should be performable with glances no longer than 2 seconds, and a total eyes-off-road time (TEORT) not exceeding 12 seconds (NHTSA, 2012). Several studies indicate that many modern vehicles fail to meet these thresholds. Tasks such as destination entry or media selection frequently exceed the recommended glance durations and are associated with elevated cognitive workload (Strayer et al., 2017).

As a result, much of the research on automotive human-machine interfaces (HMIs) has focused on objective distraction metrics, such as glance behaviour, lane keeping, and reaction times, while comparatively little attention has been paid to how drivers subjectively experience safety when interacting with these systems (Krstačić et al., 2024). Consequently, the impact of specific design decisions on perceived safety remains poorly understood.

This topic is of particular relevance to Volvo Cars, a company with a long-standing emphasis on safety. In recent Volvo models, most secondary functions are consolidated into large CSDs, reflecting a broader industry shift toward simplified interiors and screen-based control. Within this context, understanding how interface design affects perceived safety is crucial to maintaining the brand's commitment to driver wellbeing.

Accordingly, there is a growing need to understand how drivers experience safety concerning in-vehicle touch interface design. There is also a need to develop suitable evaluation approaches that can assess this experience.

1.2 Aim

This thesis explores how perceived safety during interaction with touch-based in-vehicle HMIs, specifically the CSD, can be shaped through interface design, and how this experience can be meaningfully evaluated in a driving context. The following research questions are addressed:

- **RQ1:** *How do navigation structure and button size influence the perceived safety of in-vehicle touch interfaces?*
- **RQ2:** *How suitable are existing evaluation methods for assessing and comparing the perceived safety of in-vehicle touch interface concepts?*

1.3 Delimitations

This thesis explores perceived safety within the scope of in-vehicle HMIs, without addressing broader aspects of vehicle safety such as crash avoidance, vehicle durability, or active safety systems. The study focuses specifically on touch-based interaction with the CSD, excluding other input modalities such as auditory and tactile interaction. Aspects related to physical interior design, control layout, and material choices are not considered.

The evaluation is limited to short-term interaction and user perception, rather than long-term behavioural change. While efforts were made to approximate realistic driving conditions, the testing environment was constrained by available resources and does not fully replicate the complexity of real-world scenarios.

1.4 Thesis Outline

The report is divided into seven chapters, each described briefly below:

1. **Introduction** – Provides the background, aim, research questions, and delimitations of the study.
2. **Theoretical Framework** – Introduces key concepts and previous research relevant to perceived safety and automotive interface design.
3. **Methods & Process** – Describes the overall process and the methods used throughout the project.
4. **Results** – Presents the results of the driving-based user test, including both subjective and objective data.
5. **Synthesis & Interpretation** – Interprets the results in relation to the theoretical framework and reflects on their meaning.
6. **Design & Evaluation Guidelines** – Proposes practical recommendations for designing with perceived safety in mind and offers guidance on using the evaluation methods applied.
7. **Discussion** – Reflects on the contributions, limitations, and outcomes of the thesis, and suggests directions for future work.

2 Theoretical Framework

This chapter presents theoretical concepts and prior research relevant to the thesis. It includes key definitions and frameworks related to automotive HMIs, safety, human factors, and interface design. These foundations inform the design of the user test as well as the interpretation of its results.

2.1 The Automotive Context

This section introduces how driving tasks and in-vehicle HMI systems are commonly categorised, establishing a foundation for understanding driver interaction in contemporary vehicles.

2.1.1 Primary and Secondary Driving Tasks

Driving tasks can be broadly divided into primary and secondary categories, based on their relevance to the operation of the vehicle (Angell et al., 2013).

Primary tasks encompass core functions necessary to maintain vehicle control, such as steering, braking, accelerating, and maintaining visibility. In contrast, secondary tasks involve interactions that are not essential for vehicle control but may contribute to comfort, convenience, or access to information. Examples include adjusting climate settings, entering navigation destinations, or handling communication features.

2.1.2 Primary and Secondary HMI Components

In-vehicle HMIs are similarly categorised based on their function and relevance to driving (Gáspár et al., 2014). Both primary and secondary HMI components rely on multimodal interaction, including visual, auditory, and tactile inputs and outputs, to facilitate safe and efficient vehicle use.

Primary HMI components pertain to functions that are critical for controlling the vehicle. Inputs in this category include the steering wheel, pedals, gear selector, and parking brake. Outputs typically consist of the driver information module (DIM), positioned behind the steering wheel, which presents key driving information such as speed, warnings, and engine status (see Figure 2.1). This category may also incorporate auditory alerts and voice control, allowing for hands-free engagement with certain systems.



Figure 2.1: DIM and CSD.

Secondary HMI components support the non-critical features such as climate regulation and navigation. These are primarily accessed through the centre stack display (CSD), a touchscreen interface located centrally in the dashboard. Serving both as an input and output device, the CSD enables users to interact with system menus and receive visual feedback.

2.1.3 Quick Controls

In modern Volvo vehicles, such as the EX90, the term Quick Controls refers to a cluster of frequently accessed functions presented in a consolidated, easily accessible interface view, as shown in Figure 2.2 (Volvo Cars, 2024). These controls typically include adjustments related to lighting, mirrors, and driver support features. Accessed in the CSD, the Quick Controls menu is designed to reduce interaction time by minimising the need for deep menu navigation and exposing key functions at a glance.



Figure 2.2: Quick Controls menu in the CSD.

2.2 Safety

This section examines driving safety by distinguishing between objective safety, based on measurable indicators of driver distraction, and perceived safety, which reflects the driver's subjective sense of security when interacting with an interface.

2.2.1 Objective Safety

When examining driving safety in the context of secondary tasks, driver distraction is a critical factor, influencing crash risk and remains one of the most common causes of traffic accidents (Ranney et al., 2001). In 2023, NHTSA reported 3,275 fatalities attributed to distracted driving (NHTSA, n.d.). Driving distractions are typically classified into three categories: cognitive (mental diversion from driving), visual (eyes off the road), and manual (hands off the wheel) (Li et al., 2020; Ranney et al., 2001).

Many common secondary tasks, such as text messaging, combine all three types of distraction, significantly impairing driving performance.

Research has demonstrated a clear link between engagement in secondary tasks and impaired lane-keeping ability. For example, studies by Horberry et al. (2006) and Jeong and Liu (2019) show that distractions can significantly degrade lateral control, increasing the likelihood of lane departures and subsequent collisions. As lane-keeping is a fundamental aspect of safe driving, its degradation is an important consideration when assessing driving performance.

Among the various forms of distraction, visual inattention is among the most extensively studied and has a direct association with increased crash risk. Studies have shown that glances away from the forward roadway are strong predictors of near-crashes and crashes, particularly rear-end collisions (Klauer et al., 2014; Victor, 2014). Notably, a single glance away from the road exceeding two seconds has been found to nearly double the risk of a crash (Klauer et al., 2006). Real-world data further indicate that safety risks rise once this two-second threshold is exceeded, reinforcing its value as a simple, actionable metric (Li et al., 2020).

2.2.2 NHTSA Guidelines as an Objective Standard

Because of the documented risks associated with driver distraction, NHTSA (National Highway Traffic Safety Administration) introduced comprehensive guidelines in 2012 for the design of in-vehicle electronic systems (NHTSA, 2012). These guidelines, informed by research on driver eye-glance behaviour, aim to ensure that secondary tasks can be completed with minimal visual distraction from the primary driving task.

The NHTSA framework specifies three criteria related to glance duration for evaluating whether a system poses a distraction risk during testing:

- 1. Individual glance durations:** For at least 21 of the 24 test participants, no more than 15% (rounded up) of the total number of eye glances away from the forward road scene should have durations of greater than 2.0 seconds while performing the secondary task

- 2. Mean glance duration:** For at least 21 of the 24 test participants, the mean duration of all eye glances away from the forward road scene should be less than 2.0 seconds while performing the secondary task.
- 3. Total glance time:** For at least 21 of the 24 test participants, the sum of the durations of each participant's eye glances away from the forward road scene should be less than, or equal to, 12.0 seconds while performing the secondary task one time.

These guidelines are supported by independent research. For instance, Bärghman et al. (2015) demonstrated that eliminating glances longer than two seconds substantially reduced simulated crash and injury risk, highlighting individual glance duration as a key factor. Similarly, Simons-Morton et al. (2014) found that the longest single glance provides a more reliable predictor of crash risk than total glance time, reinforcing the relevance of NHTSA's two-second threshold in safety assessments. Collectively, these evidence-based criteria offer manufacturers practical benchmarks for designing safer in-vehicle interfaces that address established crash risk factors.

2.2.3 Perceived Safety

In the context of driving, perceived safety is a driver's subjective feeling of security. Researchers have defined it in several ways. For instance, some describe it as a mental comparison between perceived risk and an acceptable level of personal risk, which is influenced by one's sense of control, the overall driving experience, and interior elements that offer a "protective effect" (Fischer et al., 2022). Others have explored it as an emotional aspect of the driver experience linked to enjoyment, noting that external factors such as traffic conditions can affect this relationship (Park & Park, 2021). Xu et al. (2018) define perceived safety as "a climate in which drivers and passengers can feel relaxed, safe and comfortable, while driving" ("Perceived Safety" section, para. 2). In contrast, Osswald et al. (2012) describe it as "the degree to which an individual believes that using a system will affect his or her well-being" (p. 55).

While this broad context is important, a more focused framework is needed to evaluate how specific HMI design elements affect the perceived safety of an interface.

Therefore, in this thesis, we draw on the work of Osswald et al. (2012), who frame perceived safety as a key factor in their Car Technology Acceptance Model (CTAM), which addresses user interaction with in-vehicle systems. Their model is particularly relevant because it directly addresses user interaction with in-vehicle systems. Based on our interpretation of their framework, perceived safety can be viewed as consisting of two interrelated dimensions:

- **Safety perception:** The user's belief about whether using the system is dangerous or safe.
- **Distraction perception:** The user's assessment of how much the system diverts attention from the primary driving task.

We adopt this interpretation as a useful lens for analysing HMI design, acknowledging that in the context of in-vehicle HMI, perceived safety encompasses both the avoidance of accidents and the ability to manage attention effectively between driving and secondary tasks.

2.2.4 A Disconnect Between Objective and Perceived Safety

While objective safety can be assessed through measurable indicators such as accident rates, lane deviation, or standardised glance duration thresholds (e.g., NHTSA's guidelines), perceived safety belongs to the subjective domain of user experience. A system may satisfy all objective safety requirements yet still feel unsafe to users. Conversely, users may feel comfortable with interactions that, according to objective measures, introduce risk.

This gap between perceived and objective safety is especially noticeable in the context of different types of driving distractions. One illustrative example relates to the physical placement of controls. Dukic et al. (2007) investigated how the positioning of in-vehicle controls influences both driver performance and perceptions of safety. Their study found that a control placed near the gear stick, although outside the typical line of sight, led to a relatively short TEORT and steering wheel deviation, suggesting good objective performance. Nevertheless, drivers rated this location as

one of the least safe, highlighting the disconnect between how safe an interaction feels and how it actually affects driving performance.

Further research highlights similar discrepancies between drivers' perceptions of secondary tasks and the tasks' objectively measured impact on driving performance. For example, research by Li et al. (2020) shows that drivers often underestimate the risks associated with certain visual distractions, such as glancing at the rear-view mirror. Although these actions feel harmless, they can negatively affect lane-keeping performance. On the other hand, some minor cognitive distractions, which are also viewed as low-risk, tend to have little measurable impact on lateral vehicle control, suggesting that the relationship between perceived and actual safety varies depending on the type of distraction, or even on the age of the user (Li et al., 2020; Dukic et al., 2006).

2.3 Human Factors

This section provides an overview of key human factor considerations that influence how drivers interact with in-vehicle touchscreens. It highlights how cognitive, perceptual, and ergonomic aspects shape attention, system usability, and, ultimately, drivers' perceived safety within the driving context.

2.3.1 Mental Models

Mental models are the internal representations people form to understand and predict how systems work. Johnson-Laird (1980) describes these models as simplified mental structures that reflect how users believe a system is organised. They help users anticipate outcomes, make decisions, and solve problems by mentally simulating the system, rather than relying solely on formal instructions or reasoning.

In HMI design, mental models are essential since they influence how users interpret and engage with an interface. When a system behaves in ways that match the user's expectations, interaction tends to feel intuitive and efficient. In contrast, if there is a mismatch between how a system operates and how users expect it to work, confusion, inefficiency, or errors may occur.

2.3.2 Workload

Mental workload refers to the amount of attentional resources an individual must allocate to meet the demands of a task (Young & Stanton, 2004). In the driving context, mental workload emerges from the interaction between the driver's capabilities (e.g., skill level, fatigue), the demands of the driving task, and environmental factors (Hart & Staveland, 1988). When a driver engages in a secondary task, such as operating an in-vehicle touchscreen, it adds to the overall workload. If the combined demands exceed the driver's available attentional capacity, cognitive overload occurs, leading to impaired performance and reduced safety (Wickens, 2008).

Multiple resource theory offers a useful framework for understanding the interference caused by dual-task situations, such as driving while interacting with a touchscreen (Wickens, 2002; Wickens, 2008). According to the multiple resource theory, cognitive resources are divided into distinct, limited pools rather than drawn from a single, undifferentiated supply. These resources are categorised along dimensions such as input modality (e.g., visual vs. auditory) and stages of processing (e.g., perceptual/cognitive vs. response). Performance declines are most pronounced when two tasks compete for the same type of resources at the same time.

The conflict between driving and in-vehicle HMI interaction exemplifies this form of resource competition. Driving primarily relies on visual, cognitive, and manual resources. Using a touchscreen directly competes for these same resource pools by diverting visual attention from the road, splitting cognitive resources required for situational awareness, and introducing manual interference with steering control. This overlap is a key factor contributing to distraction and performance deterioration, as demonstrated in numerous studies (Horrey & Wickens, 2006; Vollrath, 2004).

Understanding this resource conflict has important implications for interface design. To improve safety, in-vehicle HMI should be designed to reduce competition for the same resources by distributing demands across different channels.

2.3.3 Situation Awareness

Situation awareness refers to an individual's understanding of their environment in relation to time and task. Endsley (2000) describes it as the ability to notice important elements in the surroundings, understand their significance, and use that understanding to foresee how the situation may evolve.

This construct is especially relevant in dynamic contexts such as driving, aviation, or control systems, where operators must make decisions under changing conditions. According to Endsley, situation awareness unfolds across three levels: perceiving relevant information, making sense of it, and anticipating future states. These stages collectively support timely and accurate decision-making.

Rather than being a passive state, situation awareness involves active information gathering and integration, shaped by goals, attention, experience, and mental models. It is considered a key factor in expert performance, but it can break down due to information overload, poor interface design, or misaligned expectations.

2.3.4 Usability

In the evaluation of HMIs, usability is a central concept. Usability refers to the extent to which a user interface is easy and efficient to use and contributes to a positive user experience. It extends beyond basic functionality to include the overall quality of interaction with a system, product, or service. A thorough understanding of usability incorporates several key components that together determine whether a design meets the needs of its intended users.

According to ISO 9241-11:2018 (International Organization for Standardization [ISO], 2018), usability is "the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." This definition highlights that usability is not an inherent characteristic of a product but depends on the users, their goals, and the context in which the system is used.

Jakob Nielsen (2012) further describes usability through five key quality components:

- **Learnability:** How easily new users can accomplish basic tasks the first time they encounter the design. A system with high learnability enables users to get started with minimal instruction.
- **Efficiency:** Once users have learned the system, efficiency describes how quickly and effectively they can perform tasks. An efficient system supports productivity with minimal effort.
- **Memorability:** This refers to how easily users can return to the system after a period of not using it and re-establish their proficiency. A memorable design allows users to resume tasks without having to relearn the interface.
- **Errors:** This component concerns how frequently users make errors, how severe those errors are, and how easily they can recover from them. A well-designed system helps users avoid errors and provides clear guidance for recovery when errors do occur. This directly contributes to the overall effectiveness of the system.
- **Satisfaction:** The degree to which users find the system pleasant to use. Satisfaction reflects users' overall impressions of the interface, including their physical, cognitive, and emotional responses.

Usability evaluations should account for these factors and consider the specific context of use, as recommended by ISO 9241. It is important to recognise that usability is context-dependent. An interface that is highly usable for one group of users or in one environment may not be suitable for others.

Usability is also closely linked to utility. Utility refers to the functionality of a system or product, specifically whether it provides the features required to meet users' needs (Nielsen, 2012). A system may be highly usable but lack the necessary functions, making it ineffective for its intended purpose. Conversely, a system may offer a wide range of useful features, but if poorly designed, users may struggle to access or apply them effectively. Both usability and utility are essential to ensure that a system is not only easy to use but also fulfils its intended purpose.

2.3.5 Control and Predictability

Control and predictability are two closely related qualities that are essential for safe interaction with a touch-screen HMI. The driver must remain in control of every action and be able to anticipate how the interface will respond. Nielsen's usability heuristics reflect these same principles through the concepts of user control and freedom, consistency and standards, and visibility of system status. Together, these principles require that users can initiate, cancel, and understand each step of a dialogue without unexpected outcomes or hidden states (Nielsen, 1994).

These general principles are also embedded in the standards governing in-vehicle HMIs. ISO 9241-110:2020 (ISO, 2020) identifies controllability and conformity with user expectations as fundamental interaction principles. Similarly, ISO 15005:2017 (ISO, 2017) specifies that driver dialogues must remain predictable whenever the vehicle is in motion.

Empirical research in other safety-critical domains shows why these principles matter: when system behaviour is transparent and consistent, users form trust and rely on the automation appropriately; when behaviour is erratic, trust collapses and attention is diverted. Lee and See (2004) make this causal chain explicit in their review of automation, describing predictability as a precondition for "appropriate reliance" and, ultimately, safe performance.

2.3.6 Visual Attention When Driving

Visual attention is a fundamental component of the driving task, as it is the main way drivers take in and respond to their surroundings. Given that more than 80% of all traffic information is acquired through the visual channel, visual attention is critical for guiding driver awareness and enabling the selective concentration needed to navigate safely (Y. Liu et al., 2022).

Drivers are not only passive recipients of visual stimuli, but they actively manage their attention while driving. This regulation is a dynamic process where drivers strategically make moment-to-moment decisions about where to focus, based on the evolving context and what they are trying to achieve (J. D. Lee, 2014; Z. Liu et al.,

2021). However, the introduction of in-vehicle HMIs presents a challenge to this cognitive process. Interacting with an in-vehicle system requires that drivers divide their visual resources between the primary task of driving and a secondary task. This process is characterised by a series of glances toward the in-vehicle display, broken up by glances back at the road to maintain awareness of the surroundings (Kujala & Salvucci, 2015).

The key concern for HMI design is ensuring that design choices support safe attention switching during driving. Kujala and Salvucci (2015) propose a cognitive model to explain how drivers manage this division of attention, built on several key assumptions about visual search and time management. Central to their model is the idea that drivers are aware of the passage of time during a glance and attempt to adhere to a self-imposed time limit before returning their gaze to the road. This internal time limit is not static but is adjusted based on the perceived stability of the driving situation. If the vehicle remains stable after a glance, the driver may extend the time limit for the subsequent glance, and if the vehicle's control has degraded, the limit is shortened (Kujala & Salvucci, 2015). This active regulation in response to HMI-induced distractions demonstrates that visual attention in driving is a deliberately managed cognitive resource, allocated to meet the ongoing and future demands of the task.

2.3.7 Subtask Boundaries & Interruptibility

The design of secondary tasks significantly influences driver safety, particularly through what Salvucci and Kujala (2016) describe as task structure, meaning how a task is divided into smaller subtasks. These subtasks are separated by subtask boundaries, which serve as natural pauses or completion points within the larger task. For example, entering a segment of a phone number or typing a single word constitutes a subtask, with its completion marking a subtask boundary (Janssen et al., 2012).

In the complex multitasking environment of driving, individuals consistently manage their attention by switching between the primary driving task and the secondary HMI task at these subtask boundaries (J. Y. Lee & Lee, 2019). In practice, a driver

interacting with an in-vehicle system often completes one subcomponent of the secondary task and then pauses at a subtask boundary to redirect their visual attention to the road. This might involve typing a word before glancing up or selecting a menu item, followed by briefly monitoring the driving environment. This momentary disengagement helps drivers maintain awareness of the road before resuming the secondary task.

The structure of secondary tasks and their subtask boundaries directly affects drivers' glance strategies and, as a result, road safety. Research shows that as the duration of a secondary visual search task increases, glance durations tend to increase as well (Kujala & Salvucci, 2015). This becomes particularly problematic when a secondary task is not designed for easy interruption. If a subtask is inherently lengthy, lacks clear breakpoints, or imposes a cost for being abandoned mid-execution, such as lost input, the need to restart, or increased mental effort to reorient, drivers may be more likely to prolong their visual engagement with the HMI. Rather than distributing their attention through frequent, short glances, they may extend a single glance or a sequence of glances to complete the subtask in one attempt.

As declared, prolonged glances away from the road, especially those longer than two seconds, are widely recognised as dangerous (Bärgman et al., 2015). Consequently, in-vehicle HMI tasks that lack easily interruptible, frequent, and logically structured subtask boundaries may unintentionally encourage more hazardous glance strategies involving these extended off-road glances.

2.3.8 Physical Ergonomics of Touchscreen Interactions

The physical ergonomics of HMIs, particularly large touchscreens, is an important factor affecting driver comfort and safety. The placement of interactive elements dictates the driver's physical movements, and poorly designed layouts can lead to ergonomic challenges such as excessive hand travel, awkward postures, and body shifts that compromise driving performance.

The spatial positioning of on-screen targets directly affects the driver's reach and posture. As the distance to a target increases, so does the physical effort required.

Research by Liu et al. (2022) found that targets located on the far-right areas of a large screen resulted in less accurate inputs and required drivers to extend their arm and torso, leading to postural adjustments. These adjustments are not without consequence. For example, Ma et al. (2020) found that significant reaching movements can require the driver to move their entire body, which in turn leads to reduced vehicle control, as shown by an increase in lane departure standard deviation. Reaching for distant targets diverts both cognitive attention and physical stability away from the primary driving task.

An established principle in ergonomic HMI design is to minimise the travel distance required for the hand and fingers. Liu et al. (2022) reinforce this by distinguishing between "comfortable" and "maximum" interaction ranges. Their study showed that tasks requiring interaction at the maximum range, such as sliding to the outer edges of the screen, had a noticeably negative impact on driving performance, demonstrated by an increased standard deviation in the steering wheel angle. These findings indicate that forcing interaction with the edges of a large display introduces interference that can compromise driving safety.

In contrast, placing frequently used controls within a comfortable reach zone helps minimise these negative effects. Liu et al. (2022) also identified an "easy sliding area" in the central part of the touchscreen, where most users could perform tasks without significant postural strain. Locating high-frequency interactions within these ergonomically optimal zones can reduce both physical and cognitive workload, shorten the time it takes to complete tasks, and improve overall driving safety and comfort.

2.4 Interface Design Variables

This section outlines key interface design variables that shape driver interaction with in-vehicle touchscreens. It also examines how aesthetic and psychological elements, such as visual appeal, influence users' subjective experiences and overall satisfaction with the system.

2.4.1 Interface Structure

A key consideration when designing menus is the trade-off between breadth and depth. Designers must decide whether to present many options on a single screen (favouring breadth) or fewer options per screen, requiring users to navigate through more layers (favouring depth) (Burnett et al., 2013). Early research by Landauer and Nachbar (1985) generally suggests that broader menus enable faster target acquisition, as users can often process multiple items on one screen more quickly than navigating through successive screens.

However, the effectiveness of breadth or depth depends heavily on how menu items are organised. Burnett et al. (2013) distinguish between structured menus, where items are arranged in a predictable order (e.g., alphabetically), and unstructured menus, where items are presented without a discernible organisational logic. This distinction is critical, as a user's ability to anticipate a target's location strongly influences performance (Cockburn & Gutwin, 2009). Factors such as alphabetical ordering or user familiarity can therefore moderate the general advantage of breadth. The logical arrangement enables users to anticipate item locations and apply systematic search strategies. This advantage makes structured, broad menus particularly effective for complex interface navigation. Supporting this, Cao et al. (2025) observed significantly improved task performance with broad, grouped menus compared to deeper, hierarchical ones. However, the universal advantage of breadth has been nuanced by research showing its dependence on task type. Louveton et al. (2024), for instance, found that while maximum breadth was indeed optimal for systematic reading tasks, visual demand was lowest with intermediate breadth and depth for more cognitively complex search and memory tasks.

The situation differs notably when menus are unstructured. In these cases, users often adopt a sequential visual scan-and-evaluate strategy, examining each option individually (Landauer & Nachbar, 1985). As a result, broad unstructured menus can lead to a near-linear increase in response time, making them less efficient than narrower, deeper alternatives or at least diminishing the typical advantages of breadth. By comparison, structured layouts enhance the benefits of breadth by supporting user anticipation and accelerating target acquisition.

2.4.2 Progressive Disclosure

Progressive disclosure is an interaction design technique used to manage complexity and reduce choice overload by initially presenting only the most essential information and actions to the user, while placing more advanced or less frequently used options in a secondary view (Nielsen, 2006). A common way to implement this in HMIs is through modal windows, which appear over the main interface to provide additional or advanced functions when requested by the user (Interaction Design Foundation, 2016).

The key advantage of this approach is that it balances users' needs for both simplicity and functionality. By hiding less frequently used settings, the interface becomes easier to learn for novices and reduces the likelihood of user errors. This simplification helps users focus on core features, lowering cognitive workload and supporting more efficient decision-making (Nielsen, 2006). Expert users also benefit, as a decluttered interface allows them to avoid unnecessary scanning for rarely used options. Research by Ding et al. (2020) supports these benefits, showing that progressive disclosure can effectively highlight priority options and significantly reduce the choice overload often encountered in dense interfaces, such as smartphone home screens.

Despite these benefits, effectively applying progressive disclosure presents design challenges. The most critical is determining which features to display initially and which to defer. If too little is shown, users are forced into extra interactions, and if too much is revealed, the intended simplicity is lost (Nielsen, 2006). Without careful design, progressive disclosure can inadvertently introduce new navigational burdens, turning a tool intended to reduce complexity into a source of confusion.

2.4.3 Touch Target Size

During the performance of secondary tasks while driving, user precision is notably reduced. As a result, touch button size becomes a critical design parameter that significantly affects both usability and user performance. This principle is grounded in Fitts's Law (Fitts, 1954), which states that the time required to accurately select a target is inversely related to its width, meaning that smaller targets demand more time

and precision. Consequently, substantial research has examined optimal button sizes for in-vehicle touchscreens, resulting in various recommendations for minimum dimensions.

Several studies have investigated this question using different experimental setups. For example, Kim et al. (2014) tested five square button sizes ranging from 7.5 mm to 27.5 mm on a 17-inch monitor for a five-key input task under simulated driving conditions. They recommended a button size of at least 17.5 mm, noting usability improvements up to that dimension across different speeds. Similarly, Da Tao et al. (2018) explored a comparable size range for digit and letter input tasks and also found that medium to large square buttons (17.5 mm and above) improved user performance. Chen et al. (2013) observed only limited improvements in usability with button sizes exceeding 20 mm for the general population. However, for users with disabilities, performance continued to improve up to their tested maximum size of 30 mm, suggesting that individual factors such as motor limitations influence optimal button size.

Taking a different approach, Feng et al. (2018) examined three larger button sizes (14 mm, 24 mm, and 33 mm) combined with varying button counts (4, 8, and 15) on an 8-inch smartphone for a search-and-press task, performed while driving. Their results suggested that buttons larger than 24 mm offer usability advantages. Suh and Ferris (2019) compared two square button sizes (15 mm and 25 mm) for a five-key input task on both a smartphone and a physical keypad. While a 25 mm button generally yielded better performance, this advantage diminished when auditory feedback was introduced.

When button size falls below a certain threshold, usability metrics such as task completion time (TCT), accuracy, and subjective user satisfaction tend to decline significantly. This decline is often attributed to the absence of tactile feedback on touchscreens and the visual occlusion of the target by the user's finger (Kim et al., 2014). Research consistently shows that larger buttons are associated with improved performance (Kim et al., 2013; Feng et al., 2017). However, practical constraints such as limited screen space often require a trade-off.

2.4.4 Text Size

Research by Crundall (2016) suggests that while variations in text size height (ranging from 4 mm to 9 mm) may not affect the TEORT, they do influence the pattern of glances. Larger text sizes tend to prompt more frequent but shorter glances, whereas smaller text sizes results in fewer, longer glances. This indicates that even if the overall time spent looking away from the road remains similar, the cognitive load and nature of visual distraction can differ significantly depending on text legibility.

2.4.5 Visual Complexity

Perceived visual complexity is an important concept for understanding user interaction with visual displays, especially in information-rich environments such as in-vehicle interfaces. S. C. Lee et al. (2016) define perceived visual complexity as an operator's subjective experience of complexity when viewing such displays. Their research identifies three fundamental, interacting factors that determine this perception: the quantity of visual elements, the variety in their characteristics (such as shape and colour), and the relation, or organisational structure, between them. An increase in the number of objects or diversity in their features directly increases perceived complexity, as more elements and distinct features require greater cognitive effort to process. Conversely, strong visual hierarchies, clear groupings, and logical layouts can reduce perceived visual complexity, making even displays with many elements feel less complex and easier to navigate. This aligns with established design principles, including Gestalt laws (Interaction Design Foundation, 2016). and Nielsen's 10 usability heuristics (Nielsen, 2024). Supporting this framework, Yoon et al. (2015) found that objective measures, such as the quantity of components and icons in automotive instrument clusters, positively correlate with perceived visual complexity.

The significance of perceived visual complexity extends directly to its effects on visual search patterns and glance behaviour (Guo et al., 2022). The study by S. C. Lee et al. (2016) demonstrated a clear correlation between higher perceived visual complexity and detrimental impacts on how users scan and interact with displays. Specifically, increased perceived visual complexity led to longer time-on-task and a greater

number of fixations. This suggests that users require more time and make more distinct stops to process complex interfaces. These impacts can have significant implications, particularly in contexts such as driving, where prolonged attention diverted from the primary task can compromise safety. User sensitivity to visual complexity varies and is influenced by factors such as age. Research shows that older adults are more susceptible to performance degradation from complex interfaces. This is reflected in a greater decline in their primary task performance and a higher perceived cognitive workload compared to younger individuals (S. C. Lee et al., 2019). Understanding these components and consequences of perceived visual complexity is therefore essential for designing interfaces that support efficient and safe visual interaction.

2.4.6 Number of Display Items

Research shows that both the quantity of on-screen elements, such as buttons or menu items, and text size significantly affect driver glance behaviour. Studies by Kujala and Saariluoma (2011) and Kujala and Salvucci (2015) found that an increasing number of display items negatively impacts driver attention. Specifically, more items are associated with longer maximum glance durations, a higher frequency of off-road glances, greater TEORT, and a larger proportion of glances exceeding two seconds. Feng et al. (2018) support these findings, reporting a clear link between the number of buttons and the likelihood of tasks requiring multiple glances, along with an increase in prolonged glances.

For example, tasks involving a 2x2 button grid were often completed with a single glance. In contrast, a 3x5 grid resulted in a notable increase in tasks requiring two, three, or more glances. The research consistently suggests that designs with fewer items, such as a four-item layout (Feng et al., 2018) or a six-item list menu (Kujala & Saariluoma, 2011), are more likely to meet acceptable thresholds for minimising extended glances. This aligns with NHTSA's (2012) over-2-second glance criterion.

A limitation to consider is that these studies primarily reflect the performance of novice users during their first interactions with an unfamiliar interface. The experimental conditions did not allow participants to learn and memorise button

locations. As a result, the findings represent tasks dependent on active visual search rather than learned spatial memory. Therefore, the reported glance durations may be higher than those of experienced users who have developed familiarity with the interface layout over time.

2.4.7 Glanceability

Glanceability describes how quickly and easily a person can take in information from a display while their attention is mainly focused elsewhere, such as during driving. Blascheck et al. (2021) describe a glance as lasting anywhere from a few milliseconds to several seconds, depending on how much information needs to be processed.

According to their work, three factors influence how glanceable a display is. First, the information must be easy to see and access, without requiring large movements or taking attention too far from the main task. Second, the information should be simple and familiar, so it can be understood quickly. Third, only the information that is relevant at that moment should be shown.

As Blascheck et al. (2021) note, displays that are poorly placed or overly complex tend to cause longer glances. It can be assumed that this increases the risk of missing important events on the road. In contrast, a highly glanceable display allows drivers to quickly check information and return their focus to the road, supporting both safety and trust in the system.

2.4.8 Influence of Aesthetics

Visual design plays an important role in how users perceive the functionality of an interface. Research by Lavie et al. (2011) found that aesthetically pleasing navigation systems, especially those with clean and simple designs, were consistently rated as more usable, even though they did not improve performance. This reflects a "halo effect," where positive impressions of one feature can unconsciously influence how users judge other, unrelated qualities of the same system (Thorndike, 1920).

The influence of aesthetics is more about shaping how people feel about the system than improving measurable performance. A good-looking interface may not make

tasks faster or reduce errors, but it improves first impressions and the overall user experience. Frison et al. (2019) showed a similar effect with automated vehicle interfaces. Their study found that the visual appeal of individual design elements affected how users perceived the whole system, including their level of trust in it.

3 Methods & Process

This chapter describes the overall approach taken to explore perceived safety in in-vehicle touch interfaces. It outlines the methods used and how they were applied across the different phases of the project.

01

Pre-Study

- › Reviewed foundational literature
 - › Defined perceived safety
 - › Selected measurement tool
 - › Conducted expert interviews
- › Narrowed focus to Quick Controls

02

Research

- › Explored HMI–safety connections
- › Reviewed interface design theory
- › Conducted industry benchmarking
 - › Selected key design variables

03

Design

- › Created 4 testable HMI concepts
- › Varied navigation structure & button size
 - › Prototyped for in-vehicle testing

04

User Test

- › Conducted tests on Volvo’s closed track
 - › 19 users, 6 tasks, 4 concepts
 - › Eye-tracking & performance data
- › Perceived safety, usability & workload ratings
 - › Analysed results

05

Synthesis

- › Mapped findings back to theory
- › Derived practical design & evaluation guidelines

Figure 3.1: Process overview

3.1 Focus & Disclosure

The primary aim of this project was not to produce a single, polished design solution, but to investigate how variations in interface design influence perceived safety. All preceding phases (see Figure 3.1) of the project were therefore structured to support the execution of a controlled user test, which constituted the core focus of the work.

To enable realistic interaction in a driving environment, interactive prototypes were developed in collaboration with Volvo Cars' prototype team. As the test vehicle included future interior elements and systems, some details cannot be disclosed. Instead, to maintain context while preserving confidentiality, the interface concepts shown in this report are visualised using a design language found in currently available vehicles.

3.2 Pre-Study

The initial stages of the project were characterised by a broad and partially undefined brief. Although the overarching focus on perceived safety in automotive HMIs had been established, its definition, scope, and practical application remained unclear. To address this, a pre-study was conducted, combining a structured literature review with semi-structured expert interviews to establish a foundation for the work ahead.

3.2.1 Literature Review on Perceived Safety

The purpose of the initial literature review was to examine how perceived safety has been defined and measured in prior research, particularly in the context of automotive interaction.

Literature reviews serve to establish a baseline of existing knowledge by identifying key concepts, theoretical models, and methodological approaches. This process typically includes scoping the domain, generating keywords, retrieving sources from academic databases, and synthesising findings through thematic analysis or categorisation (Booth et al., 2016).

Searches were conducted using Google Scholar, Web of Science, and Scopus, focusing on terms such as *perceived safety*, *trust in automotive systems*, *driver workload*, and other related concepts. Boolean operators were used to improve precision and relevance.

The reviewed literature included peer-reviewed journal articles, conference proceedings, and research reports addressing both general perceptions of safety and domain-specific factors such as in-vehicle HMI-related stressors and cognitive load. These insights informed the selection of a working definition of perceived safety and a suitable measurement tool for the user test, namely five perceived safety items from Osswald et al.'s (2012) car technology acceptance model (CTAM).

3.2.2 Expert Insights

In parallel with the initial literature review, a series of semi-structured interviews were conducted with experts in safety, automotive UX and HMIs. Participants included representatives from Volvo Cars and academic researchers from Chalmers University of Technology.

Interviews are a qualitative method used to explore participants' experiences, interpretations, and perspectives. While structured interviews follow a fixed script and unstructured formats are more exploratory, semi-structured interviews, used in this study, offer a balance of consistency and flexibility by following predefined themes while allowing for follow-up questions and emerging insights (Kvale & Brinkmann, 2015).

The interviews aimed to explore how perceived safety is interpreted and addressed in practice, and to identify interface areas where it is considered especially relevant. In addition to the interviews, complementary discussions with Volvo Cars specialists were held to map out systems in which perceived safety concerns were most prominent.

Several areas were considered, including the DIM and assisted driving features. However, the CSD, and specifically the Quick Controls menu, was repeatedly highlighted as a point of concern. Quick Controls was ultimately selected as the

primary focus due to its frequent mention, its feasibility for prototyping and testing, and the opportunity it offered to examine meaningful design trade-offs.

3.3 Research

Following the pre-study, the project advanced into a theory-driven research phase aimed at informing the selection of design variables for the subsequent user test. This phase consisted of a focused literature review and a benchmarking study, both intended to identify design-relevant patterns, tensions, and opportunities for exploring perceived safety. The insights also contributed to the formulation of later design guidelines.

3.3.1 Literature Review on HMI Design

The second literature review explored how interface design choices may influence perceived safety in in-vehicle HMIs.

Searches were once again conducted using Google Scholar, Web of Science, and Scopus, combining terms such as *user interface design*, *infotainment systems*, *perceived safety*, *automotive HMI design*, *driver distraction*, and *CSD interaction*.

3.3.2 Benchmarking

To complement the literature review, a benchmarking study was carried out to examine how contemporary vehicles implement Quick Controls or equivalent interface menus. Benchmarking refers to the systematic collection and comparison of specific features across systems to identify recurring patterns, performance gaps, and opportunities for improvement (Stapenhurst, 2009).

In this study, twelve vehicles from a diverse range of manufacturers were analysed, including both premium and emerging market brands such as BMW, Mercedes-Benz, Zeekr, Lotus, Tesla, Hyundai, and Xpeng (see Figure 3.2). The goal was to capture the current landscape of industry practice, which also informed the selection of design variables.



Figure 3.2: Benchmarking.

3.4 Design

As previously stated, the aim of the design phase was not to arrive at a final solution, but to systematically vary interface characteristics to examine their influence on perceived safety. This section outlines the selected variables, presents the resulting interface concepts, and concludes with the rationale behind key design decisions.

3.4.1 Design Variables

Navigation structure and *button size* were selected as the primary variables, based on patterns and tensions that emerged during the research phase.

Navigation structure refers to whether settings are displayed in a flat (broad) layout, where all options are immediately visible, or a nested (deep) layout, which requires the user to first select a category (e.g., Exterior lights) before accessing options in a modal window.

Button size denotes the dimensions of the touch targets used in both flat and nested layouts. Three sizes were tested, each exceeding the minimum thresholds recommended in prior automotive usability:

- **Small:** 20×20 mm – aligned with the lower bound of usability recommendations.
- **Medium:** 35×35 mm – selected for being slightly above the highest recommendation found.
- **Large:** 50×50 mm – chosen to be significantly larger while still accommodating the interface layout.

3.4.2 Design Concepts

Below are descriptions of the four interface concepts, designed in Figma in collaboration with UX designers at Volvo Cars, depicted in Figures 3.3–3.6

Flat Small (FS)

- Employs a flat layout where all Quick Control options are displayed on the initial screen, requiring no additional steps.
- Uses compact 20×20 mm touch targets.
- Text height is 3.7 mm for category labels and 2.9 mm for option buttons.

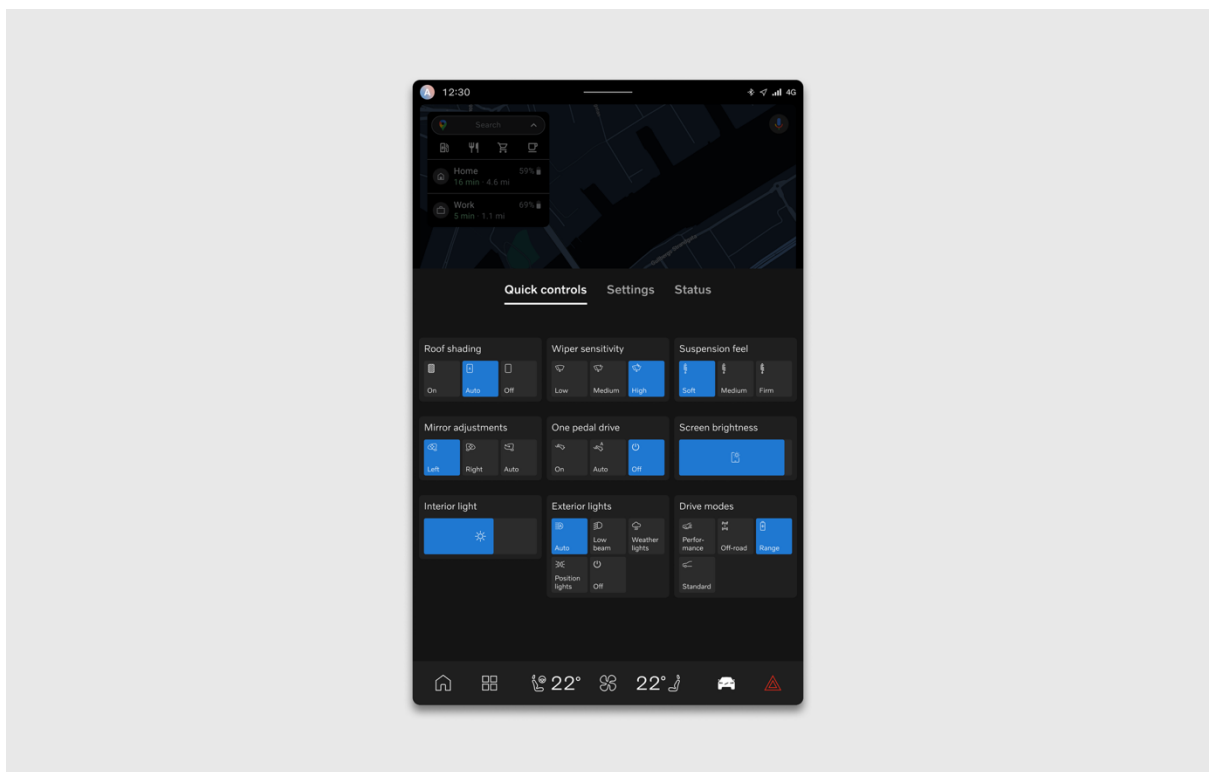


Figure 3.3: Concept FS.

Nested Small (NS)

- Applies the nested structure, requiring category selection before accessing options in a modal window.
- On all nested concepts, the initial category button measures 68×35 mm.
- A status indicator is displayed in the bottom-right corner of each category button, showing the currently selected option.
- Reduces on-screen complexity through progressive disclosure, but increases the number of interaction steps.
- Maintains 20×20 mm touch targets within modal to isolate the effect of structure from size.
- Uses the same text heights as FS, with 3.7 mm for category labels and 2.9 mm for option buttons and status indicators.

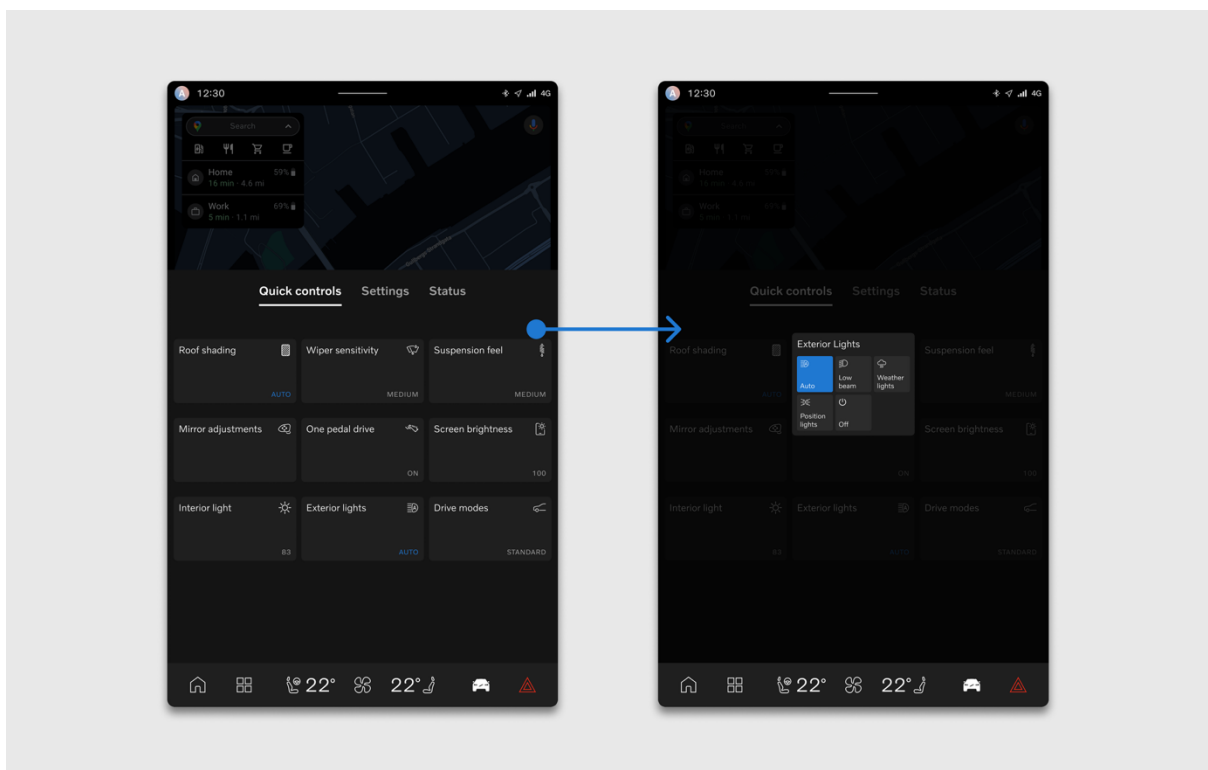


Figure 3.4: Concept NS.

Nested Medium (NM)

- Builds upon the nested layout of NS, increasing touch targets to 35×35 mm within the modal.
- Isolates the impact of a moderate-sized increase within a consistent interaction flow.
- Text height of 6.5 mm for category labels, 5.0 mm for option button text, and 2.9 mm for status indicators.

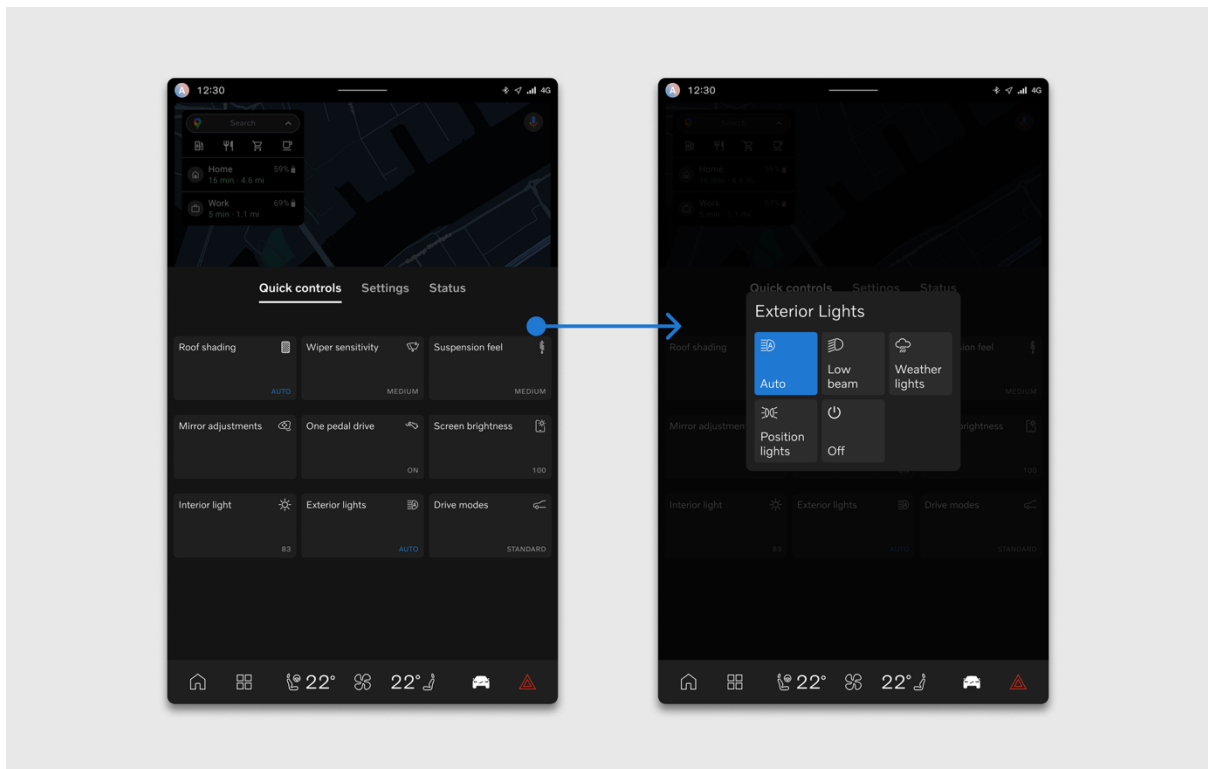


Figure 3.5: Concept NM.

Nested Large (NL)

- Continues the nested structure, further enlarging touch targets to 50×50 mm within the modal.
- Preserves interaction logic to focus on size as the only changing variable.
- Text height of 9.3 mm for categories, 7.3 mm for option buttons, and 2.9 mm for status indicators.

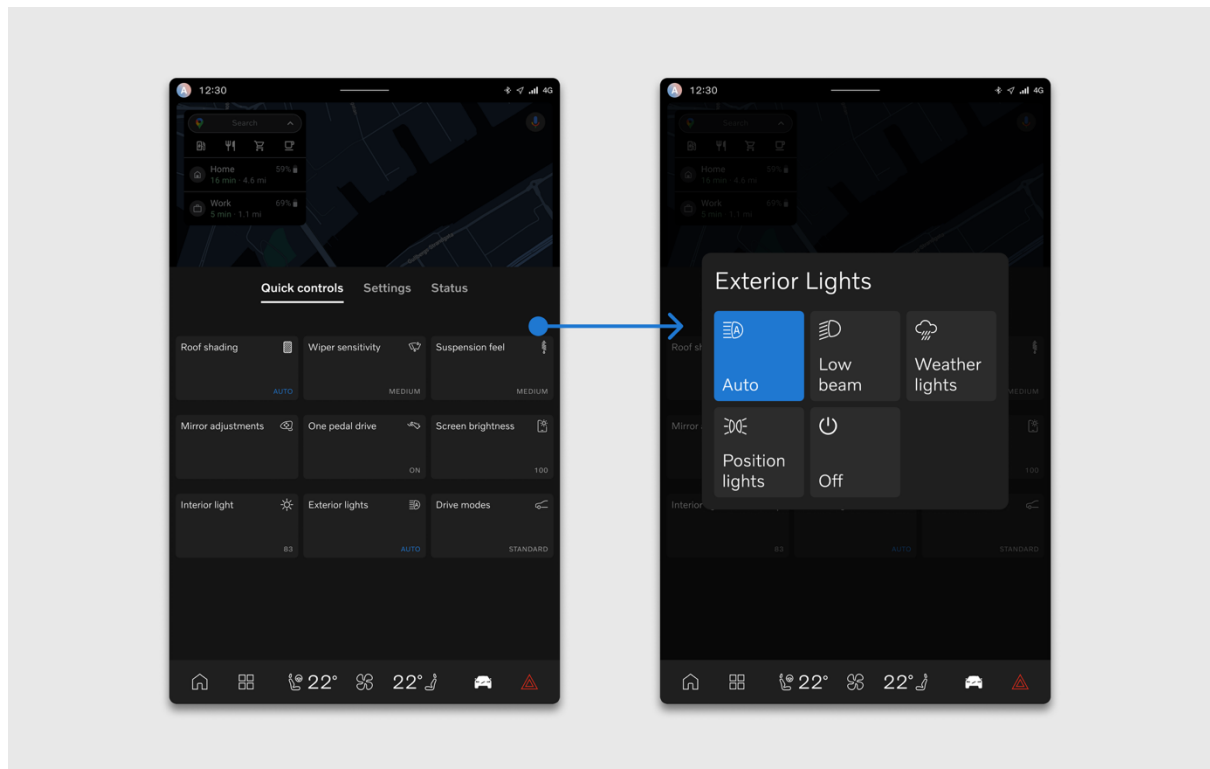


Figure 3.6: Concept NL.

All four concepts contained the same categories and options, with consistent category placement across interfaces (see Figure 3.7).

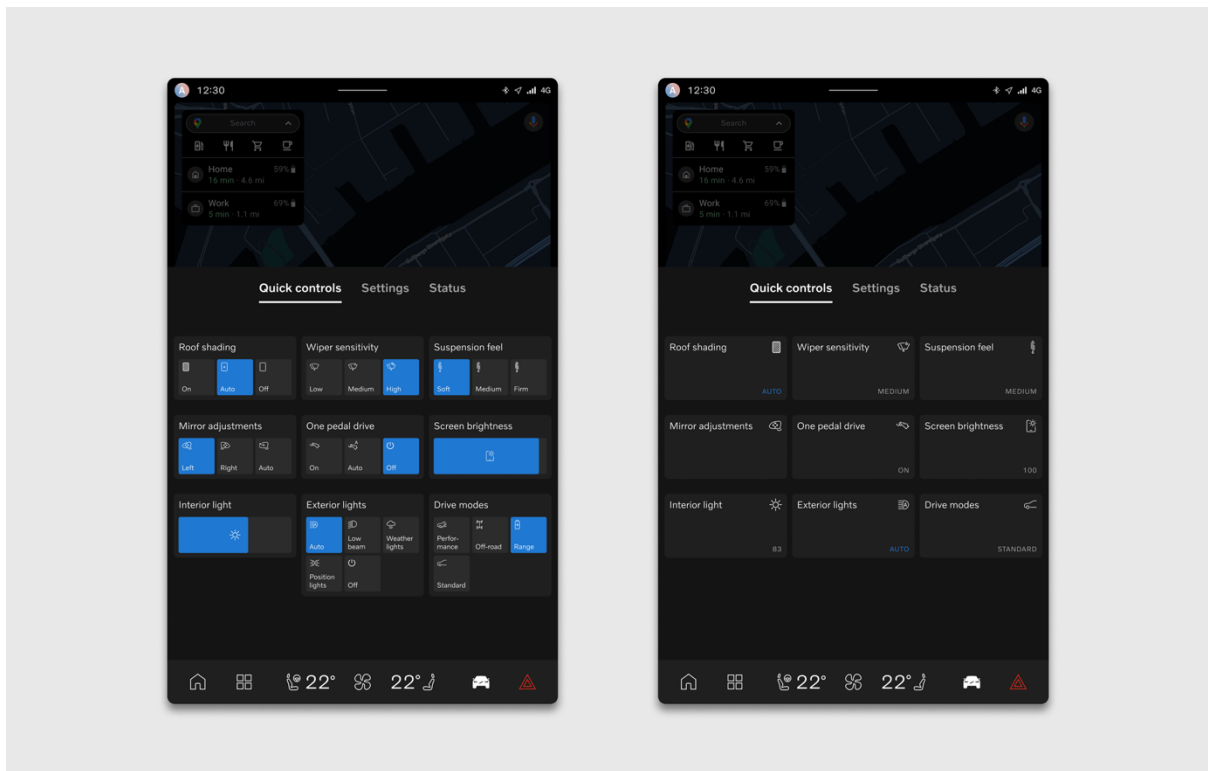


Figure 3.7: Category placement.

3.4.3 Design Context & Limitations

The selection and number of interface functions, comprising 9 categories and 26 options, were not defined as experimental variables. Instead, they were based on components and patterns observed in existing Quick Controls implementations.

Other interface characteristics, such as colour palette and typography, were not manipulated. As these aspects also fell outside the study's scope, the design followed Volvo's future design language in terms of visual style. As stated, for confidentiality reasons, the visual expression in this report is consistent with vehicles currently available to the public.

Apart from the Quick Controls page, being the only modified part of the system, the rest of the interface functioned and could be navigated like a typical future production vehicle. This approach ensured high fidelity and contributed to a realistic test

experience. While the interface allowed for natural navigation, most buttons were not connected to live vehicle systems. This was not deemed a limitation, as the study focused on interaction within the interface rather than the underlying functionalities.

3.4.4 Design Rationale

Button size, text, and spacing were treated as a cohesive design bundle rather than isolated variables. This approach acknowledged that changes in button size inherently affect other elements within modal containers and allowed for high-fidelity design without unrealistic combinations, such as large modals paired with small text, which could have affected the participants' aesthetic judgements of the interface.

To isolate the effect of navigation structure, FS and NS featured identical button sizes. This enabled a direct comparison between flat and nested layouts while keeping all other variables constant, except for the inherent layout differences introduced by the use of modals.

To assess the impact of button size, the three nested variants (NS, NM, NL) maintained a consistent interaction structure while varying only the size of the touch targets. This approach enabled the effect of button size to be evaluated independently within a consistent navigational flow.

3.5 User Test

The user test followed the design phase and is described in detail in this section. It covers the contextual setup, data collection methods, test procedure, and subsequent analysis.

3.5.1 Test Environment & Setup

The user test was conducted at Volvo Cars' proving ground (VCT Test Track) in Torslanda. The track consists of two straight sections (1.8 km each) connected by a turnaround (see Figure 3.8) and was selected to provide a consistent, low-distraction driving environment. By eliminating external factors such as traffic, pedestrians, and variable road conditions, the controlled setting allowed for uniform task execution

and reduced data variability. A maximum of three vehicles were present on the track at any time, further supporting test reliability and replicability.



Figure 3.8: VCT Test Track in Torslanda, Gothenburg.

Participants drove a modified Polestar 2 provided by Volvo Cars (see Figure 3.9). The four Quick Controls concepts were displayed on a laptop mounted to replicate the native position of the CSD in the dashboard. The prototype was developed and integrated with support from Volvo’s prototype team. A MacBook Pro 16” (M1, 2021), operated by one of the test leaders from the back seat, was used to switch between concepts during the test. To minimise glare and improve visibility, the laptop screen was covered with an anti-reflective film.

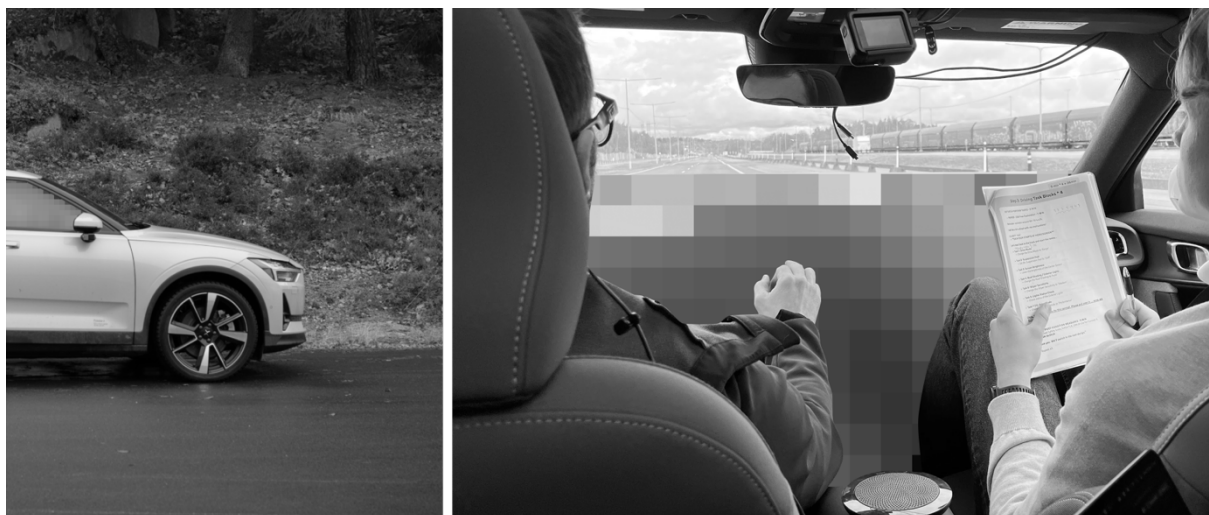


Figure 3.9: Prototype Car and In-vehicle Setup During Test.

The test setup also included the following equipment:

- Tobii Pro Glasses 3 eye-tracking system
- GoPro Hero 12 for video recording
- Jabra SPEAK 510 speakerphone for audio capture

Two test leaders were present in the vehicle throughout each session. One facilitated the test protocol and guided the participant through the tasks, while the other documented observations, monitored the eye-tracking system, managed interface transitions, and ensured participant safety.

3.5.2 Data Collection

A combination of objective eye-tracking metrics, subjective scales, and qualitative methods such as think-aloud and interviews were employed to collect a robust and triangulated dataset. This multi-method approach supported the broader objective of the study, i.e. to explore whether perceived safety can be intentionally designed for in the context of in-vehicle touch interfaces, and whether it can be meaningfully measured as a distinct quality of user experience. By comparing perceived safety ratings with usability, workload, and behavioural indicators (eye-tracking metrics), the study aimed to assess whether perceived safety functions as an independent construct or primarily reflects general performance-related factors.

Objective Measures

Eye-tracking data were collected using Tobii Pro Glasses 3 to provide objective insight into visual attention and cognitive demand. The Tobii system employs binocular dark-pupil tracking and an integrated scene camera to capture gaze behaviour in naturalistic environments (Tobii, n.d.). The key metrics analysed included:

- **Total eyes-off-road time (TEORT):** The cumulative time the participant's gaze was directed toward the CSD and away from the forward road scene while completing a task
- **Task completion time (TCT):** The total time required to complete each task

- **Glance count:** The number of distinct visual visits to the CSD per task
- **Average glance duration:** The mean duration of each glance toward the CSD, reflecting how long participants typically looked away from the road
- **Misclicks:** instances where participants had the correct intention but unintentionally tapped outside the intended touch target

Subjective Measures

Standardised questionnaires provide a validated means of collecting consistent and comparable data by translating abstract constructs into quantifiable items (Malhotra, 2006). In this study, three instruments were used. See Appendix C for the full scales:

- **Perceived safety scale:** A five-item scale derived from the CTAM by Osswald et al. (2012) was used to assess how safe, focused, and in control participants felt when interacting with the concepts. Items addressed perceived danger, attentional demand, distraction from driving, confidence during use, and ease of use without visual attention. Each item was rated on a 7-point Likert scale (1 = Strongly Disagree, 7 = Strongly Agree).
- **NASA Task Load Index (NASA-TLX):** A multidimensional tool for evaluating perceived workload across six factors: mental, physical, and temporal demand, effort, frustration, and perceived performance. Each factor was rated on a 21-point scale (1 = Very Low, 21 = Very High, with one exception) (Hart & Staveland, 1988).
- **UMUX-Lite:** Introduced by Lewis et al. (2013), this short-form usability measure consists of two items evaluating perceived usefulness and ease of use. Responses were recorded on a 7-point Likert scale (1 = Strongly Disagree, 7 = Strongly Agree). UMUX-Lite has been shown to correlate strongly with longer usability instruments.

The Likert scale, developed by Rensis Likert (1932), is a widely used psychometric tool for capturing attitudinal data through fixed-point agreement ratings. Its numerical structure enables both interpretability and statistical analysis of subjective experience.

When completing the questionnaires, participants were instructed to verbalise their thoughts using a think-aloud protocol. Originally developed within cognitive psychology, think-aloud is a well-established method in usability research for accessing users' cognitive processes during task performance (Nielsen, 1993). Verbalisation allowed researchers to better understand participant reasoning, decision points, and real-time reactions.

Participants also engaged in brief semi-structured interviews. These explored topics such as perceived sources of stress or comfort, interpretation of questionnaire items, and general impressions regarding clarity and interaction flow. Additionally, participants were asked to indicate preferences across several dimensions:

- Overall preferred concept
- Perceived safest concept
- Safest button size (small, medium or large)
- Safest structure (flat or nested)

These qualitative insights added interpretive depth to the quantitative data and supported a more comprehensive understanding of user experience patterns.

3.5.3 Participants

Participants were recruited from Volvo Cars' internal test participant pool, consisting of employees who voluntarily enrol in internal research initiatives. Both Swedish- and English-speaking employees were eligible to participate. Participants were screened for eye-tracking compatibility. Individuals using bifocal lenses, those who had recently undergone eye surgery, or those known to have factors affecting eye-tracking accuracy were excluded.

A large number of employees from the internal pool were contacted, resulting in an initial group of 23 participants. Following a few cancellations, a total of 19 participants completed the user test. All were active drivers between the ages of 25 and 60, with regular experience using touch-based in-vehicle interfaces. Approximately half were current Volvo drivers (model year 2017 or newer).

Table 1: User test participants (TPs).

TP	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Age	48	38	36	50	36	29	46	37	48	49	37	26	35	36	52	52	45	47	43
Gender	F	M	M	F	M	M	M	M	F	F	M	M	M	M	M	M	M	M	M

The final sample comprised 14 male and 5 female participants, representing a broad spectrum of driving and interface experience. Although an even gender distribution was intended, the final composition was ultimately constrained by the demographics of the available internal test group.

3.5.4 Test Design & Procedure

A within-subjects design was employed, where each participant completed six tasks on each of the four interface concepts. This design, also referred to as a repeated measures design, involves collecting multiple observations from the same participants under different conditions (Kim, 2010). It enables direct comparisons within individuals, reducing variability caused by individual differences and increasing the sensitivity of statistical analyses. While particularly effective for evaluating multiple conditions with a limited sample size, this approach can introduce order effects such as learning, fatigue, or carryover.

To mitigate these risks, the order of both interface concepts and tasks was counterbalanced across participants (see Figure 3.10). Counterbalancing systematically varies the sequence of conditions to prevent biases related to task order (Brooks, 2012).

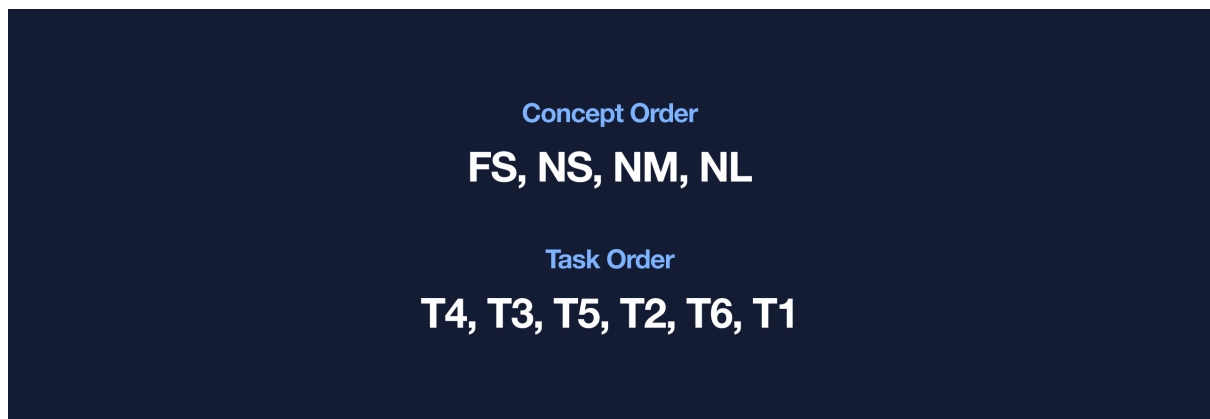


Figure 3.10: Example of concept and task order for one participant.

Test Procedure

Each test session lasted approximately 90 minutes and followed a consistent structure guided by a scripted protocol (see Appendix B). This ensured methodological rigour and minimised variability between sessions.

Introduction and Setup

Participants were welcomed at a designated meeting point within the Volvo Cars area and asked to read and sign a consent form. During the drive to the VCT Test Track, they received a verbal briefing covering the study's objectives, procedures, and safety considerations.

Upon arrival, participants were seated in the driver's seat, and the test leader began equipment setup, including calibration of the eye-tracking system. A short walkthrough of the vehicle's controls and in-test safety instructions was also provided before starting the session.

Familiarisation

To reduce novelty effects and ensure participants were comfortable with the setup, they were first given time to explore the initial interface concept while the vehicle remained stationary. This was followed by a warm-up lap without any tasks, allowing them to become acquainted with the vehicle, test route, and overall environment.

Test Blocks

Participants completed six standardised interaction tasks (T1–T6; see Figure 3.11) for each of the four interface concepts. Each concept was tested during a separate lap. Tasks were verbally prompted by the facilitator at appropriate moments to ensure safety and consistency. All tasks began from the CSD home screen.

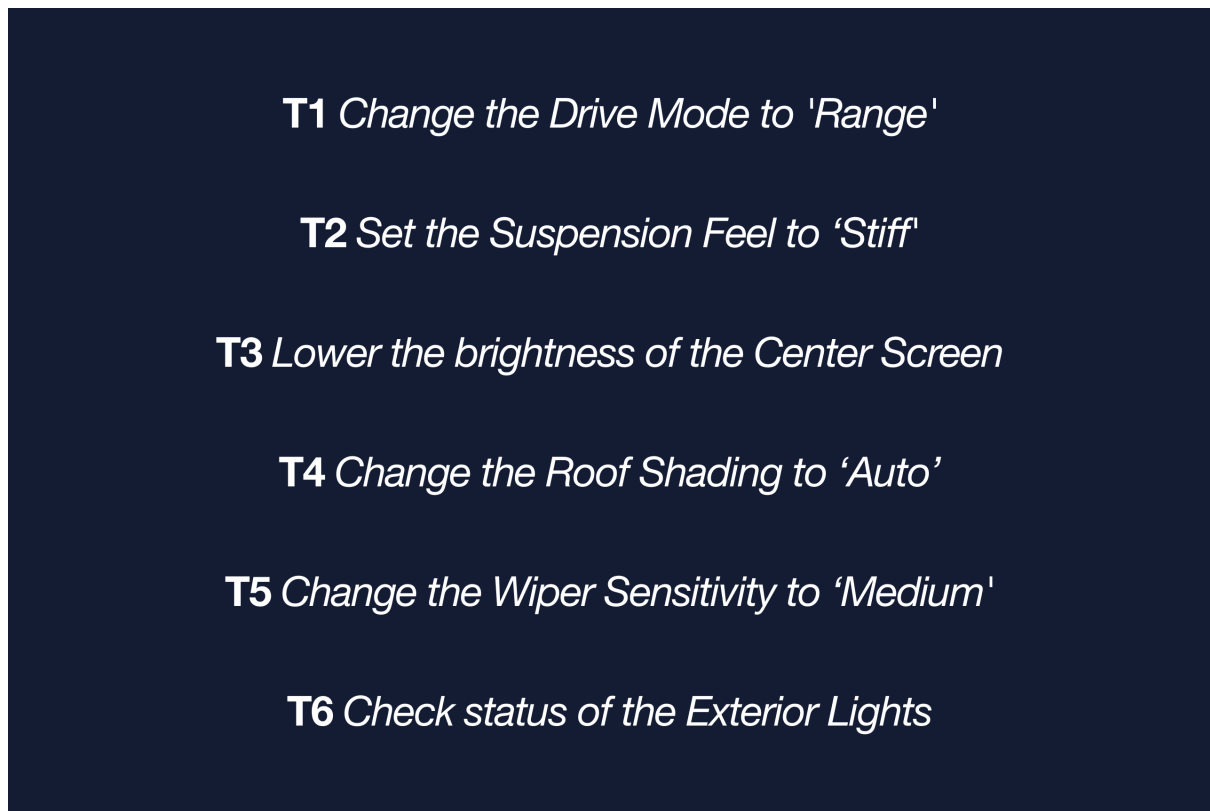


Figure 3.11: The six interaction tasks.

After each lap, the vehicle was parked, and participants completed the three subjective scales (perceived safety, UMUX-Lite and NASA-TLX).

Post-Test Interview

Upon completing the test blocks for all four concepts, participants took part in a brief semi-structured interview. As outlined previously, the discussion covered perceived sources of stress or comfort, clarity, interaction flow, and participants' rankings of the concepts.

3.5.5 Data Analysis

This section outlines how data from the user test was processed and analysed. A software issue occurred during the first eight sessions, affecting only the three nested interface concepts (NS, NM, NL). Specifically, when a modal overlay was active, pressing “Home” did not return the user to the main screen unless the modal was manually closed first. While this issue did not affect the eye-tracking data, it may have influenced participants’ subjective responses.

To ensure transparency, results from subjective measures are reported separately for subgroups TP1–8 and TP9–19, in addition to the aggregated dataset. However, due to the limited size of each subgroup and the absence of consistent differences, no inferential comparisons between them were conducted.

Furthermore, participant 19’s eye-tracking data was excluded due to a recording failure, though their questionnaire responses were retained.

Objective Quantitative Data

For the objective eye-tracking metrics (TEORT, TCT, Glance Count, and Average Glance Duration), median values were calculated per task and then averaged across tasks to generate concept-level aggregates.

Compliance with the second NHTSA guideline, for average glance duration, was also assessed. Violations were counted per concept. Applied to this study with a sample of 18, a violation occurred if three or more participants exceeded the threshold for a given task and concept.

The other two NHTSA guidelines were not evaluated. The first, which concerns the total number of individual glances per task, could not be assessed due to limitations in the dataset, which only included averaged durations. The third, specifying a maximum TEORT of 12.0 seconds per task, was not considered critical in this context. All participants remained well within this limit, as the tasks performed in Quick Controls were brief and required only minimal TCTs.

Misclicks were analysed separately. For each concept, an error rate was calculated by dividing the number of misclicks by the total number of task attempts. This percentage reflects the frequency of erroneous inputs. For instance, an error rate of 6% indicates that six misclicks would be expected for every 100 tasks performed using that concept.

Subjective Quantitative Data

Responses from the perceived safety scale, NASA-TLX, and UMUX-Lite were analysed using a within-subjects approach. For the perceived safety scale, items 4 and 5 were first reverse-coded to align with the directionality of the remaining items. All five items were then reverse-coded so that higher scores reflected greater perceived safety. UMUX-Lite scores were converted to a 0–100 scale, by standard practice, to improve comparability and interpretability. NASA-TLX scores were also reverse-coded, meaning that higher values indicated lower perceived workload.

Additional subjective inputs, including rankings and preference selections, were analysed descriptively to support interpretation and argumentation related to interface preferences and patterns in perceived safety.

Statistical Significance Testing

To determine whether the observed differences between interface concepts were statistically significant, a series of tests were conducted using Python (relevant scripts noted in parentheses).

First, a Friedman test (`scipy.stats.friedmanchisquare`) was used to examine differences between the three nested concepts (NS, NM, NL) across both objective metrics (TEORT, TCT, glance count, average glance duration) and subjective metrics (perceived safety, NASA-TLX, UMUX-Lite). The Friedman test is a non-parametric alternative to repeated-measures ANOVA, appropriate for within-subject designs where normality cannot be assumed. It ranks values within each participant and tests whether rank distributions differ across conditions (Pereira et al., 2015). Statistical significance was evaluated at $\alpha = .05$.

Second, a Wilcoxon signed-rank test (`scipy.stats.wilcoxon`) was used to compare FS against the average of the three nested variants. This enabled a broader flat-versus-nested comparison using the same set of metrics. The Wilcoxon test assesses whether the median difference between paired observations differs from zero, making it suitable for repeated-measures data without normality assumptions (Lowry, 1998). The same significance threshold ($\alpha = .05$) was applied.

For the subjective metrics, additional pairwise Wilcoxon tests were conducted between the three nested concepts (NS, NM, NL), motivated by near-significant findings in the Friedman test and visible differences in participant ratings. No pairwise comparisons were conducted for the objective metrics, as the Friedman test revealed no significant differences across nested variants, and visual inspection of the raw data indicated highly similar distributions.

To explore relationships between the subjective scales (perceived safety, NASA-TLX, UMUX-Lite), repeated measures correlation analyses (`pingouin.rm_corr`) were conducted pairwise between perceived safety and each of NASA-TLX and UMUX-Lite. This method estimates associations between variables measured multiple times within individuals by removing between-subject variance. The resulting correlation coefficient (r), ranging from -1 to +1, is accompanied by a significance test indicating the strength and direction of the relationship (Bakdash & Marusich, 2017). Statistical significance was again evaluated at $\alpha = .05$.

Qualitative Data

A qualitative synthesis was conducted based on detailed notes taken during each test session. Audio recordings were selectively consulted to verify specific quotes or clarify ambiguous observations; however, full transcriptions were not produced.

To analyse the material, an affinity diagram was constructed. Related observations and participant quotes were grouped by thematic similarity, allowing clusters to emerge organically. These clusters formed broader categories reflecting key aspects of driving performance, interaction behaviour, usability, and perceived safety. This method supports a grounded and open-ended approach to analysis, where structure

arises directly from the data (Hartson & Pyla, 2012). The resulting insights were used to support and contextualise the quantitative findings

3.6 Synthesising & Design Guidelines

Drawing from both the user test and the preceding research, a set of design guidelines was formulated to inform future work on perceived safety in digital in-vehicle interaction. Together, they provide a practical foundation for designing and evaluating safety-critical interactions, with particular relevance to Quick Controls within the CSD.

4 Results

This chapter outlines the main findings from the user test, divided into quantitative metrics and qualitative insights. Results are presented by interface concept to highlight patterns in performance, user experience, and perceived safety.

4.1 Quantitative Results

This section presents the quantitative results of the user test, covering both objective performance metrics from eye-tracking and the subjective responses collected via post-condition questionnaires and ranking tasks.

4.1.1 Objective Measures

Results are presented for the five eye-tracking-based measures: TEORT, TCT, glance Count, average glance duration and misclicks. As noted, one participant was excluded due to a recording failure, leaving 18 participants in the objective dataset. Compliance with the second NHTSA guideline is also reported to contextualise glance behaviour relative to established safety guidelines.

Total Eyes-Off-Road Time (TEORT)

As shown in Figure 4.1, FS yielded the lowest average TEORT at 2.56 seconds, followed by NL at 2.94, NS at 3.00, and NM at 3.06. No statistically significant differences were found between the three nested concepts ($p = .513$, Friedman test), and the comparison between FS and the nested average also did not reach significance ($p = .102$, Wilcoxon test).

Although the difference was not statistically significant, FS consistently produced slightly lower TEORT values across tasks, a trend also reflected in the near-significant Wilcoxon result.

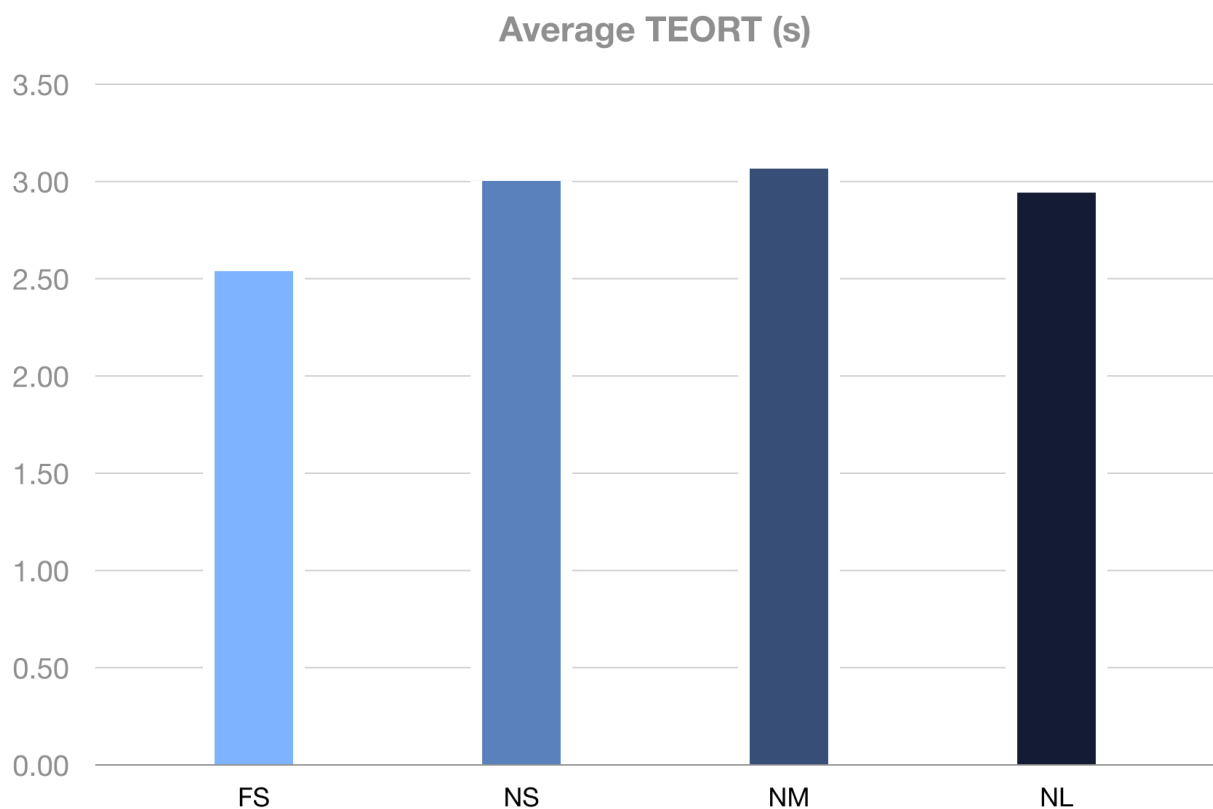


Figure 4.1: Average TEORT per concept.

Task Completion Time (TCT)

Figure 4.2 shows that tasks were completed fastest using FS, with an average duration of 3.00 seconds. The nested designs resulted in longer durations, with NL at 3.67 seconds, NS at 3.71, and NM at 3.89. No significant differences were found between the nested concepts ($p = .680$), but FS was significantly faster than the nested average ($p = .038$).

These results indicate a time-saving benefit for FS in terms of interaction speed. This was one of three objective measures where FS outperformed the other concepts with a statistically significant margin.

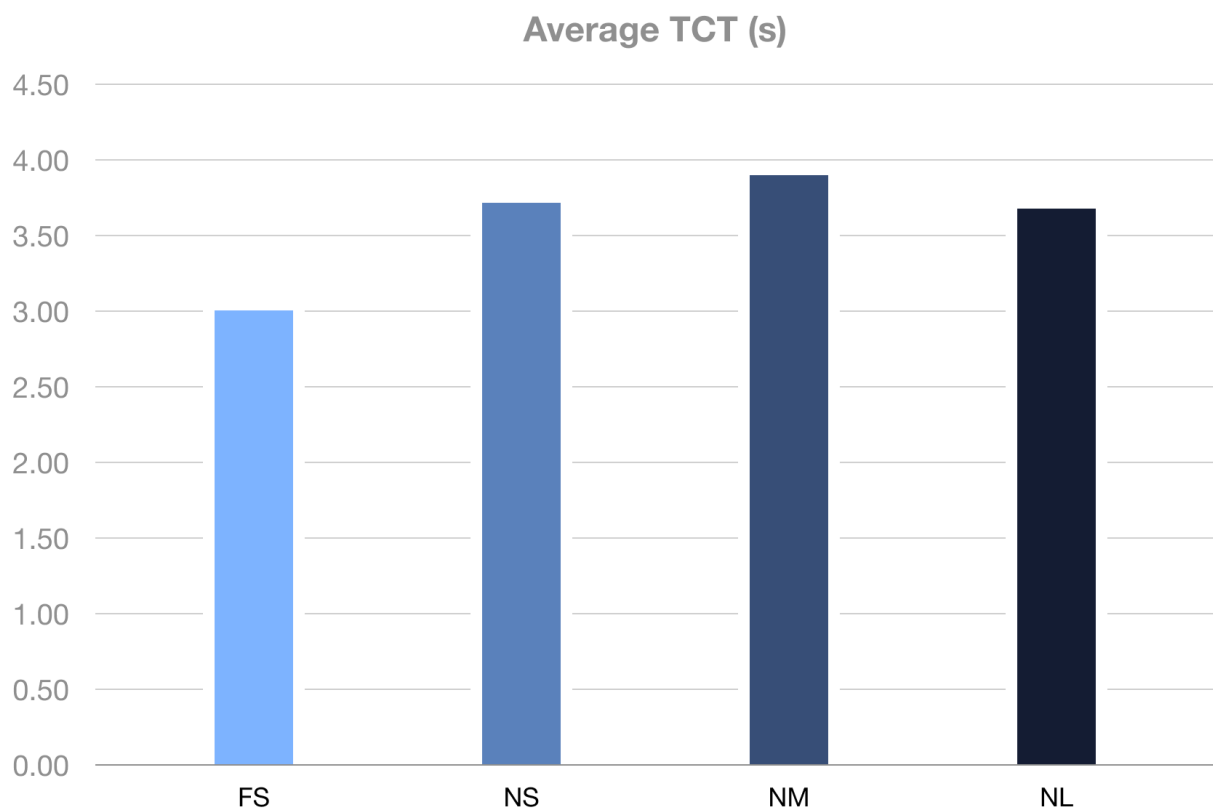


Figure 4.2: Average TCT per concept.

Glance Count

As illustrated in Figure 4.3, FS resulted in the fewest average visits to the CSD, at 1.92 per task. The nested designs led to more frequent glances, with NS at 2.33, NL at 2.42, and NM at 2.58. Differences among the nested concepts were not statistically significant ($p = .179$), but the contrast between FS and the nested average was highly significant ($p = .00025$), indicating a large effect.

The lower glance count for FS reflects its flat navigation structure, which involved fewer interaction steps and thus required fewer glances to complete each task.

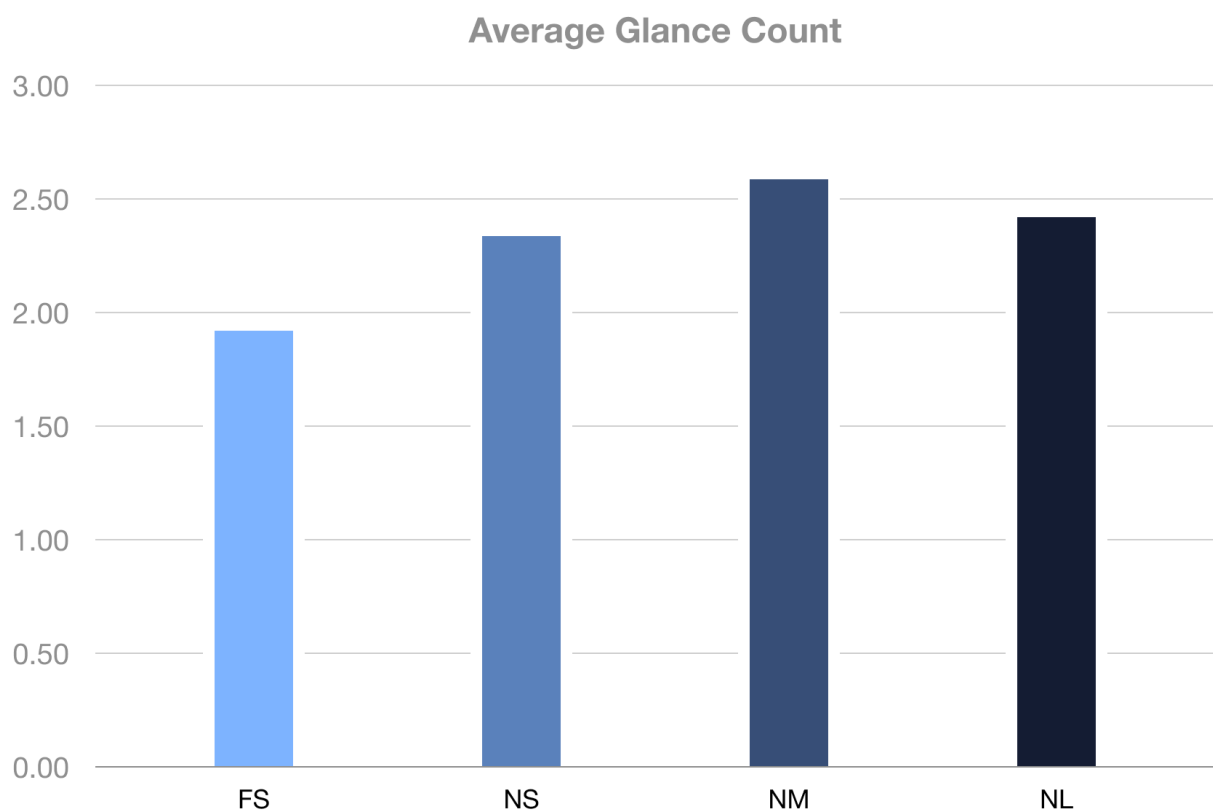


Figure 4.3: Average CSD visit count per concept.

Average Glance Duration

As shown in Figure 4.4, FS resulted in the longest average glance duration to the CSD at 1.21 seconds. The nested concepts had shorter durations, with NL at 0.94 seconds, NS at 0.99, and NM at 1.00. Differences between the nested designs were not significant ($p = .607$), but FS differed significantly from the nested average ($p = .00025$), indicating a large effect.

While FS prompted fewer glances overall, each glance tended to be longer. This may suggest that more information was processed per visit, possibly due to layout density or fewer navigation steps.

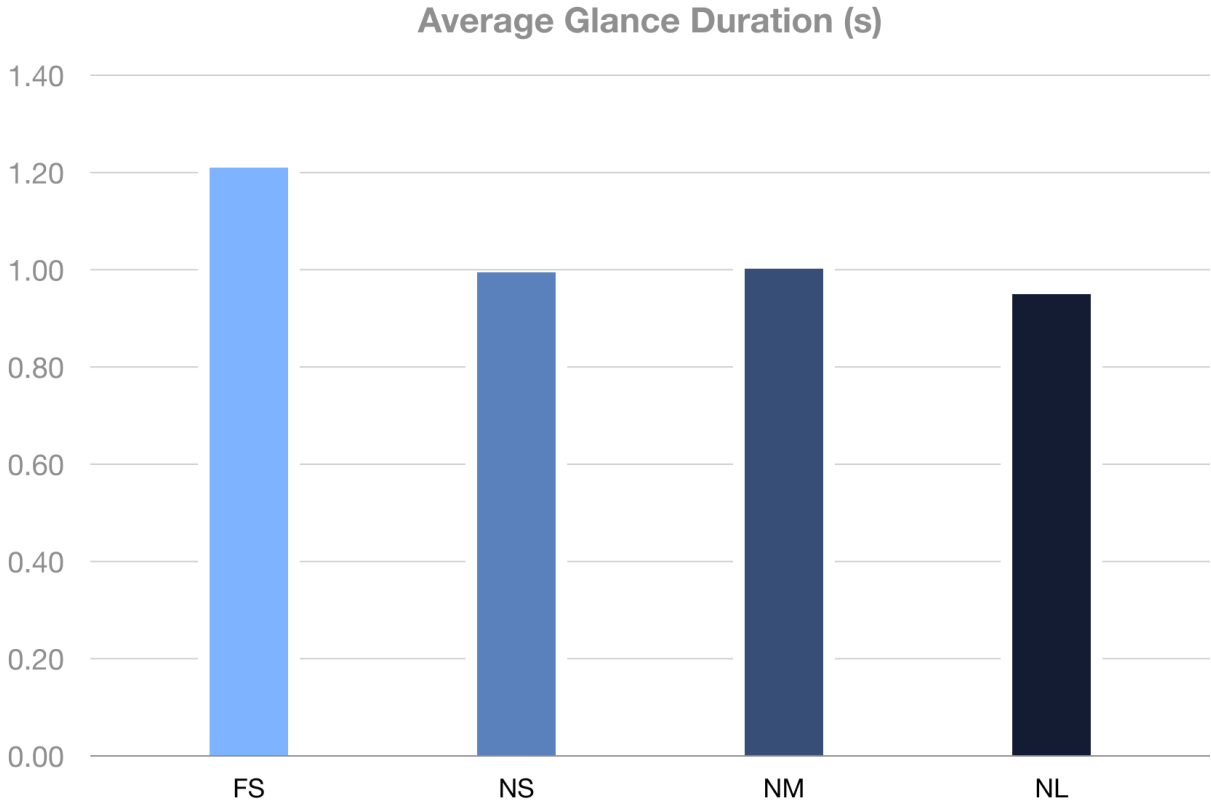


Figure 4.4: Average CSD glance duration per concept.

Misclicks & Error Rate

As shown in Figure 4.5, the highest error rate was recorded for NS at 10.19%, followed by FS at 6.48, NM at 4.63, and NL at 3.70. Overall, misclicks were relatively rare across all concepts but occurred most frequently during Task 3, particularly for NS and NM.

This pattern appears to reflect differences in button size. NS and FS, which both featured small buttons, showed more frequent errors than NM and NL.

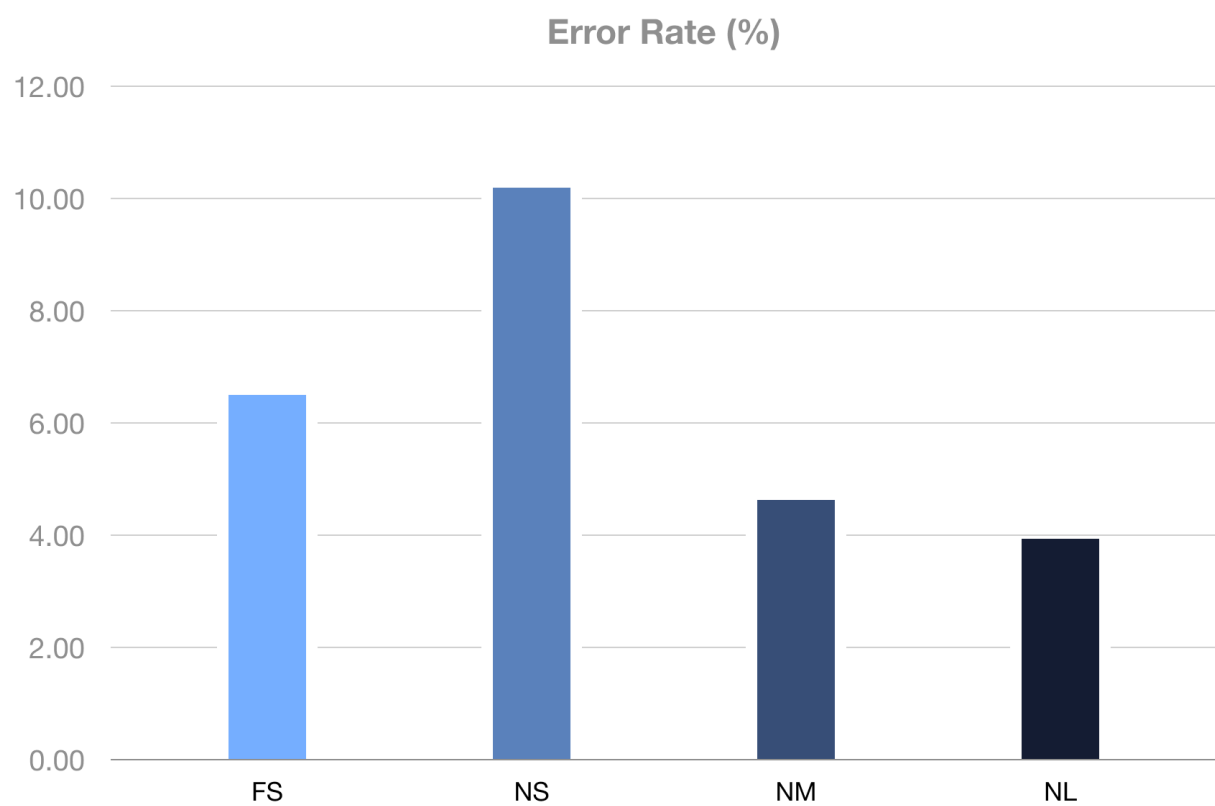


Figure 4.5: Average error rate per concept.

Compliance with NHTSA Guideline 2

FS showed the highest number of average glances above the 2-second threshold, with 14 in total, compared to 4 each for NL and NS, and 3 for NM (see Appendix D for a full breakdown). This resulted in two violations for FS and none for the remaining concepts (see Figure 4.6). That is, for two separate tasks, three or more participants exceeded the threshold when using FS.

These findings support the interpretation that although FS reduced the number of glances and overall task time, it also encouraged longer individual glances.

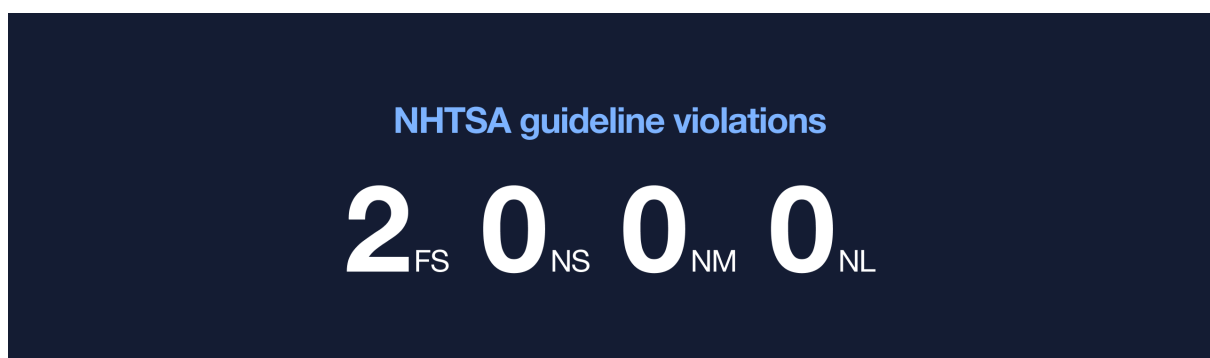


Figure 4.6: Number of NHTSA Guideline 2 violations per concept.

4.1.2 Subjective Measures

This section presents participants' subjective evaluations of each interface concept, based on the perceived safety scale, UMUX-Lite, and NASA-TLX. The results presented reflect the full sample. Disaggregated data for participants TP1–8 and TP9–19 are available in Appendix D.

In addition to the scale ratings, the section includes participants' concept and structure rankings, as well as an analysis of correlations between the subjective measures.

Perceived Safety Scores

As shown in Figure 4.7, NM received the highest average perceived safety score at 4.04, followed by NL at 3.98, FS at 3.83, and NS at 3.76. The Friedman test comparing the three nested concepts (NS, NM, NL) revealed no statistically significant differences overall ($p = .165$). Similarly, when comparing FS to the average of the nested concepts, no significant difference was observed ($p = .922$, Wilcoxon test), indicating that overall perceived safety did not differ meaningfully between the structural approaches.

However, the pairwise comparisons among the nested concepts revealed more nuanced distinctions. The Wilcoxon test showed that NM was rated significantly higher than NS ($p = .048$). The comparison between NS and NL approached significance ($p = .140$), while NM and NL were rated similarly ($p = .691$).

Taken together, these findings indicate that although average ratings were relatively close, NS was consistently perceived as less safe, particularly in comparison to NM. This pattern suggests that touch target size influenced perceived safety even when all sizes exceeded recommended usability guidelines. While FS trailed NM and NL in average ratings, it remained broadly comparable and did not differ significantly from the nested group as a whole.

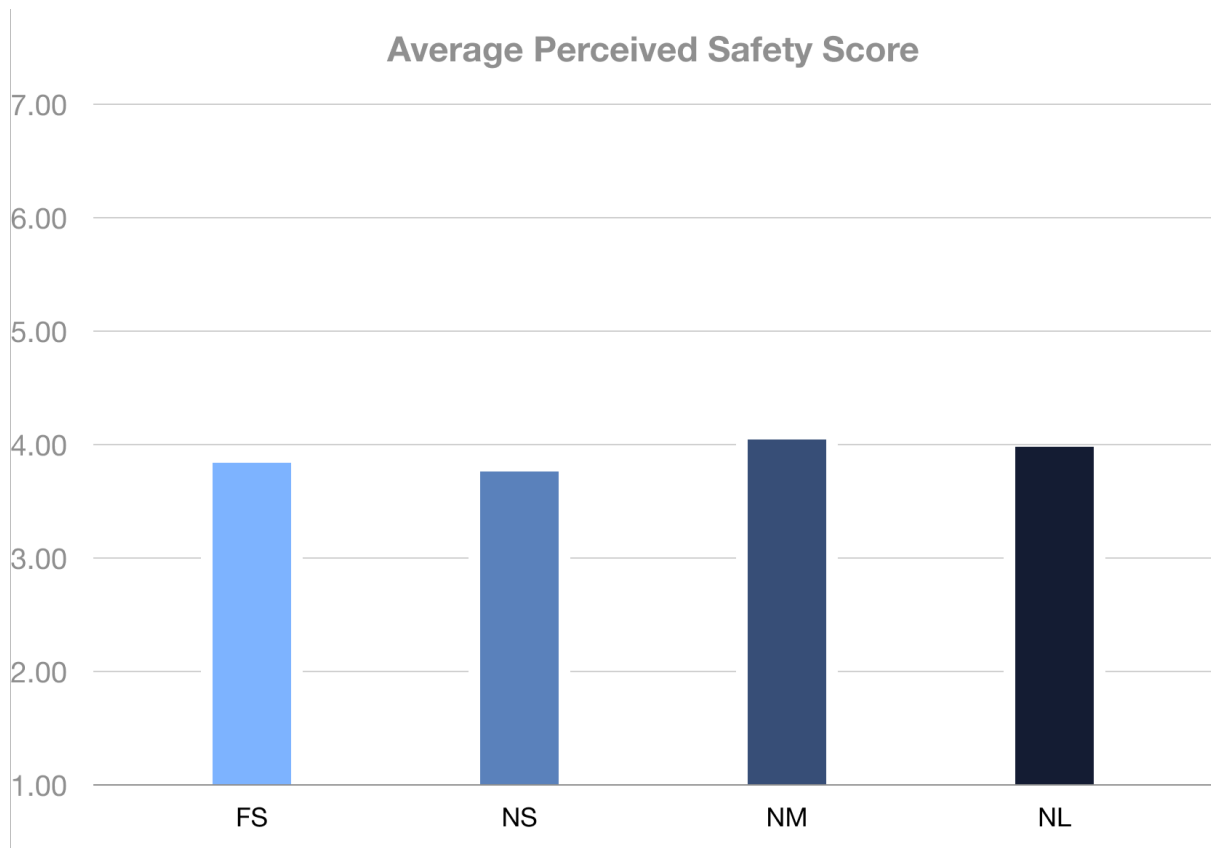


Figure 4.7: Average Perceived Safety scores per concept.

NASA-TLX Scores

The NASA-TLX results, presented in Figure 4.8, showed that NL received the highest average score at 16.49, followed by NM at 16.41, FS at 16.29, and NS at 15.56. This indicates that NL was associated with the lowest perceived workload, while NS was associated with the highest. The Friedman test revealed no statistically significant differences between the nested concepts ($p = .119$), and FS did not differ significantly from the average of the nested group ($p = .332$, Wilcoxon test).

However, the pairwise comparisons revealed a more differentiated pattern. NL was rated significantly lower in workload than NS ($p = .018$), corresponding to a large effect. The contrast between NM and NS approached significance ($p = .113$), while NM and NL were rated similarly ($p = .865$).

These results show that NS was associated with higher perceived workload than the other nested alternatives, likely due to its combination of small buttons and added interaction depth. In contrast, NL and NM were experienced as less demanding, suggesting that thoughtful button sizing can help reduce workload without compromising interaction performance. FS also performed well but did not differ significantly from the nested average.

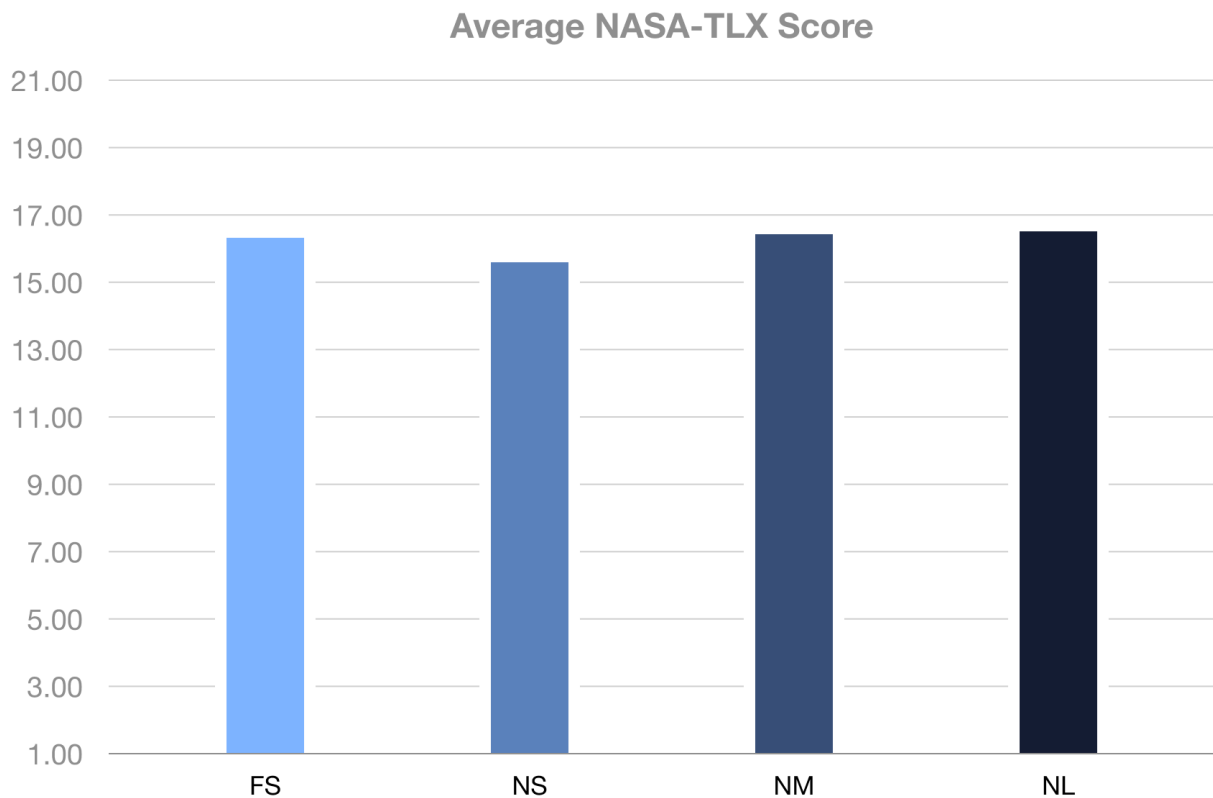


Figure 4.8: Average NASA-TLX scores per concept.

UMUX-Lite Scores

Figure 4.9 displays the usability scores derived from the UMUX-Lite scale. FS received the highest average rating at 78.51, followed by NL at 77.19, NM at 76.32, and NS at 71.93. No statistically significant differences were found between the three nested designs ($p = .242$, Friedman test), and FS did not differ significantly from the nested average ($p = .246$, Wilcoxon test), indicating broadly similar usability perceptions across the concepts.

However, a closer examination of the nested designs revealed that NL was rated significantly higher than NS ($p = .035$), corresponding to a large effect. The comparison between NS and NM approached significance ($p = .122$), while NM and NL showed no meaningful difference ($p = .719$).

These results suggest that NS was perceived as less usable than the other nested designs, particularly NL. This reinforces the role of button size in shaping usability judgments. Although FS topped the usability rankings, its difference from the nested group was not statistically significant.

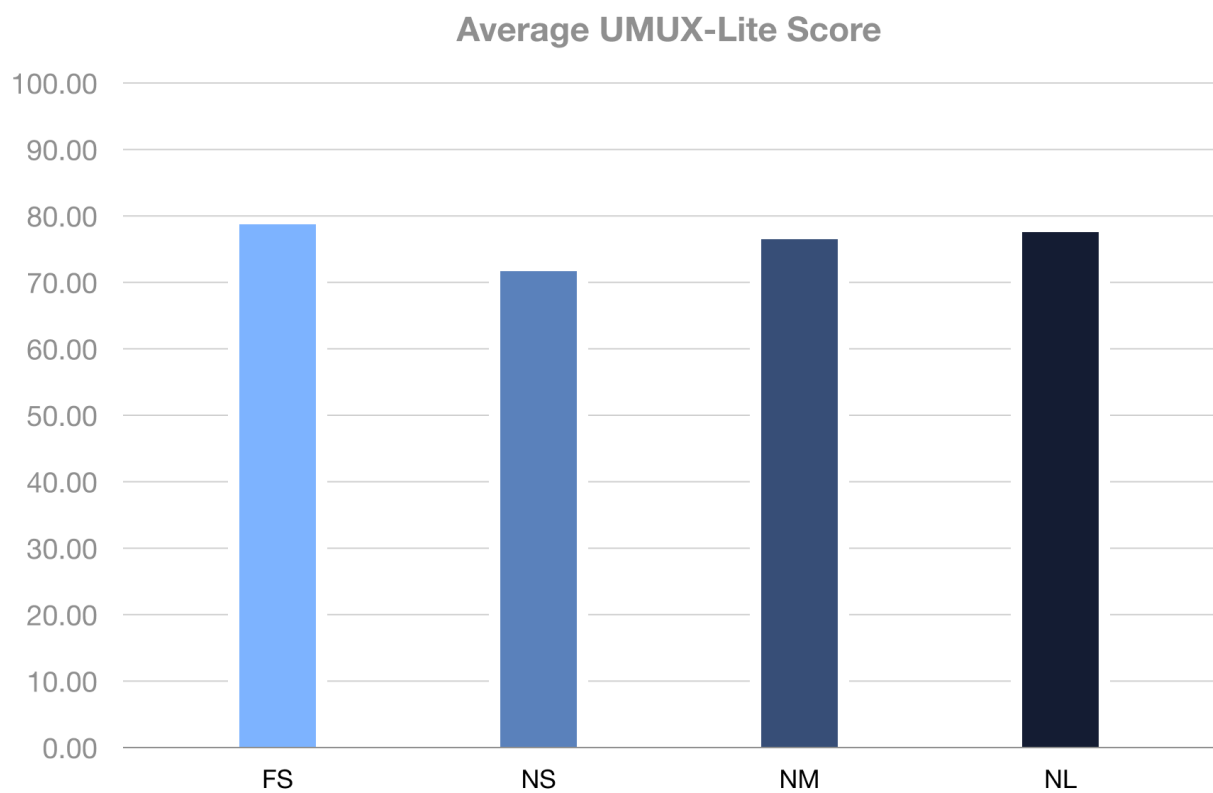


Figure 4.9: Average UMUX-Lite scores per concept.

Concept & Structure Rankings

FS was most frequently preferred overall, receiving 13 first-place rankings (see Figure 4.10). It was followed by NM with 5 and NL with 1. NS was not ranked first by any participant and was most commonly placed last. The full table of rankings is available in Appendix D.

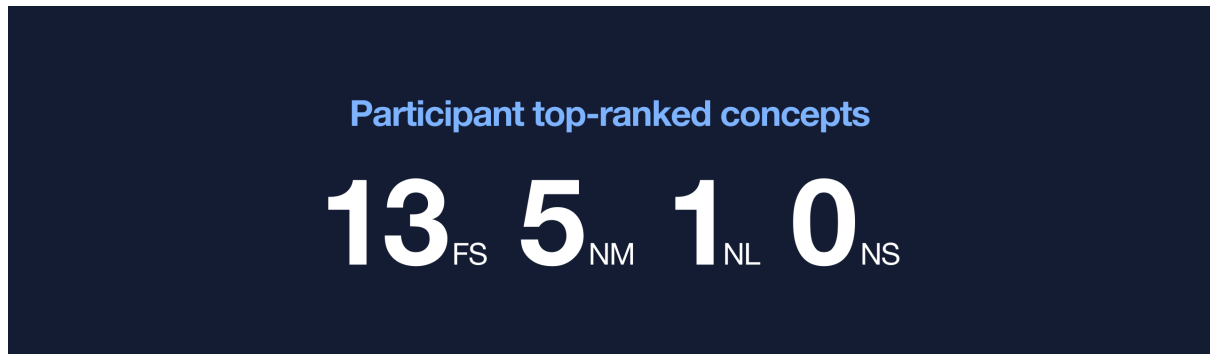


Figure 4.10: Number of top-rankings per concept.

When asked which interface structure, flat or nested, felt safest, responses were nearly evenly split (see Figure 4.11). Ten participants selected the flat structure, while nine chose the nested alternative. This indicates that although FS was the most preferred concept overall, perceptions of structural safety were more evenly distributed.



Figure 4.11: Participant responses to the question “Which interface structure felt safest?”.

In contrast, responses regarding button size revealed a clearer trend. The majority of participants (12 out of 19) identified the large (NL) buttons as the safest, followed by medium (NM) with 6 votes (see Figure 4.12). Only one participant selected the small (NS) buttons. This strong preference for larger touch targets highlights their perceived safety advantage, a finding that aligns with observed patterns in misclick frequency and subjective workload ratings.

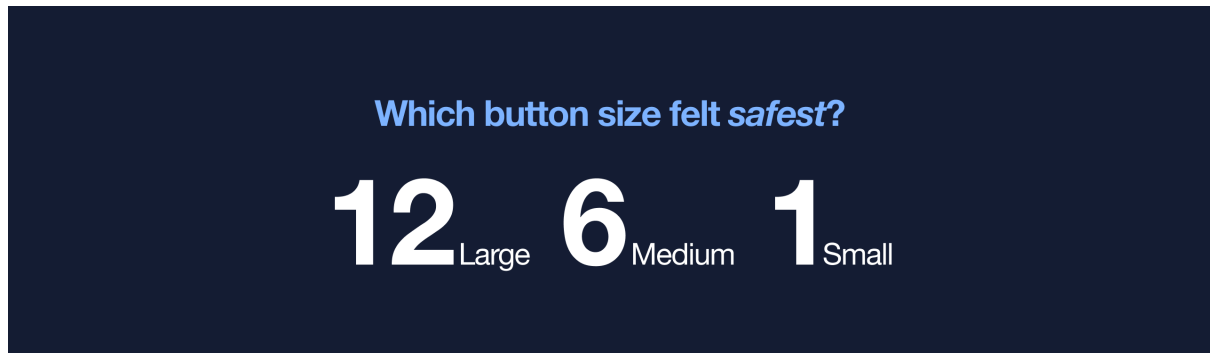


Figure 4.12: Participant responses to the question “Which button size felt safest?”

Correlation Between Scales

The repeated measures correlation analyses between the subjective scales showed that both pairings were positively and significantly correlated, with p values below .001 and moderately strong effect sizes. This suggests that participants' ratings tended to shift in the same direction across interface concepts, as shown in Figures 4.13 and 4.14.



Figure 4.13: Repeated measures correlation between perceived safety and UMUX-Lite.

The strongest relationship was observed between perceived safety and UMUX-Lite ($r = 0.642$, $p = 5.594e-08$). A similarly strong correlation was found between perceived safety and NASA-TLX ($r = 0.608$, $p = 4.102e-07$). Since NASA-TLX scores were reverse-coded, this indicates that higher perceived safety was associated with lower perceived workload (i.e. higher NASA-TLX values).



Figure 4.14: Repeated measures correlation between perceived safety and NASA-TLX.

Together, these findings suggest that when an interface felt safe, it was also more likely to be perceived as usable and less mentally taxing. The consistency between measures supports the interpretation that these aspects of user experience are closely connected.

4.2 Qualitative Results

This section presents the qualitative results from the user test. Instead of reviewing each concept individually, the findings are organised around key themes that emerged from the affinity diagramming of participant feedback. These themes, including visual demand, cognitive load, physical interaction, and other subjective or contextual influences, offer insight into the factors that shape drivers' perceptions of safety when interacting with in-vehicle touch interfaces. The analysis examines how these factors influenced participants' views on the tested concepts and their overall sense of safety.

Maintaining Vehicle Control

Participants' sense of safety was closely tied to whether interacting with the system could lead to a loss of vehicle control or other risky driving situations. They indicated that any interaction perceived to result in outcomes such as lane deviations or swerving would reduce their confidence in the system's safety. One participant described how interaction with the system affected their driving:

“When I looked back up from the screen and had moved off centre, it felt unsafe.”

(TP19)

This example shows how noticing a driving error after interacting with the interface can influence a user's perception of safety. In this case, the unintended lane drift suggests that any system requiring significant attention, to the point of affecting vehicle control, was likely to be seen as unsafe.

Eyes-Off-Road Time

A recurring theme in participants' accounts was that a “safer” interface is one that keeps TEORT to a minimum. Several participants described how spending more time looking away from the road made them feel less safe and more distracted.

“Looking away from the road too much feels dangerous.” (TP11)

“It felt a bit unsafe and uncomfortable, I had to look away a lot to find things.” (TP9)

These accounts suggest that extended periods of looking away from the road contribute to discomfort and a reduced sense of control. When an interface demands prolonged TEORT, it interferes with drivers' ability to monitor their surroundings, which in turn undermines their sense of safety.

Glanceability

The notion of “glanceability” emerged as a practical way to address the TEORT dilemma. Participants described displays that allowed status information to be understood in a single glance as contributing to a greater sense of safety.

“It feels safer when it's more glanceable while looking at the road.” (TP14)

This view suggests that visual clarity, achieved through high contrast, legible typography and clear status indicators, helps drivers access necessary information while maintaining situation awareness. These accounts indicate that immediate visual understanding, rather than just clean and minimalistic design alone, plays an important role in supporting drivers' sense of safety.

Task Segmentation

Several participants expressed that in the flat designs, interrupting the search before completing the task was perceived as problematic. They described how looking back at the road before pressing the correct button often resulted in losing their orientation, forgetting the button's location, or having to restart the entire search. To avoid this, they tended to keep their eyes on the interface until the task was finished, which contributed to prolonged glance durations with the flat design.

In contrast, the ability to divide interactions into smaller, distinct steps, with natural pauses to check the road, was seen by many as a helpful way to manage attention and maintain a sense of safety. This approach allowed drivers to shift their attention between the interface and the road in a more controlled and manageable way, reducing feelings of distraction.

“It's easier to split the task up into different steps. Press once, look up, and then look down again.” (TP17)

This comment illustrates how task segmentation can make interactions feel less intrusive and better aligned with the constant attentional demands of driving. It enables users to remain engaged with both the driving task and the secondary interaction, without feeling pressured to maintain continuous focus on the screen.

Quick Task Completion

The efficiency of an interaction, particularly the ability to complete a task quickly and with fewer clicks, was often associated by participants with a greater sense of safety. Faster interactions were seen as reducing TEORT, which in turn contributed to feeling safer. In contrast, nested designs that required additional clicks were frequently viewed as inefficient and distracting. One participant summarised this perspective:

“Avoiding extra clicks feels safer.” (TP16)

Another participant pointed to specific aspects of the nested design that contributed to distraction, linking them directly to the number of required actions:

“The distracting part of the nested design is the number of clicks, opening and closing the modal.” (TP19)

These accounts suggest that from some users’ perspective, each additional action represents a potential point of distraction and increased cognitive effort. As a result, tasks that feel inefficient or unnecessarily prolonged are also likely to be perceived as less safe.

Distraction Cues

Participants frequently described specific system elements that diverted their focus from the search task as distracting, which they linked to reduced safety. This could be due to visual design choices, such as the use of colour for highlighting active settings in FS, for example.

“Even on NS, the small text was easier to read [compared to FS] because there weren’t as many distractions on the page. There is no blue [highlights] that takes attention away from the text.” (TP12)

The quote indicates that excessive visual emphasis can divert attention from critical search targets, thereby increasing mental effort and diminishing perceived safety.

Visual Angle Away From the Road

The perceived safety of an interaction was influenced not only by the duration of a glance but also by its direction and angle away from the road ahead. Participants described it as uncomfortable and more strenuous to look towards the far side of the screen. One participant articulated how this affected the interaction effort.

“Options more in the left side are easier to find because you need to move your eyes less off the road. Those on the far right are more difficult to find.” (TP19)

This sentiment was echoed by another participant who found interacting with distant elements to be a source of unease.

“Felt unsafe in general. It was less bothersome to change roof shading and wiper sensitivity, but more bothersome to change the three [settings] that were furthest away.” (TP13)

These observations suggest that interface layouts requiring significant lateral gaze shifts or physical repositioning may be perceived as particularly unsafe, potentially because they necessitate more substantial disengagement from the forward driving view.

Ease of Learning

Participants described feeling safer when the interface was easy to learn. Such interfaces required less cognitive effort while driving, allowing participants to operate them with minimal distraction. In contrast, when participants had difficulty remembering the location of controls or understanding how an interaction worked, their sense of safety decreased, as they felt distracted or uncertain while trying to operate the system.

“I was fumbling around to find the exterior light control since I could not remember where it was from before.” (TP15)

This example suggests a connection between learnability and perceived safety. When the layout or interaction is unfamiliar, drivers face additional memory demands that compete with the attention needed for driving. In contrast, intuitive and easily learned designs reduce cognitive load and allow drivers to stay focused on the primary task.

Visual Search Effort

Extensive visual search was viewed as both stressful and risky. When some participants confronted a more cluttered screen, such as in FS, they reported longer search times and heightened anxiety.

“Searching too much feels dangerous.” (TP11)

This comment illustrates how search effort translates into perceived risk. Interfaces that reduce simultaneous on-screen options, or that provide clear categorical groupings, could therefore foster a calmer interaction and a stronger sense of control.

Ease of Reach

The physical effort required to reach controls was described as a critical issue for some participants, directly impacting their sense of safety. Difficulty reaching targets, particularly on the far side of the screen, could necessitate shifts in posture or gaze that participants felt were distracting and unsafe. One participant noted that this forced them to physically reorient themselves in a way that felt hazardous.

“Somewhat dangerous, I feel like I lose my focus a bit. I shift my gaze and physically move far from where I'm sitting.” (TP10)

Another participant specifically connected this physical strain to the layout of the interface.

“For the ones in the 3rd column, I had to turn my shoulder more to see them.” (TP19)

These experiences suggest that poor ergonomics can have a direct and negative impact on perceived safety. Interfaces that force drivers into awkward or extensive

movements may be judged as unsafe because the physical action itself is a source of distraction and compromised driving posture.

Movement Minimisation

Participants expressed preferences for interactions requiring minimal physical movement, suggesting that reduced finger travel distance could contribute to both comfort and perceived safety. This preference appeared particularly relevant when comparing different button sizes. One participant explicitly connected movement distance to safety perceptions:

"NM felt safer because of the reach issue mainly. When bigger, you need to travel longer distances with your finger" (TP17)

This observation implies that larger interface elements, such as in NL, while potentially easier to see and hit, may introduce safety concerns if they require extensive physical movements to navigate. The trade-off between target size and required movement distance suggests that optimal interface design must balance multiple physical interaction factors to maximise perceived safety.

Effort to Hit Targets

One of the strongest points of agreement among all participants was that the difficulty of accurately hitting a touch target directly negatively impacted perceived safety. The small button sizes required more concentration, increased the likelihood of making an error, and heightened the driver's anxiety. The fear of making a mistake was a significant concern:

"I'm worried about pressing wrong. I'm afraid of having to do it twice." (TP16)

This anxiety stems from the knowledge that an error necessitates a second attempt, effectively doubling the interaction time and the associated visual distraction. The cognitive load of aiming for a small target while simultaneously managing the vehicle created a sense of being overwhelmed, further degrading perceived safety. As the same participant explained:

"I had difficulty hitting the right function. I felt uncertain since keeping track of the speed, the road, and then choosing right on the screen. It became a lot at once."

(TP16)

This indicates that from the user's perspective, small touch targets are a significant usability and safety issue, as they demand more visual focus and increase the likelihood of miss-hits, which in turn necessitates repeated, distracting interactions.

Personal Preferences & Trade-offs

Aesthetic impressions occasionally overrode technical safety merits. While large icons enhanced usability, some found them visually off-putting.

"NL felt very safe, but almost a little childish." (TP18)

This tension illustrates that users negotiate between functional reassurance and stylistic comfort, suggesting that safety gains must be delivered alongside an appearance perceived as mature and credible.

Previous Experience

Prior exposure to other in-vehicle HMI systems set the baseline for evaluation. Familiar reference points coloured how "innovative" or "safe" the new prototypes felt.

"All concepts are better than the current one in [my car] anyway." (TP9)

This remark indicates that comparative mental models guide acceptance: if a new design feels at least as intuitive as an existing system, its safety is more readily assumed.

Preference for Physical Controls

For some participants, there appeared to be an underlying bias against touchscreens for in-vehicle controls. For these participants, touchscreens were viewed with scepticism irrespective of layout quality.

"Having a touchscreen that takes focus is not entirely safe. Compared to a knob, this takes a lot more attention." (TP9)

Such comments hint at a deeper modality bias: tactile hardware promises non-visual verification, and its absence can diminish trust in a purely digital solution.

Environmental Factors

Practical, real-world concerns such as screen glare and ambient lighting conditions were cited by participants as factors that affected the usability of the interface. It was noted that such conditions could significantly impair the visibility of crucial elements like modal edges, text, and on-screen controls.

“Unclear if it [the glare] impacted driving, but it would probably have felt safer with something clearer [like more contrast or a brighter interface].” (TP9)

This comment suggests that visibility issues caused by environmental factors can degrade usability and, consequently, perceived safety. The inability to see the interface increases frustration and the cognitive effort required for an interaction, which appears to undermine the user’s sense of safety.

Real-World Extrapolation

When assessing safety, participants often reported that they were mentally extrapolating how the system would perform in more demanding, real-world driving scenarios. Their evaluations were not limited to the controlled test environment but were based on anticipated performance in challenging situations like busy traffic or on winding roads.

“Slightly less safe because there are small boxes you must hit. If the road is winding, then perhaps it feels less safe.” (TP18)

Another participant considered the implications of even brief distractions in a more complex environment.

“If I were on a busy road, those split seconds could be dangerous.” (TP17)

These statements indicate that users’ mental models of how the system will function under pressure are a key component of their safety assessments. A system's

perceived safety is therefore heavily dependent on its perceived robustness in the context of realistic and challenging driving conditions.

5 Synthesis & Interpretation

This chapter discusses the implications of the results for in-vehicle touch interface design. The findings are interpreted in light of perceived safety, usability, and user behaviour. Connections to existing research are made to contextualise the outcomes and support conclusions about interface structure and user experience.

5.1 Designing Quick Controls

The findings from this study carry significant implications for the design of CSD Quick Controls, particularly concerning the trade-offs between interface structure (flat or nested) as seen in Figure 5.1 and the sizing of interactive elements. This chapter dissects these implications, bridging the empirical results to actionable insights for HMI designers.

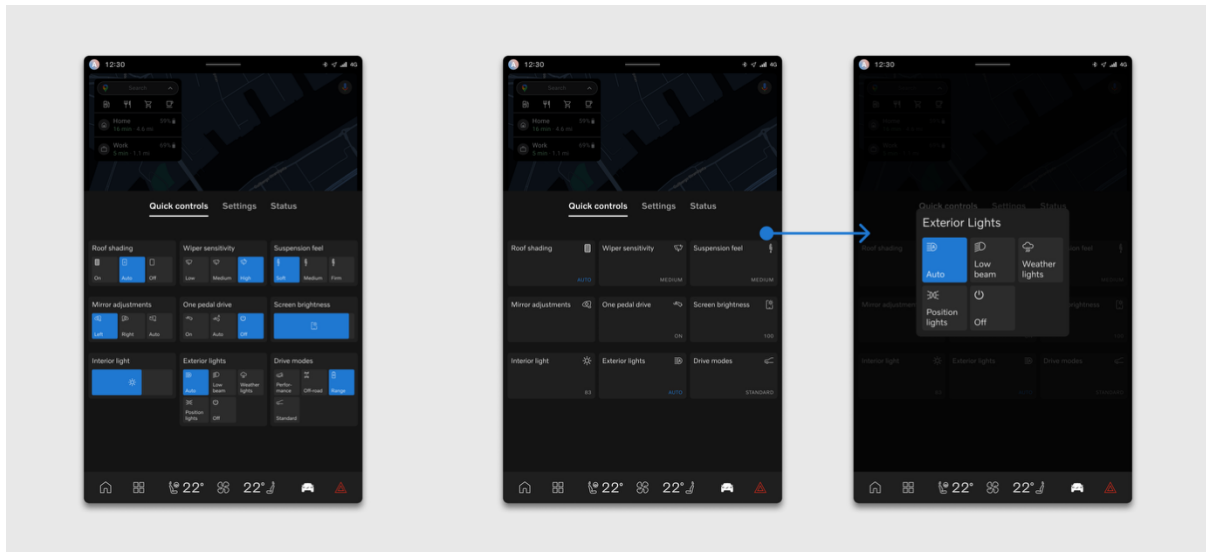


Figure 5.1: Flat or nested.

5.1.1 Flat or Nested: Navigating the Trade-offs

The choice between a flat HMI structure, which presents all controls on a single screen, and a nested structure, which organises controls into categorised modals, is a critical design decision that affects driver interaction, objective safety outcomes, and subjective user perceptions.

Flat Structure

The study revealed a considerable user preference for the immediacy of flat structures, as exemplified by FS, indicated by 13 out of 19 participants choosing it as their preferred choice. This preference is largely driven by the desire for rapid task completion and minimised interaction steps (taps), which many participants mentioned in interviews. This aligns with the objectively shorter TCT recorded for FS,

a finding consistent with research indicating faster search-based tasks in structured, broad layouts (Nicolas et al., 2024). Indeed, the high overall preference and top "safest" ranking bestowed upon FS by a significant portion of participants suggest users often equate this directness and interaction speed with efficiency and, consequently, safety. This perception of having all the control options visible and accessible at the same time fosters a sense of control for some users and was associated with shorter TEORT.

However, this user's inclination towards flat structures is not without hidden costs, particularly concerning glance behaviour and potential distraction. Despite its popularity and efficiency, FS was associated with significantly higher average glance durations, reflected in that 2 out of 6 tasks failed to meet the second NHTSA glance guideline, compared to no failed tasks for any of the nested concepts. This critical finding highlights a potential dichotomy between subjectively perceived safety and objectively riskier glance patterns. While participants described the flat interface as feeling faster and more direct because of the shorter TCTs, it may inadvertently encourage prolonged visual attention away from the roadway, a concern supported by studies linking the number of menu items to longer glance times (Feng et al., 2018). This specific divergence between subjective user experience and objective safety metrics in HMI design is critical. As noted by Li et al. (2020), the broader principle that individuals' perception of risk can differ significantly from objective measures underscores why such potential discrepancies demand careful consideration from HMI designers.

Moreover, qualitative feedback also indicated that participants perceived FS as initially cluttered and experienced a need to expend greater visual search effort. While familiarity over time might mitigate these issues, the initial experience can be overwhelming. As Burnett et al. (2013) demonstrated, the efficiency of broad menu structures heavily depends on users' ability to anticipate item locations, a characteristic of well-structured menus. For broad flat interfaces, such as FS, an absence of a clear, rapidly learnable structure could lead to excessive visual demand, similar to their findings for broad unstructured menus. Thus, successful flat designs

necessitate meticulous attention to visual hierarchy, logical grouping of functions, and ensuring the structure becomes predictable with minimal use.

Nested Structure

Conversely, nested structures demonstrated clear advantages, particularly in supporting safer glance strategies. The average glance duration results from the user tests indicated that the nested concepts encouraged a pattern of more frequent but shorter glances toward the CSD compared to the flat concept. This behaviour aligns with the research on driver attention, which suggests that drivers naturally redirect their focus to the road at subtask boundaries (J. Y. Lee & Lee, 2019). A nested design inherently introduces such boundaries. For example, after selecting a menu category, a driver is likely to glance back at the road to assess safety before proceeding to the next step of the interaction. By breaking a complex search into a sequence of simpler, clearly separated choices, the nested structure promotes this safer, more interruptible glance pattern. The nested design also reduces perceived visual clutter by presenting a limited and clearly defined set of options in the initial category screen and within modal windows. This reduction in visual clutter directly contributes to shorter average glance durations for each of the two subtasks (Yoon et al., 2015).

This effect was also reflected in participant feedback. Several participants described taking a glance at the road after selecting the initial function, before continuing the search for the correct setting in the modal. Dividing the interaction into smaller steps was perceived as a helpful way to manage attention and maintain a sense of safety. It allowed for more controlled shifts of focus between the interface and the road. By encouraging more frequent re-engagement with the road, this glance pattern also helps drivers avoid the prolonged off-road glances that have been linked to increased accident risk (Bärgman et al., 2015; Klauer et al., 2006).

Despite these objective benefits, the primary drawback identified for nested structures was the so-called "tap penalty," referring to the additional screen taps required. This was a consistent source of frustration, with many participants describing the extra interaction steps as inefficient, burdensome, and distracting. This perception contributed to a lower overall preference for the nested concepts when

compared directly with the flat concept, even when the nested designs featured larger buttons. For a nested approach to be justified, its benefits, such as reduced initial perceived visual complexity and the potential for larger targets within modals, must outweigh both the perceived and actual costs of the additional interaction steps. If categorisation is not intuitive, leading users to frequently navigate to incorrect categories, the resulting frustration and increased task time can undermine the safety benefits associated with glance segmentation.

5.1.2 The Impact of Button Size in Nested Interfaces

Within the evaluated nested concepts, the physical size of the final option buttons within the modal influenced misclick rates, perceived physical effort, and, crucially, subjective assessments of safety and usability. This was evident even though the button size variations between the three nested concepts did not yield statistically significant differences in the objective distraction metrics.

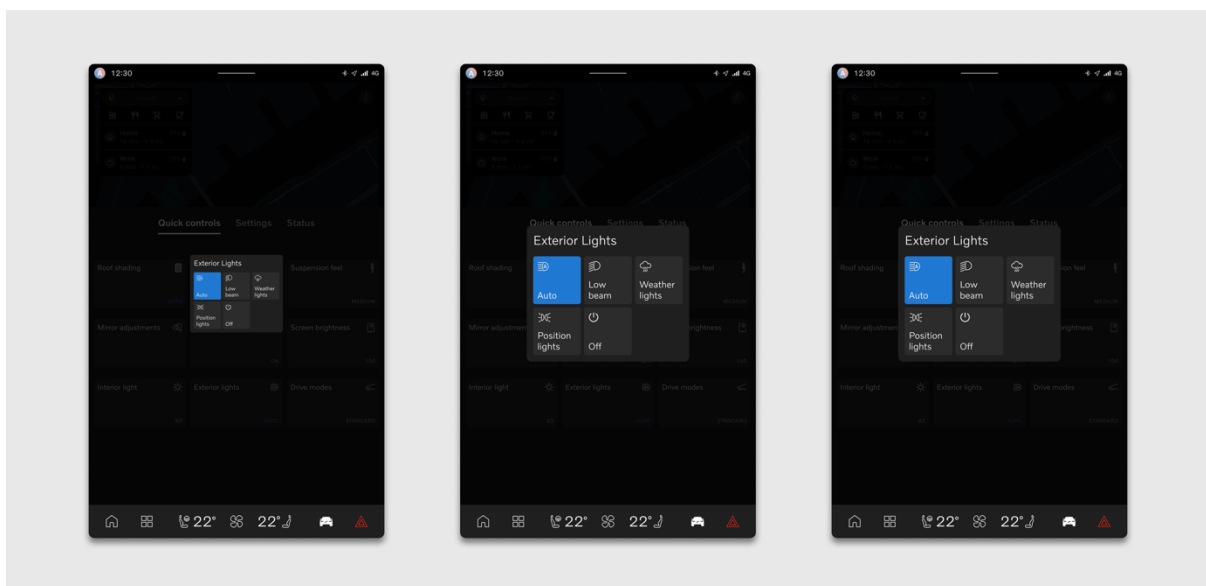


Figure 5.2: Modal button size comparison between NS, NM and NL.

Small Buttons (NS)

NS (20 x 20 mm buttons) was consistently the least preferred and was perceived as the least safe option. This perception was confirmed by the results from the perceived safety scale, where NS ranked lowest and was rated significantly less safe than NM.

Notably, this strong negative reception occurred despite the button sizes exceeding some of the minimum recommended dimensions for touchscreens in driving contexts (Kim et al., 2014; Tao et al., 2018) and despite no objective performance disadvantages compared to the other two nested concepts in terms of overall task metrics, apart from a higher percentage of misclicks. These findings emphasise the need for caution when specifying small touch targets for in-vehicle interfaces.

Furthermore, NS received the most critical qualitative feedback regarding the difficulty participants faced in accurately pressing buttons and reading associated labels. The text height, a known factor affecting legibility (Crundall et al., 2016), likely contributed to the lower perceived safety of this concept. Participants frequently described the increased cognitive load and visual search effort needed to locate and select these smaller elements, which diminished their confidence and comfort during interaction. Aesthetic factors also appeared to influence perceptions of NS. Many participants commented on the ample unused space surrounding the small modal, which made the buttons seem "ridiculously small." Interestingly, this perception of poor space utilisation was more marked for NS than for FS in the qualitative feedback, despite both concepts having identical button and text sizes. This suggests that context and user expectations significantly influence aesthetic and usability judgments. These observations align with previous research indicating that perceived aesthetics are often closely linked to perceived usability ratings, even when objective performance differences are minimal (Lavie et al., 2011; Frison et al., 2019).

Large Buttons (NL)

In contrast, NL was frequently identified in both qualitative feedback and preference rankings as having the safest button size. This view was shared by 12 out of 19 participants. Participants described the buttons as being "impossible to miss" and offering high legibility, which could benefit older drivers or those with reduced fine motor control (Chen et al., 2013). These findings align with the intuitive assumption that larger buttons improve target acquisition while driving by reducing the precision required, which can result in fewer misclicks and greater confidence in successful first-time activation, as noted in the interviews.

A notable finding from the study was that participants reported feeling uncertain when using smaller button sizes, as they felt the need to visually track their finger to the target. In contrast, larger buttons gave them the confidence to look away from the interface sooner, without waiting for visual confirmation. However, the glance data did not show shorter average glance durations for larger buttons within the nested modals compared to the other two nested button sizes. Therefore, this specific anticipatory glance behaviour for NL could not be corroborated by the glance metrics in this study.

Despite the strong subjective safety preference for NL, the larger button size did not translate into statistically significant improvements in objective task metrics, compared to the other nested concepts. The only significant differences were that NL performed better than NS on the NASA-TLX and UMUX-Lite scores, and had a lower percentage of misclicks. When comparing usability and workload scores to NM, these differences were not statistically significant. On the perceived safety scale, NL received high ratings, but these did not differ significantly from those for NM or NS. One possible explanation for the gap between subjective safety preference for NL and the lack of objective performance benefit is that participants' perceived safety may be linked to reduced anxiety about misclicking, rather than to measurable reductions in task time or glance activity.

NL also presented practical and aesthetic challenges. Some participants described the large buttons as "childish" or "exaggeratedly big." Additionally, larger buttons occupy more screen space. In this specific design, their size made it more difficult to close the modal by tapping outside of it, and interacting with the buttons inside required greater hand movement. These drawbacks likely contributed to the finding that, when participants who preferred a nested interface for their own vehicle were asked to select a concept in a general preference ranking, all six participants who preferred a nested design chose NM over NL.

Medium Buttons (NM)

NM often emerged as a well-regarded compromise, particularly when practical application and aesthetics were considered alongside perceived safety. A key finding

from the perceived safety scale was that increasing button size from small to medium resulted in a statistically significant improvement in perceived safety. However, this benefit plateaued, as no further significant gain was observed when increasing button size from medium to large. This indicates that, in terms of perceived safety, NM achieved a level comparable to NL while being significantly better than NS. This plateau effect was also reflected in the qualitative interview data.

NM performed well in the objective measures within the nested group and received positive subjective feedback during the interviews, suggesting that it represented an optimal trade-off for many users. It appeared to offer a favourable balance between the improved targetability associated with larger buttons and the screen efficiency and aesthetic neatness of smaller elements.

Importantly, the strong preference for NM in the "choice for their car" scenario supports its suitability as a balanced solution. All six participants who preferred a nested interface for their own vehicle selected NM. While NL buttons were rated as subjectively "safest" by more participants when considered in isolation, NM was preferred for integration into a vehicle due to its ability to deliver high perceived safety (comparable to NL according to the perceived safety scale) without the aesthetic and ergonomic drawbacks of NL, such as increased hand movement and difficulty closing the modal.

Notably, this preference for NM occurred despite its buttons being considerably larger than some established recommendations for in-vehicle HMIs. This highlights the importance of factors such as perceived safety, legibility, and aesthetic balance in shaping user preference. The findings suggest that while improvements to objective safety metrics may plateau at smaller button sizes, perceived safety can be optimised differently, and practical considerations ultimately drive user preferences for in-vehicle HMI designs.

Overarching Implication for Button Size in Modals

A key takeaway from the study is the absence of statistically significant differences in overall TEORT, TCT, average glance duration, or average number of glances

between the three nested concepts. This indicates that increasing button size beyond recommended minimums within modals primarily benefits users by reducing interaction errors, lowering frustration and effort, and enhancing perceived safety and ease of use. These improvements, however, do not appear to significantly influence overall glance strategies or task completion times, which are largely shaped by the underlying decision to employ a nested structure.

5.1.3 Designing For Perceived Safety

This section distils the synthesis into concrete considerations and key takeaways.

Aligning Perceived and Objective Safety

The findings from this study highlight a central design dilemma in automotive HMI development: choosing between a flat or a nested navigation structure for secondary controls. The results do not indicate a universally superior solution but rather reveal a complex trade-off that led to a nearly even split in user preference. Nine participants preferred FS for its perceived safety advantages, while ten favoured the nested concepts. This division suggests that the most suitable choice depends on individual user psychology and priorities.

At the heart of this dilemma is a trade-off between interaction speed and attentional management. Participants who preferred the flat structure valued shorter TCTs and the sense of direct control that comes from having all options visible at once. In contrast, those who favoured a nested structure were willing to accept an extra interaction step, often referred to as the "tap penalty," in exchange for shorter average glances, better interruptibility, and the potential for larger, more easily targeted buttons within modals. The decision-making process for each user appears highly personal, shaped by factors such as risk tolerance, driving confidence, and sensitivity to visual complexity and cognitive load.

A key goal for any safety-critical HMI is to ensure that the interface perceived as safer by the driver is also supported by objective safety evidence. The findings of this study suggest a disconnect in this alignment for the flat design. While many users associated the faster interaction of the flat structure with greater efficiency and safety,

the objective glance data pointed to a different outcome. As declared, the flat concept was linked to longer average glance durations and multiple task failures relative to NHTSA's 2-second glance guideline. Users may feel safer because the task is completed more quickly, yet they may also engage in riskier glance behaviours to achieve that speed.

In contrast, the nested concepts demonstrated a closer alignment between perceived and objective safety. This navigation structure appears to support a safer glance strategy by introducing clear subtask boundaries (J. Y. Lee & Lee, 2019). By dividing the interaction into distinct steps, such as selecting a category followed by selecting a function, the interface creates natural pause points that encourage drivers to return their gaze to the road. This segmentation may foster a sense of controlled engagement and promote a safer glance pattern, characterised by more frequent, shorter glances.

From this perspective, the so-called "tap penalty" associated with nested designs can be reconsidered. Rather than viewing it as a usability flaw, it may be understood as a small interaction cost that contributes to reducing the safety risks associated with prolonged glances away from the road. For designers, this reframes the decision from a straightforward usability trade-off to a more deliberate safety consideration.

When to Favour a Nested Structure

The evidence suggests that the decision to implement a flat or nested structure is highly context-dependent. The benefits of a nested design are not universally applicable, and its "tap penalty" must be justified by clear advantages in managing attention and complexity. Based on the findings of this study, a nested approach can be considered a suitable and potentially safer option under the following conditions:

1. **High information density:** As the number of functions within a menu increases, a flat structure may become visually cluttered, which can lead to longer visual search times and increased cognitive load. If a flat design results in glance durations that consistently exceed the 2-second NHTSA

threshold, the cognitive and visual search demands may outweigh the modest interaction cost of a nested system. In such situations, a well-organised nested structure is likely to provide a safer alternative.

2. **User sensitivity to visual clutter:** For user groups that may be more sensitive to visual complexity, such as older adults (S. C. Lee et al., 2019), the simplified initial screen of a nested design can help reduce the risk of users feeling overwhelmed. Several participants in this study expressed such concerns when first interacting with the flat design, suggesting that a nested structure may better support these groups.
3. **Accessibility for specific user groups:** For drivers with motor impairments or those who lack confidence in their fine motor control while driving, a nested design may offer benefits. By reducing the number of visible options at any one time, it can shorten the initial search process and free up screen space within modals for larger, easier-to-press targets. This may contribute to improved usability and perceived safety.
4. **Prioritising a conservative glance strategy:** Where the design goal is to minimise the likelihood of prolonged glances away from the road, a nested structure appears to offer advantages. From a safety policy perspective that emphasises strict adherence to NHTSA guidelines, accepting a slight increase in total task time in exchange for a safer glance pattern, characterised by shorter, more frequent glances, may be an appropriate compromise. It is important to note, however, that this consideration applies specifically to secondary tasks performed while driving. For interactions carried out when the vehicle is stationary, glance-related safety concerns are not relevant, and the superior speed and efficiency of a flat design may be preferable.

When to Favour a Flat Structure

Although the flat structure was associated with certain objective risks in this study, it may still be a viable categorize option under specific conditions where its potential drawbacks can be effectively mitigated:

- **Lower-density interfaces:** For menus containing relatively few functions, a flat layout could be an effective choice. With a limited number of items, visual search time is reduced to a level where the risk of prolonged glances becomes minimal, allowing the benefits of direct access to outweigh the disadvantages.
- **High learnability and frequent use:** If it is expected that users will interact with the menu frequently enough to quickly learn and memorise the spatial location of the functions, the challenge of visual search may be significantly reduced. Once users become familiar with the layout, they can rely on spatial memory rather than active searching, making the flat structure efficient without the same risk of extended glance durations.

If a designer chooses to implement a flat structure for use while driving, it is important to apply clear organisational principles, logical grouping, and high visual contrast to reduce visual clutter and make the interface as easy to search as possible. Most importantly, such a design should be thoroughly tested to ensure it does not encourage glance durations that exceed recommended safety thresholds. If it does not meet these safety requirements, a nested structure should be considered as an alternative.

Prioritising Attentional Management

Overall, the findings of this study indicate that while some users may be initially drawn to the perceived speed and simplicity of a flat interface, a well-designed nested structure may offer a more considered approach to supporting driver attention in most non-trivial use cases while the vehicle is in motion. Opting for a nested structure represents a deliberate design choice that places emphasis on managing driver attention, even if it involves a modest increase in interaction steps. For designers

working within a safety-critical context, accepting this small interaction cost may be an appropriate trade-off to help reduce the likelihood of prolonged glances away from the road.

The Role of Button Size

While the structural design of an interface affects interaction flow and glance behaviour, the physical size of touch targets plays a particularly important role in shaping the user's immediate experience. The results of this study indicate that button size directly influences perceived safety, interaction confidence, and workload, even when overall task completion times remain largely unaffected.

A key insight is that simply meeting the minimum size recommendations found in existing literature does not guarantee a positive user experience. The small buttons (NS), for example, exceeded several established minimums but were consistently rated as the least preferred and least safe option. This was reflected both in subjective feedback and objective performance data. The small buttons resulted in a higher rate of misclicks and increased user frustration. Participants reported difficulty reading labels and accurately selecting targets, while the excessive unused space in the modal made the buttons appear disproportionately small, diminishing the overall aesthetic impression and perceived usability.

These findings demonstrate that users do not evaluate button size in isolation. A design may comply with formal size recommendations yet still fail if it feels awkward, leads to errors, or appears visually unbalanced within the interface. To avoid such shortcomings, designers should aim for optimal rather than merely acceptable dimensions.

The study identified a clear "sweet spot" for button size. Increasing the size from small (NS) to medium (NM) buttons led to the most significant improvement in perceived safety across all size comparisons. However, increasing the size further, from medium (NM) to large (NL) buttons, produced no additional statistically significant gains in perceived safety. Instead, some participants found the large buttons visually unappealing or described them as "childish." Others noted

ergonomic drawbacks, such as increased hand movement and difficulty closing the modal.

The medium buttons (NM) ultimately provided the most favourable trade-off between usability, visual comfort, and spatial efficiency. Among participants who preferred a nested interface design, all selected the medium buttons for use in their own vehicles. This option successfully balanced ease of target selection, high perceived safety, and an aesthetically pleasing layout without introducing the ergonomic or visual disadvantages associated with excessively large buttons.

Based on these findings, a button size of approximately 35x35 mm is recommended for primary interactive elements within in-vehicle modals. This dimension optimises usability and user confidence while preserving space efficiency and visual balance. Importantly, button size in this context does not primarily affect task completion time but plays a critical role in reducing user uncertainty. By enabling quick and accurate selections without prolonged visual confirmation, appropriately sized buttons help preserve cognitive and visual resources for the primary task of driving.

5.2 Evaluation Methods

One central methodological finding of this study is the limited ability of standardised subjective scales, specifically the perceived safety scale, NASA-TLX, and UMUX-Lite, to statistically differentiate between the four evaluated concepts. This lack of quantitative distinction stands in stark contrast to the clear and decisive preferences expressed in the qualitative data, and this was further corroborated by objective eye-tracking data, which detected significant variations in glance behaviour between the designs.

Most notably, the scales found no significant safety difference between FS and NS, yet participant rankings identified FS as the safest concept by a strong majority (12 out of 19) and NS as the least safe by an overwhelming majority (16 out of 18). While the perceived safety scale did identify one isolated difference, rating NM as significantly safer than NS, it failed to capture the much larger and more critical distinction between the designs users perceived as best and worst.

This discrepancy between the undifferentiated quantitative results and the rich, nuanced qualitative feedback is the central problem this chapter addresses. The following sections will explore the potential reasons for this gap and discuss the resulting implications for evaluating perceived safety in future HMI design.

5.2.1 Explaining the Lack of Differentiation

A significant challenge in this study was selecting and applying measurement instruments capable of detecting subtle perceptual differences between the HMI concepts. While established scales like the NASA-TLX, UMUX-Lite, and the perceived safety scale from CTAM are well-validated, their design and intended application context may limit their sensitivity to capture the nuanced variations explored in this research.

The perceived safety scale, for example, was originally developed by Osswald et al. to predict a user's intention to adopt new in-vehicle technologies. In contrast, this study did not introduce a new technology but rather compared design variations within an established system (the CSD). The scale's focus on broad adoption decisions may make it less effective at capturing finer experiential differences between concepts that do not fundamentally alter a user's overall safety perception. This limitation is potentially amplified by the use of the perceived safety items in isolation, separate from the full CTAM framework, which holistically assesses user acceptance. Using these items for a comparative evaluation between similar concepts departs from their intended use for evaluating a single, complete system.

Similarly, the broad applicability of the NASA-TLX and UMUX-Lite scales, while valuable for general assessments of workload and usability, can be a drawback when attempting to discern minor variations between closely related prototypes. The general nature of these instruments may make them less capable of capturing the specific interaction nuances that differentiate the HMI concepts, especially when effect sizes are modest. When judged against the high-level criteria of these scales, the design differences between the evaluated concepts may not have been distinct enough to register significant variance.

Moreover, interacting with any in-vehicle touchscreen while driving is an inherently demanding dual task (Wickens, 2018). Participants must manage vehicle operation and maintain situational awareness while simultaneously engaging with the interface. When evaluated against this complex and safety-critical backdrop, the specific interaction differences between the HMI concepts may have appeared relatively minor relative to the questions in the scales. For instance, a participant might find smaller buttons in one concept frustrating but may not rate the system as fundamentally "dangerous" or "unsafe" on a general safety scale if it did not directly lead to a perceived loss of control or a hazardous situation within the controlled test environment. Consequently, while qualitative feedback may have highlighted specific usability issues, these nuances may not have been substantial enough to influence the broader quantitative ratings of safety and workload as measured by the selected instruments.

5.2.2 Predictive Relationships

While the primary analysis showed that the NASA-TLX and UMUX-Lite could not statistically differentiate between the HMI concepts, a secondary correlation analysis revealed their significant value in another capacity: they are strong indicators of perceived safety.

This analysis demonstrated a clear, strong, and statistically significant association between the measures. Higher reported usability (high UMUX-Lite score) was strongly associated with higher perceived safety ($r = 0.642$, $p = 5.594e-08$), and lower perceived workload (high NASA-TLX score) was strongly associated with higher perceived safety ($r = 0.608$, $p = 4.102e-07$).

The strength of these correlations underscores a fundamental relationship between critical user experience dimensions. It suggests that usability and mental workload are not independent of safety perceptions; rather, they are deeply interconnected. An interface that is perceived as easy to use and not mentally taxing is also very likely to be perceived as safe. This consistency provides robust evidence that these constructs are closely linked in the user's experience.

5.2.3 Measuring Perceived Safety

The methodological findings of this study offer several considerations for evaluating HMI designs, particularly regarding perceived safety. The results suggest that discrepancies between quantitative scale scores and qualitative feedback may reveal limitations in relying solely on standardised measurement tools. These observations highlight the value of a more comprehensive evaluative approach. A central aspect of such an approach is recognising that perceived safety and objective safety are separate. Perceived safety cannot be assumed to accurately reflect measurable, objective safety, as individual differences and subjective factors play a significant role in shaping user perceptions. This distinction is important, as misalignment between the two may have practical consequences: for example, a system that feels safe but is objectively unsafe could encourage risky user behaviour, while a system that is objectively safe but feels unsafe might experience lower acceptance (Osswald et al., 2012). Although this study does not establish a direct causal relationship, the findings indicate that aligning perceived and objective safety should be a consideration in the evaluation process, where possible.

With regard to measuring perceived safety, the results highlight certain limitations associated with standardised scales. While these tools can be useful for capturing general attitudes toward a system or for comparing substantially different design concepts (such as physical buttons versus touchscreens), they may be less effective in detecting subtle distinctions between more closely related alternatives. One possible explanation is that broad or imprecise survey items might unintentionally reflect a user's general impression of the overall context rather than their response to specific design variations. For more nuanced comparisons, it may be beneficial to complement existing scales with tailored questions that address particular aspects believed to influence perceived safety, such as confidence, task segmentation, or perceived task completion speed.

The inclusion of qualitative data proved to be especially valuable in this context. In the present study, interview responses provided more detailed insights into participants' experiences, including which design aspects influenced their sense of safety and the perceived magnitude of those effects. While qualitative methods do

not offer the same statistical generalisability as quantitative tools, they can reveal patterns, underlying reasoning, and contextual factors that structured scales may overlook. This type of information can support more informed, fine-grained design decisions, especially when comparing alternatives with subtle differences.

A key conclusion from this study is that established usability and workload metrics can serve as indicators for perceived safety. The data shows a strong and statistically significant association where higher usability and lower workload consistently correspond with higher levels of perceived safety. This finding has direct practical implications, suggesting that designers can leverage these more common metrics to gain indirect, yet valuable, insights into perceived safety throughout the design process. While this study provides a strong foundation, future research should aim to validate and quantify this predictive relationship across a broader range of contexts.

6 Design & Evaluation Guidelines

This chapter brings together the study's findings into concrete design recommendations for in-vehicle interfaces. Based on insights from both theory and empirical results, it outlines key considerations for layout, information hierarchy, interaction behaviour, and perceived safety. It also highlights approaches for measuring and evaluating perceived safety.

6.1 Design Guidelines

These guidelines are based on user testing and are intended to support HMI designers in balancing objective safety measures with the user's subjective experience when developing in-vehicle quick controls. The primary goal is to align the user's sense of safety with verifiable safety outcomes.

6.1.1 Flat or Nested

The choice between a flat and nested structure reflects a key trade-off between interaction speed and attentional management. This decision should be made deliberately, considering the design's main objective and the specific use context.

Prioritise Flat Structures for Stationary Interactions

When the vehicle is stationary, the risks associated with long glances are minimal. In this context, the direct access and speed of flat designs are preferable, as the "click penalty" of a nested structure provides little added benefit.

Favour Nested Structures as the Safer Default for In-Motion Tasks

Although users may prefer the speed of flat interfaces, nested designs can encourage safer glance behaviour by dividing tasks into smaller, more manageable steps. The additional tap should be viewed not as a drawback but as a minor interaction cost that helps reduce prolonged glances away from the road.

Choose Nested Structure if the Primary Goal Is to Minimise Long Single Glances

This study found that nested designs, across all tested button sizes, resulted in fewer long glances (>2.0 seconds) compared to flat designs. If the main objective is to keep single glance durations below recommended safety thresholds (e.g., NHTSA's 2-second guideline), a nested structure offers a more reliable solution.

Opt for Flat Structures to Minimise Task Completion Time and Eyes-Off-Road Time

In this study, flat designs generally resulted in shorter task completion times (TCT) and total eyes-off-road times (TEORT), suggesting greater efficiency when overall time-based performance metrics are prioritised.

Consider Nested Structures for Unstructured or Infrequently Used Menus

For menus containing many items or for functions used infrequently, nested designs help reduce initial visual clutter and cognitive load, making the interface feel less overwhelming.

Favour Flat Structures When Controls Are Logically Organised and Locations Can Be Quickly Learned

Research suggests that well-organised, broad (flat) menus can be highly effective once users have learned item locations. In cases where functions are used frequently and item positions can be memorised, flat structures may offer both speed and efficiency while mitigating the risks of prolonged visual search identified in this study.

Consider Nested Structures for Users Sensitive to Visual Clutter or With Specific Visual or Motor Needs

Users who prefer reduced visual clutter or who benefit from larger targets may find nested designs better suited to their needs and perceptions of safety.

Favour Flat Structures for Users Who Prioritise Interaction Speed and Direct Access

If minimising clicks and achieving the fastest possible interaction is the user's primary concern—and they are comfortable managing a denser layout—a flat structure may be preferable.

6.1.2 Designing Nested Interfaces

Within a nested structure, button size and layout play a critical role in shaping user confidence and perceived safety, even if they do not significantly affect overall task completion times.

Ensure Modal Buttons Are Appropriately Sized for the Available Space

Buttons that appear "contextually too small" or fail to utilise the available space effectively can reduce perceived safety and lead to negative aesthetic evaluations.

Expect Perceived Safety to Increase When Button Size Grows From Small to Medium

Larger targets in this size range (e.g., 20 × 20 mm to 35 × 35 mm) can reduce the likelihood of misclicks and improve confidence in accurate selection, contributing to greater feelings of safety.

Recognise That Perceived Safety Benefits Plateau Beyond a Medium Button Size

This study found no statistically significant improvement in perceived safety when button size increased from medium (35 mm) to large (50 mm).

Avoid Overly Large, "Cartoonish," or "Childish" Button Designs

Excessively large buttons may carry negative associations (e.g., with interfaces intended for older adults or assistive devices) and can lead to user aversion.

6.1.3 Designing Flat Interfaces

If a flat structure is selected for an in-motion interface, its potential to induce prolonged glances must be actively mitigated through careful design.

Prioritise Learnable Structure and Clear Organisation

Users must be able to quickly locate items in flat interfaces. Logical grouping, consistent placement, and clear visual hierarchy help make the layout predictable and reduce visual search time.

Use Colour Purposefully and Sparingly

Colour should be employed primarily for indicating hierarchy and system state, rather than simply for visual differentiation. Excessive or varied use of colour increases visual clutter and can distract from critical information.

Manage Information Density to Support Safe Glance Times

An excessive number of items increases glance durations. Design testing should ensure glance times remain within safety standards (e.g., under 2 seconds). If the interface becomes overloaded, consider relocating less frequently used items to submenus or switching to a nested structure.

Ensure High Text Legibility

Text size, contrast, and font selection should be optimised for quick readability. Poor legibility increases both glance duration and cognitive load.

6.2 Evaluation Guidelines

The following guidelines outline recommended approaches for evaluating perceived safety in HMI design. They emphasise the importance of using a combination of qualitative insights, standardised measures, and contextual testing to ensure both objective and perceived safety are thoroughly assessed.

Prioritise Qualitative Insights

Methods such as interviews and observations reveal important nuances in perceived safety that standardised scales, such as the CTAM-PS, often overlook. Verbal feedback can provide richer, more detailed insights than abstract numerical ratings.

Use Standardised Scales Critically

Validated perceived-safety scales, such as the CTAM-PS, are useful for broad comparisons, for example between different interface types or when evaluating a single design. However, their general wording can miss subtle design differences and tends to capture users' overall impressions rather than specific details. As such, they are less suitable for comparing closely related design variants.

Develop Specific, Contextual Questions

Design evaluation questions that target momentary safety perceptions during specific tasks (e.g., "How confident did you feel hitting that button?"), rather than relying solely on generalised safety statements.

Measure Both Objective and Perceived Safety

Objective safety should be assessed using relevant safety metrics, rather than relying on perceived safety as a proxy. The goal is to seek alignment between objective and perceived safety. A system that feels safe but is objectively unsafe can encourage risky behaviour, while one that is objectively safe but feels unsafe may suffer from poor user acceptance. Both constructs should therefore be explicitly addressed during evaluation.

Utilise Comparative Ranking for Direct Preference Judgements

Asking users to rank concepts based on perceived safety or overall preference often produces clearer, more meaningful comparisons than relying on absolute scale ratings, particularly when evaluating subtle design differences.

Leverage Usability and Workload as Predictive Indicators of Perceived Safety

Usability and workload are not merely supplementary metrics; they are closely linked to perceived safety. Interfaces with high usability (e.g., high UMUX-Lite scores) and low mental workload (e.g., low NASA-TLX scores) are more likely to be perceived as safe. These established metrics can be used as reliable, indirect indicators of perceived safety throughout the design process.

7 Discussion

This chapter reflects on the project's aims, results, and methodology. It considers how specific aspects of the process and test design may have influenced the findings and outlines key ethical considerations and suggestions for future research.

7.1 Aim & Objectives Revisited

This thesis aimed to explore whether perceived safety can be intentionally influenced through interface-level design choices in in-vehicle touch-based HMIs, and how this experience can be meaningfully measured. This aim was addressed by answering two research questions through a process involving literature reviews, expert interviews, interface design and prototyping, and a user test conducted in a real driving context.

- **RQ1:** *How do navigation structure and button size influence the perceived safety of in-vehicle touch interfaces?*

This question was examined by systematically varying button size across the three nested concepts, while keeping the flat concept (FS) constant. Within the nested group, larger buttons were generally associated with higher perceived safety ratings, suggesting that button size played a supportive role in more complex structures. However, the consistently strong preference for FS, despite its small buttons, indicates that structural simplicity and direct access can outweigh the perceived benefits of larger touch targets. The findings suggest that both structure and button size affect perceived safety, but their relative importance depends on the specific interaction context and user priorities.

- **RQ2:** *How suitable are existing evaluation methods for assessing and comparing the perceived safety of in-vehicle touch interface concepts?*

This question was addressed through a mixed-method approach that included subjective ratings, eye-tracking data, and participant interviews. While the perceived safety scale offered a general overview, it showed limited sensitivity in distinguishing between similar interface concepts. To complement this, the semi-structured interviews proved essential for uncovering user insights that the quantitative measures alone could not capture. NASA-TLX and UMUX-Lite scales revealed strong correlations between low workload, high usability, and higher perceived safety (i.e., higher scores), suggesting their potential to provide supporting insights into safety-related concerns. Eye-tracking further enriched the evaluation by revealing distinct

differences in glance patterns between concepts. It showed that participants often had glance durations exceeding recommended limits, which is objectively unsafe behaviour, even when they subjectively rated an interface as safe.

Together, these findings underscore the importance of combining subjective and behavioural measures to reliably assess perceived safety in realistic driving conditions.

7.2 Methodological Reflections

This section discusses how the chosen process, methods, and project constraints may have shaped the outcome. The focus is limited to selected aspects of the methodology that were particularly influential or are considered valuable to reflect on.

7.2.1 Pre-Study

Although the pre-study included both internal and external input, the final scope focusing on perceived safety concerning Quick Controls was primarily shaped by internal discussions. Most interviews were held with Volvo Cars employees, many of whom had long tenure at the company. These conversations guided the project toward areas deemed valuable or relevant to Volvo Cars, rather than emerging from a broader exploration of the automotive domain.

As a result, perspectives from other manufacturers, adjacent industries, or academic design contexts were only briefly considered. This may have constrained the range of initial directions and limited exposure to alternative ways of framing digital in-vehicle interaction.

While the literature review addressed in-vehicle HMI interaction broadly, it offered limited theoretical justification for focusing specifically on Quick Controls. The decision was ultimately driven more by internal priorities at Volvo Cars than by academic rationale.

7.2.2 User Test

The user test provided valuable insights into how interface layout can influence perceived safety, although several methodological factors help explain why the statistical analyses yielded limited significant results.

Test Design

The participant group was relatively diverse, but broader variation would have been beneficial. Due to budget constraints, external recruitment was not possible, and participants were therefore limited to Volvo Cars employees. The inclusion of eye-tracking further reduced the eligible pool, resulting in a sample composed predominantly of men. While no demographic comparisons were conducted, it is plausible that factors such as gender, age, or driving experience influenced participants' perceptions and behaviours.

Sample Size and Generalizability

With 19 participants, the study operated under practical constraints and had limited statistical power. This likely contributed to the absence of significant effects in several subjective measures.

Although the NHTSA guidelines are based on a sample size of 24 participants, only 18 participants were included in the objective dataset for this study. As a result, the threshold for defining a guideline violation had to be adjusted from four to three participants. This reduced margin increases the influence of individual variation and may limit the robustness of conclusions drawn from the compliance analysis.

Furthermore, while the within-subjects design enabled consistent comparisons across conditions, it may have introduced learning or ordering effects. Even with counterbalancing, repeated exposure to different layouts could have shaped participants' later evaluations. This risk was particularly relevant among the nested variants, all of which began with the same initial Quick Controls screen.

Participant-Related Variability

Individual participant variability is a common challenge in subjective user studies, as it can introduce noise that obscures the effects of specific design interventions. In the present study, this issue was likely amplified by the combination of a cognitively demanding test environment and natural differences among participants in risk perception and technical familiarity.

Observations of participants' rating behaviour support this interpretation. Many appeared to form an overall impression early in the study, reporting either a general sense of safety or consistent apprehension across all HMI concepts. The concept each participant encountered first often seemed to establish a reference point, with ratings for subsequent concepts showing only minor deviations from that initial evaluation. This suggests that individual factors, such as pre-existing attitudes toward in-vehicle touchscreens, general risk perception related to driving, or prior experience with vehicle HMIs, may have established a stable personal baseline for subjective evaluations. The observed rating patterns reflect high variability between participants and low variability within each participant's set of ratings.

It is therefore plausible that the subtle design differences between the HMI concepts were overshadowed by more dominant factors, such as the broader demands of the driving task itself.

Test Environment and Setup

Several limitations were present in the test environment. Since all driving took place on a straight and predictable track without other vehicles or external hazards, the setting likely reduced the salience of safety-related interface differences. In more complex driving scenarios, such as urban traffic or curved roads, interface design might have had a greater impact.

Visibility issues were also reported, particularly on the right side of the screen during sunny conditions. The anti-reflective film applied to the display proved insufficient in direct sunlight, and several participants noted difficulty seeing elements near the screen edges. Additionally, weather varied across test days, ranging from bright sun

to light rain, which may have subtly affected visual performance and interaction comfort.

The tasks used in the study were based on typical Quick Controls interactions and were generally short and procedural. While representative of real use cases, the low cognitive demand may have limited the extent to which layout differences influenced both perceived safety and objective performance.

7.3 Ethical Considerations

Evaluating perceived safety in in-vehicle HMIs, as well as the outcomes of this thesis, involves several ethical considerations. These primarily concern ensuring participant safety during testing and the risk of design recommendations being misinterpreted or applied outside their intended context.

7.3.1 Participant Safety During Testing

Conducting tests in a real vehicle enhanced ecological validity but also required participants to perform secondary tasks, temporarily diverting attention from driving. While all interactions were brief, carried out at moderate speeds, and supervised on a closed track, the intentional introduction of distraction remains an ethical concern. A driving simulator would have eliminated this risk, but may not have captured participants' sense of safety as effectively. Ultimately, the method choice represented a trade-off between ecological validity and participant well-being. To address this, a thorough procedural risk analysis was carried out before the first test session, and all participants were fully informed and gave their consent with an understanding of the test's conditions and associated risks.

7.3.2 Transparency in Design Interpretation

Another ethical consideration concerns how the findings and design recommendations might be interpreted. Perceived safety should not be conflated with actual, objective safety. Interfaces that feel "safe" may still encourage interaction patterns that compromise driver attention. The aim of this study was not to promote increased engagement in secondary tasks but to improve understanding of how

interface design influences user perception, so that trade-offs can be considered more consciously. The results must be communicated with care and not oversimplified in applied design work.

7.4 Recommendations for Future Work

While the study contributed meaningful insights into the influence of HMI design on perceived safety, further investigation is needed across several areas.

7.4.1 Refining Evaluation Methods

As previously noted, the perceived safety scale showed limited sensitivity when comparing relatively similar interface concepts. Future studies could explore its applicability in more disruptive system changes, such as alternative steering layouts, physical controls, or new display paradigms, where differences may be more pronounced.

There is also a need to develop or adapt tools that better capture subtle design differences. This might involve incorporating more focused questions that target specific aspects of the user experience, such as confidence *during* interaction or the effort required for visual search. Such refinements could support a more nuanced understanding of how perceived safety is shaped.

7.4.2 Extending Context and Conditions

Future research could explore how perceived safety is influenced under more demanding driving conditions or in high-fidelity simulators with increased visual and cognitive load. This would enable a more detailed examination of how interface layout interacts with external complexity, such as navigating intersections or responding to unexpected traffic events. Similar studies could also be conducted during partially automated driving, for example with Assisted Cruise Control or lane-keeping systems, where attention is shared between the driver and the vehicle. These contexts may provide deeper insight into when perceived safety becomes most relevant and how it shifts with changes in task workload and situational awareness.

7.4.3 Studying Longer-Term Use

This study focused on first-time use. However, perceived safety is likely to evolve as users become more familiar with the interface or encounter edge cases. A natural next step would be a short longitudinal study in which participants use the same interface over several weeks in real-world driving. This could reveal how preferences and expectations shift over time, and whether increased familiarity enhances or undermines the sense of safety. It may also clarify when layout complexity starts to affect learnability, particularly in flat designs with high information density.

7.4.4 Exploring Alternative Interaction Models

Finally, future work could look beyond visual-manual touch input. A Quick Controls menu that includes tactile or multimodal elements, such as physical feedback or voice control, may help reduce visual demand and increase user confidence. These approaches could be especially relevant in conditions where visual attention is limited, or for users with different abilities or preferences. It may also be valuable to explore hybrid systems that provide direct access to frequent settings while grouping others into a modal structure.

References

- Angell, L., Perez, M., & Garrott, W. R. (2013, November). *Explanatory material about the definition of a task used in NHTSA's driver distraction guidelines, and task examples* (Report No. DOT HS 811 858). National Highway Traffic Safety Administration.
- Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, 8, 456. <https://doi.org/10.3389/fpsyg.2017.00456>
- Bärgman, J., Lisovskaja, V., Victor, T., Flannagan, C., & Dozza, M. (2015). How does glance behavior influence crash and injury risk? A 'what-if' counterfactual simulation using crashes and near-crashes from SHRP2. *Transportation Research Part F: Traffic Psychology and Behaviour*, 35, 152–169. <https://doi.org/10.1016/j.trf.2015.10.011>
- Blascheck, T., Bentley, F., Choe, E. K., Horak, T., & Isenberg, P. (2021). Characterizing glanceable visualizations: From perception to behavior change. In *Mobile data visualization*. Chapman and Hall/CRC.
- Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review* (2nd ed.). SAGE.
- Brooks, J. L. (2012). Counterbalancing for serial order carryover effects in experimental condition orders. *Psychological Methods*, 17(4), 600–614. <https://doi.org/10.1037/a0029310>

- Burnett, G. E., Lawson, G., Donkor, R., & Kuriyagawa, Y. (2013). Menu hierarchies for in-vehicle user-interfaces: Modelling the depth vs. breadth trade-off. *Displays*, 34(4), 241–249. <https://doi.org/10.1016/j.displa.2013.07.001>
- Burnett, G. E., Summerskill, S. J., & Porter, J. M. (2004). On-the-move destination entry for vehicle navigation systems: Unsafe by any means? *Behaviour & Information Technology*, 23(4), 265–272. <https://doi.org/10.1080/01449290410001669950>
- Cao, J., Chu, H.-L., Chen, Y.-L., Chu, C.-H., Huang, Y.-Y., & Lee, Y.-J. (n.d.). Enhancing in-car interface efficiency: The influence of menu configuration on cognitive load and visuospatial memory. *Ergonomics*, 0(0), 1–17. <https://doi.org/10.1080/00140139.2025.2470951>
- Chen, K. B., Savage, A. B., Chourasia, A. O., Wiegmann, D. A., & Sesto, M. E. (2013). Touch screen performance by individuals with and without motor control disabilities. *Applied Ergonomics*, 44(2), 297–302. <https://doi.org/10.1016/j.apergo.2012.08.004>
- Cockburn, A., & Gutwin, C. (2009). A predictive model of human performance with scrolling and hierarchical lists. *Human–Computer Interaction*, 24(3), 273–314. <https://doi.org/10.1080/07370020902990402>
- Crundall, E., Large, D. R., & Burnett, G. (2016). A driving simulator study to explore the effects of text size on the visual demand of in-vehicle displays. *Displays*, 43, 23–29. <https://doi.org/10.1016/j.displa.2016.05.003>
- Ding, G.-J., Hwang, T. K. P., & Kuo, P.-C. (2020). Progressive disclosure options for improving choice overload on home screen. In T. Ahram & C. Falcão

- (Eds.), *Advances in usability, user experience, wearable and assistive technology* (pp. 173–180). Springer International Publishing. https://doi.org/10.1007/978-3-030-51828-8_23
- Dukic, T., Hanson, L., & Falkmer, T. (2006). Effect of drivers' age and push button locations on visual time off road, steering wheel deviation and safety perception. *Ergonomics*, *49*(1), 78–92. <https://doi.org/10.1080/00207540500422320>
- Dukic, T., Hanson, L., Holmqvist, K., & Wartenberg, C. (2005). Effect of button location on driver's visual behaviour and safety perception. *Ergonomics*, *48*(4), 399–410. <https://doi.org/10.1080/00140130400029092>
- Endsley, M. (2000). Situation awareness analysis and measurement, chapter theoretical underpinnings of situation awareness. *A Critical Review*, 3–33.
- Feng, F., Liu, Y., & Chen, Y. (2018). Effects of quantity and size of buttons of in-vehicle touch screen on drivers' eye glance behavior. *International Journal of Human-Computer Interaction*, *34*(12), 1105–1118. <https://doi.org/10.1080/10447318.2017.1415688>
- Fischer, L., Holder, D., Weiss, B., Kiessling, J., Reichelt, F., & Maier, T. (2022). Factors influencing the perception of safety in automated vehicles interiors. *Intelligent Human Systems Integration (IHSI 2022): Integrating People and Intelligent Systems*, *22*(22). <https://doi.org/10.54941/ahfe100944>

- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391. <https://doi.org/10.1037/h0055392>
- Frison, A.-K., Wintersberger, P., Riener, A., Schartmüller, C., Boyle, L. N., Miller, E., & Weigl, K. (2019). In UX we trust: Investigation of aesthetics and usability of driver-vehicle interfaces and their impact on the perception of automated driving. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3290605.3300374>
- Gáspár, P., Szalay, Z., & Aradi, S. (2014). *Highly automated vehicle systems* (Report No. 978-963-313-173-2). Department of Mechatronics, Optics, and Mechanical Engineering Informatics, Budapest University of Technology and Economics.
- Global Growth Insights. (2025, June 2). *Center stack display market size, share, growth trends, and forecast 2023–2033*. <https://www.globalgrowthinsights.com/market-reports/center-stack-display-market-107921>
- Guo, F., Chen, J., Li, M., Lyu, W., & Zhang, J. (2022). Effects of visual complexity on user search behavior and satisfaction: An eye-tracking study of mobile news apps. *Universal Access in the Information Society*, 21(4), 795–808. <https://doi.org/10.1007/s10209-021-00815-1>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 239–250). North-Holland.

Hartson, R., & Pyla, P. S. (2012). *The UX book: Process and guidelines for ensuring a quality user experience*. Morgan Kaufmann.

Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention*, 38(1), 185–191. <https://doi.org/10.1016/j.aap.2005.09.007>

Interaction Design Foundation. (n.d.). *What are the Gestalt principles?* Retrieved June 23, 2025, from <https://www.interaction-design.org/literature/topics/gestalt-principles?srsltid=AfmBOoq89Ju2ME5oLWAjyOAxuQfNpMt9RbaIS55WHZs4IkwUXE9wlaJm>

Interaction Design Foundation. (2025). *What is progressive disclosure?* Retrieved June 23, 2025, from <https://www.interaction-design.org/literature/topics/progressive-disclosure>

International Organization for Standardization. (2017). *Road vehicles – Ergonomic aspects of transport information and control systems – Dialogue management principles and compliance procedures* (ISO Standard No. 15005:2017). ISO.

International Organization for Standardization. (2018). *Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts* (ISO Standard No. 9241-11:2018).

- International Organization for Standardization. (2020). *Ergonomics of human-system interaction – Part 110: Interaction principles* (ISO Standard No. 9241-110:2020). ISO.
- Janssen, C. P., Brumby, D. P., & Garnett, R. (2012). Natural break points. *Human Factors*, 54(5), 715–728. <https://doi.org/10.1177/1555343411432339>
- Jeong, H., & Liu, Y. (2019). Effects of non-driving-related-task modality and road geometry on eye movements, lane-keeping performance, and workload while driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 60, 157–171. <https://doi.org/10.1016/j.trf.2018.10.015>
- Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, 4(1), 71–115. https://doi.org/10.1207/s15516709cog0401_4
- Kim, H., Kwon, S., Heo, J., Lee, H., & Chung, M. K. (2014). The effect of touch-key size on the usability of in-vehicle information systems and driving safety during simulated driving. *Applied Ergonomics*, 45(3), 379–388. <https://doi.org/10.1016/j.apergo.2013.05.006>
- Kim, J.-S. (2010). Within-Subjects Design. In *Encyclopedia of Research Design* (pp. 1639–1644). SAGE. <https://doi.org/10.4135/9781412961288>
- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-Car Naturalistic Driving Study data* (Report No. DOT HS 810 594). U.S. Department of Transportation, National Highway Traffic Safety Administration.

- Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine*, *370*(1), 54–59. <https://doi.org/10.1056/NEJMsa1204142>
- Krstajić, R., Žužić, A., & Orehovački, T. (2024). Safety aspects of in-vehicle infotainment systems: A systematic literature review from 2012 to 2023. *Electronics*, *13*(13), 2563. <https://doi.org/10.3390/electronics13132563>
- Kujala, T., & Saariluoma, P. (2011). Effects of menu structure and touch screen scrolling style on the variability of glance durations during in-vehicle visual search tasks. *Ergonomics*, *54*(8), 716–732. <https://doi.org/10.1080/00140139.2011.592601>
- Kujala, T., & Salvucci, D. D. (2015). Modeling visual sampling on in-car displays: The challenge of predicting safety-critical lapses of control. *International Journal of Human-Computer Studies*, *79*, 66–78. <https://doi.org/10.1016/j.ijhcs.2015.02.009>
- Kvale, S., & Brinkmann, S. (2015). *InterViews: Learning the craft of qualitative research interviewing* (3rd ed.). SAGE.
- Landauer, T. K., & Nachbar, D. W. (1985). Selection from alphabetic and numeric menu trees using a touch screen: Breadth, depth, and width. *SIGCHI Bulletin*, *16*(4), 73–78. <https://doi.org/10.1145/1165385.317470>
- Lavie, T., Oron-Gilad, T., & Meyer, J. (2011). Aesthetics and usability of in-vehicle navigation displays. *International Journal of Human-Computer Studies*, *69*(1), 80–99. <https://doi.org/10.1016/j.ijhcs.2010.10.002>

- Lee, J. D. (2014). Dynamics of driver distraction: The process of engaging and disengaging. *Annals of Advances in Automotive Medicine*, 58, 24–32.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Lee, J. Y., & Lee, J. D. (2019). Modeling microstructure of drivers' task switching behavior. *International Journal of Human-Computer Studies*, 125, 104–117. <https://doi.org/10.1016/j.ijhcs.2018.12.007>
- Lee, S. C., Hwangbo, H., & Ji, Y. G. (2016). Perceived visual complexity of in-vehicle information display and its effects on glance behavior and preferences. *International Journal of Human-Computer Interaction*, 32(8), 654–664. <https://doi.org/10.1080/10447318.2016.1184546>
- Lee, S. C., Kim, Y. W., & Ji, Y. G. (2019). Effects of visual complexity of in-vehicle information display: Age-related differences in visual search task in the driving context. *Applied Ergonomics*, 81, 102888. <https://doi.org/10.1016/j.apergo.2019.102888>
- Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE: When there's no time for the SUS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)* (pp. 2099–2102). Association for Computing Machinery. <https://doi.org/10.1145/2470654.2481287>
- Li, Z., Wang, C., Fu, R., Sun, Q., & Zhang, H. (2020). What is the difference between perceived and actual risk of distracted driving? A field study on a real

- highway. *PLOS ONE*, 15(4),
e0231151. <https://doi.org/10.1371/journal.pone.0231151>
- Liu, Y., Wang, X., Chen, L., Liu, S., Han, J., Shi, H., & Zhong, F. (2022). Driver's visual attention characteristics and their emotional influencing mechanism under different cognitive tasks. *International Journal of Environmental Research and Public Health*, 19(9),
5059. <https://doi.org/10.3390/ijerph19095059>
- Liu, Z., Yuan, W., & Ma, Y. (2021). Drivers' attention strategies before eyes-off-road in different traffic scenarios: Adaptation and anticipation. *International Journal of Environmental Research and Public Health*, 18(7),
3716. <https://doi.org/10.3390/ijerph18073716>
- Lowry, R. (1998). *Concepts and applications of inferential statistics* (Subchapter 12a: The Wilcoxon signed-rank test). <http://vassarstats.net/textbook/ch12a.html>
- Ma, J., Gong, Z., Tan, J., Zhang, Q., & Zuo, Y. (2020). Assessing the driving distraction effect of vehicle HMI displays using data mining techniques. *Transportation Research Part F: Traffic Psychology and Behaviour*, 69, 235–250. <https://doi.org/10.1016/j.trf.2020.01.016>
- Malhotra, N. (2006). Questionnaire design and scale development. *The Handbook of Marketing Research: Uses, Misuses, and Future Advances*.
- Malhotra, N. K. (2006). Questionnaire design and scale development. In R. Grover & M. Vriens (Eds.), *The handbook of marketing research: Uses, misuses, and future advances* (pp. 83–94). SAGE.

- National Highway Traffic Safety Administration. (n.d.). *Distracted driving*. Retrieved June 27, 2025, from <https://www.nhtsa.gov/risky-driving/distracted-driving>
- National Highway Traffic Safety Administration. (2012, February 16). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices* (Docket No. NHTSA-2010-0053). U.S. Department of Transportation. https://www.nhtsa.gov/sites/nhtsa.gov/files/distraction_npfg-02162012.pdf
- Nicolas, L., Rod, M., & Thomas, E. (2024). *Designing touchscreen menu interfaces for in-vehicle infotainment systems: The effect of depth and breadth trade-off and task types on visual-manual distraction* (No. arXiv:2404.11469). arXiv. <https://doi.org/10.48550/arXiv.2404.11469>
- Nielsen, J. (1993). *Usability engineering*. Academic Press.
- Nielsen Norman Group. (n.d.). *10 usability heuristics for user interface design*. Retrieved June 23, 2025, from <https://www.nngroup.com/articles/ten-usability-heuristics/>
- Nielsen Norman Group. (n.d.). *Progressive disclosure*. Retrieved June 23, 2025, from <https://www.nngroup.com/articles/progressive-disclosure/>
- Nielsen Norman Group. (n.d.). *Usability 101: Introduction to usability*. Retrieved June 24, 2025, from <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>
- Osswald, S., Wurhofer, D., Trösterer, S., Beck, E., & Tscheligi, M. (2012). Predicting information technology usage in the car: Towards a car technology acceptance model. *Proceedings of the 4th International Conference on*

- Automotive User Interfaces and Interactive Vehicular Applications*, 51–58. <https://doi.org/10.1145/2390256.2390264>
- Park, D. E., & Park, S.-E. (2021). Factors affecting perceived safety and enjoyment based on driver experience. *Transportation Research Part F: Traffic Psychology and Behaviour*, 83, 148–163. <https://doi.org/10.1016/j.trf.2021.10.006>
- Pereira, D. G., Afonso, A., & Medeiros, F. M. (2015). Overview of Friedman’s test and post-hoc analysis. *Communications in Statistics — Simulation and Computation*, 44(10), 2636–2653. <https://doi.org/10.1080/03610918.2014.931971>
- Ranney, T. A., Garrott, W. R., & Goodman, M. J. (2001, June 4). *NHTSA driver distraction research: Past, present, and future*. International Technical Conference on Enhanced Safety of Vehicles. <https://saemobilus.sae.org/papers/nhtsa-driver-distraction-research-past-present-future-2001-06-0177>
- Salvucci, D. D., & Kujala, T. (2016). Balancing structural and temporal constraints in multitasking contexts. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 38, 2501–2506. <https://escholarship.org/uc/item/4dc1w60k>
- Simons-Morton, B., Guo, F., Klauer, S. G., Ehsani, J. P., & Pradhan, A. K. (2014). Keep your eyes on the road: Young driver crash risk increases according to duration of distraction. *The Journal of Adolescent Health*, 54(5 Suppl), S61–S67. <https://doi.org/10.1016/j.jadohealth.2013.11.021>

- Stapenhurst, T. (2009). *The benchmarking book: A how-to guide to best practice for managers and practitioners*. Butterworth-Heinemann.
- Strayer, D. L., Cooper, J. M., Goethe, R. M., McCarty, M. M., Getty, D., & Biondi, F. (2017). *Visual and cognitive demands of using in-vehicle infotainment systems*. AAA Foundation for Traffic Safety. <https://aaafoundation.org/wp-content/uploads/2017/11/VisualandCognitive.pdf>
- Suh, Y., & Ferris, T. K. (2019). On-road evaluation of in-vehicle interface characteristics and their effects on performance of visual detection on the road and manual entry. *Human Factors*, 61(1), 105–118. <https://doi.org/10.1177/0018720818790841>
- Tao, D., Yuan, J., Liu, S., & Qu, X. (2018). Effects of button design characteristics on performance and perceptions of touchscreen use. *International Journal of Industrial Ergonomics*, 64, 59–68. <https://doi.org/10.1016/j.ergon.2017.12.001>
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1), 25–29. <https://doi.org/10.1037/h0071663>
- Tobii. (n.d.). *Advanced wearable eye tracker built for research*. (n.d.). Retrieved May 2, 2025, from <https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3/>
- Victor, T., Strategic Highway Research Program Safety Focus Area, Transportation Research Board, & National Academies of Sciences, Engineering, and Medicine. (2014). *Analysis of naturalistic driving study data: Safer glances*,

driver inattention, and crash risk. Transportation Research Board. <https://doi.org/10.17226/22297>

Volvo Cars. (2024, April 25). *EX90 centre display views*. Volvo Support MT.

Retrieved May 15, 2025,

from <https://www.volvocars.com/mt/support/car/ex90/article/47d2c97fd33effd3c0a8cc3718c999b7-655ba57eb6bd2eedc0a8b0971b5b532a-8664b2fa77a7e089c0a8296870d1a409/?search=quick%20controls&page=0>

Xu, Z., Zhang, K., Min, H., Wang, Z., Zhao, X., & Liu, P. (2018). What drives people to accept automated vehicles? Findings from a field experiment. *Transportation Research Part C: Emerging Technologies*, 95, 320–334. <https://doi.org/10.1016/j.trc.2018.07.024>

Yoon, S. H., Lim, J. H., & Ji, Y. G. (2015). Perceived visual complexity and visual search performance of automotive instrument cluster: A quantitative measurement study. *International Journal of Human–Computer Interaction*, 31(12), 890–900. <https://doi.org/10.1080/10447318.2015.1069661>

Appendices

Appendix A – Participant Survey

Personal Details

- First and surname
- Email
- Phone number

Additional Information

- To ensure accurate eye-tracking, please confirm the following:
 - You do not wear bifocal or progressive lenses.
 - You have not had any type of eye surgery.
 - You do not have cataracts.
 - You do not have any eye movement or alignment disorders.
 - You can either avoid wearing glasses during the test or use contact lenses.
 - You can avoid wearing heavy make-up, mascara, fake eyelashes, or coloured contact lenses during the session.
- Does all of the above apply to you? (select an option)
 - Yes / No

Demographic Questions

- Age
- Gender (select an option)
 - Male female other / I prefer not to say

Appendix B – Test Description

Test Procedure

1. Meet Participant
2. Introduction & Transport to Track
3. Getting Settled in the Car
4. Exploration & Final Setup
5. Warm-up Drive (No Tasks)
6. Task Execution
7. Post-Condition Measures
8. Post-Drive Interview
9. Return to PVE & Wrap-up

Detailed Description

1. Meet Participant

- Meet the participant at the PVE reception and briefly explain that the session will be conducted in a car.
- Have the participant sign the GDPR consent form.
- Escort the participant to the vehicle and seat them in the front passenger seat.

2. Introduction & Transport to Track

- While the test leader drives to the track, ask:
 - What car model they drive.
 - How often they drive.
 - Whether they are used to touchscreen interfaces.
- Introduce the goal of the study:
 - Explain that the focus is on the user experience of "Quick Controls" on the center display.

- Emphasize usability and perceived safety during driving.
- Clarify that the participant is not being tested-only the interface.
- Provide a brief overview of the session structure:
 - Parked exploration, warm-up drive, four tested designs, post-task feedback, and final interview.
- Show a printed map of the test track and explain the loop structure.

3. Getting Settled in the Car

- Guide the participant to the driver's seat and help them get comfortable.
- Store personal belongings in the back seat and check for loose objects in the cabin.
- Introduce prototype car:
 - Vehicle is a modified Polestar 2.
 - New screen, steering wheel, and control stalks.
 - No airbag, but approved safety measures are in place.
- Show gear stalk, stalk functions (blinkers, wipers), and confirm everything is safe and ready.
- Explain that the session will be recorded (camera + eye-tracking).

4. Exploration and Eye-Tracking Setup

- Set the interface to the first design and allow 2 minutes of free exploration while parked.
- Fit and calibrate the eye-tracking glasses.

5. Warm-Up Drive (First Concept Only)

- Instruct the participant to drive one lap without any tasks.
- Emphasize speed limits (60–70 km/h) and lane keeping.

6. Task Execution

- Before each task set, remind the participant:
- Start each task from the Home screen.
- Maintain consistent speed and lane control.
- You will provide instructions directly while driving.

Task List:

- **Task 1:** Change Drive Mode to 'Range'.
- **Task 2:** Set Suspension Feel to 'Stiff'.
- **Task 3:** Lower Center Screen brightness.
- **Task 4:** Change Roof Shading to 'Auto'.
- **Task 5:** Set Wiper Sensitivity to 'Medium'.
- **Task 6:** Check status of Exterior Lights.

- At the end of the lap, instruct the participant to pull over.

7. Post-Condition Measures

- While parked, have the participant complete:
 - CTAM Perceived Safety Scale.
 - NASA-TLX.
 - UMUX-Lite.
- Encourage brief think-aloud reflections before proceeding to the next design.

(Repeat Section 6–7 for each of the four interface concepts.)

8. Post-Test Interview

- Overall Experience & Initial Comparison
 - When you think back to the four interfaces you tried while driving, what's the first thing that comes to mind?
 - Did any one of the designs feel safer to use while driving? What made it feel that way?
 - Was there any design that felt less safe? What gave you that feeling?
 - Did any design feel more distracting? What was distracting about it?

- **Flat or Nested Structure**
 - When you used the flat design, how did it feel to find and use controls while driving?
 - How about the nested designs? How did it feel to find and use controls while driving?
 - Focusing only on safety, which structure felt safer: flat or nested? What made it feel that way?

- **Button Size, Within Nested Concepts**
 - Now just thinking about the nested designs, they had different button sizes on the last step in each setting: small, medium, large.
 - How did those different sizes affect how you used the interface after choosing a setting?
 - And again thinking about safety, which button size felt safest to use while driving: small, medium or large?

- **Preference and Final Suggestions**
 - If you had to pick one of these designs for your own car, which one would you choose, and why?
 - Do you have any other ideas or suggestions to improve the design, based on what you tried today?

- **Ranking the Designs**
 - Could you rank the 4 designs from most preferred (1) to least preferred (4) for driving use?

9. Return to PVE and Wrap-Up

- Drive the participant back to the office.
- Thank them for participating.
- Invite any final questions and explain how their input contributes to the study.

Appendix C – User Test Scales

Perceived Safety items from Osswald's CTAM

Rate how much you agree with each statement based on your experience.

	Strongly disagree	Disagree	Somewhat disagree	Neutral	Strongly agree	Agree	Strongly agree
<i>I believe that using the system is dangerous.</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>Using the system requires increased attention.</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>The system distracts me from driving.</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>I feel safe while using the system.</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>I can use the system without looking at it.</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

UMUX-Lite

Rate how much you agree with each statement based on your experience.

Strongly disagree Disagree Somewhat disagree Neutral Strongly agree Agree Strongly agree

The design concept is easy to use.

The design concept's capabilities meet my requirements.

NASA Task Load Index (NASA-TLX)

Hart and Staveland's NASA Task Load Index (TLX) method assesses workload on five 7-point scales. Increments of high, medium, and low estimates for each point result in 21 gradations on the scales.

Mental Demand *How mentally demanding was the task?*

Very Low Very High

Physical Demand *How physically demanding was the task?*

Very Low Very High

Temporal Demand *How hurried or rushed was the pace of the task?*

Very Low Very High

Performance *How successful were you in accomplishing what you were asked to do?*

Very Low Very High

Effort *How hard did you have to work to accomplish your level of performance?*

Very Low Very High

Frustration *How insecure, discouraged, irritated, stressed, and annoyed were you?*

Very Low Very High

Appendix D – Results – Full Data

NHTSA Second Guideline Violations

Task	FS	NS	NM	NL
Task 1	4,00	1,00	0,00	2,00
Task 2	4,00	1,00	0,00	2,00
Task 3	2,00	1,00	2,00	0,00
Task 4	1,00	0,00	1,00	0,00
Task 5	2,00	1,00	0,00	0,00
Task 6	1,00	0,00	0,00	0,00
Total:	14,00	4,00	3,00	4,00

Perceived Safety Scores

All Test Participants (TP1-TP19)

Item	FS	NS	NM	NL
<i>I believe that using the system is dangerous.</i>	4,84	4,37	4,74	4,79
<i>Using the system requires increased attention.</i>	3,32	3,21	3,58	3,42
<i>The system distracts me from driving.</i>	3,68	3,68	3,79	3,79
<i>I feel save while using the system.</i>	5,00	5,05	5,26	5,11
<i>I can use the system without looking at it.</i>	2,32	2,47	2,84	2,79
Average:	3,83	3,76	4,04	3,98

TP1-TP8

Item	FS	NS	NM	NL
<i>I believe that using the system is dangerous.</i>	5,00	4,71	4,86	5,14
<i>Using the system requires increased attention.</i>	3,71	3,43	3,57	3,57
<i>The system distracts me from driving.</i>	4,00	4,00	4,14	4,00
<i>I feel save while using the system.</i>	5,43	5,57	6,29	5,43
<i>I can use the system without looking at it.</i>	2,86	3,57	3,86	3,57
Average:	4,20	4,26	4,54	4,34

TP9-TP19

Item	FS	NS	NM	NL
I believe that using the system is dangerous.	4,55	4,00	4,55	4,45
Using the system requires increased attention.	3,18	3,18	3,64	3,55
The system distracts me from driving.	3,55	3,55	3,64	3,73
I feel safe while using the system.	4,64	4,64	4,55	4,82
I can use the system without looking at it.	2,00	1,91	2,36	2,45
Average:	3,58	3,45	3,75	3,80

UMUX-Lite Scores

All Test Participants (TP1-TP19)

Item	FS	NS	NM	NL
The design concept is easy to use.	79,82	76,32	80,70	82,46
The design concept's capabilities meet my requirements.	77,19	67,54	71,93	71,93
Average:	78,51	71,93	76,32	77,19

TP1-TP8

Item	FS	NS	NM	NL
The design concept is easy to use.	73,81	73,81	76,19	78,57
The design concept's capabilities meet my requirements.	78,57	57,14	64,29	57,14
Average:	76,19	65,48	70,24	67,86

TP9-TP19

Item	FS	NS	NM	NL
The design concept is easy to use.	83,33	77,27	83,33	84,85
The design concept's capabilities meet my requirements.	77,27	74,24	77,27	81,82
Average:	80,30	75,76	80,30	83,33

NASA-TLX Scores

All Test Participants (TP1-TP19)

Item	FS	NS	NM	NL
Mental Demand	14,37	12,89	14,37	14,39
Physical Demand	17,55	16,47	16,89	18,03
Temporal Demand	16,37	15,74	16,00	17,26
Performance	17,37	17,42	17,89	17,37
Effort	15,42	15,89	16,74	15,84
Frustration	16,63	14,95	16,58	16,03
Average:	16,29	15,56	16,41	16,49

TP1-TP8

Item	FS	NS	NM	NL
Mental Demand	15,63	14,00	15,75	16,38
Physical Demand	18,44	18,13	18,50	18,13
Temporal Demand	17,45	17,25	18,50	18,25
Performance	16,69	15,88	16,38	15,88
Effort	16,63	17,25	17,00	17,13
Frustration	17,75	16,50	17,13	16,13
Average:	17,10	16,50	17,21	16,98

TP9-TP19

Item	FS	NS	NM	NL
Mental Demand	13,45	12,09	13,36	12,95
Physical Demand	16,91	15,27	15,73	17,95
Temporal Demand	15,59	14,64	14,18	16,55
Performance	17,86	18,55	19,00	18,45
Effort	14,55	14,91	16,55	14,91
Frustration	15,82	13,82	16,18	15,95
Average:	15,70	14,88	15,83	16,13

Concept & Item Rankings

General Concept Ranking – From Most Preferred (1) to Least Preferred (4)

Test Participant	1 Most Preferred	2	3	4 Least Preferred
TP1	FS	NM	NS	NL
TP2	NM	NL	FS	NS
TP3	FS	NM	NL	NS
TP4	FS	NM	NL	NS
TP5	FS	NM	NS	NL
TP6	FS	NM	NL	NS
TP7	FS	NM	NL	NS
TP8	FS	NL	NM	NS
TP9	FS	NL	NM	NS
TP10	NM	NS	FS	NL
TP11	NM	FS	NL	NS
TP12	NM	NL	NS	FS
TP13	NL	NM	FS	NS
TP14	FS	NL	NM	NS
TP15	NM	NS	NL	FS
TP16	FS	NM	NL	NS
TP17	FS	NL	NM	NS
TP18	FS	NM	NS	NL
TP19	FS	NL	NM	NS



CHALMERS
UNIVERSITY OF TECHNOLOGY