CHALMERS
UNIVERSITY OF TECHNOLOGY

Adoption of Voice User Interfaces in Controlling In-Vehicle Features

Understanding Key Influencing Factors and Creating Design Guidelines for Replacing Physical Controls

Master's thesis in Industrial Design Engineering

SIYUAN HU & JIAWEI TAO

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE

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

MASTER'S THESIS 2025

Adoption of Voice User Interfaces in Controlling In-Vehicle Features

Understanding Key Influencing Factors and Creating Design Guidelines for Replacing Physical Controls

Siyuan Hu
Jiawei Tao

Department of Industrial and Materials Science
Division of Design & Human Factors
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025

Adoption of Voice User Interfaces in Controlling In-Vehicle Features
Understanding Key Influencing Factors and Creating Design Guidelines for Replacing Physical Controls
Siyuan Hu
Jiawei Tao

Supervisor: Fredrick Ekman, Design & Human Factors, Industrial and Materials Science, Chalmers University of Technology
Examiner: Bijan Aryana, Design & Human Factors, Industrial and Materials Science, Chalmers University of Technology

Master's Thesis 2025
Department of Industrial and Materials Science
Division of Design & Human Factors
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: The example concept of the VUI touchpoint using the established design guideline.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2025

Abstract

Contemporary vehicles are usually equipped with numerous interior features with complicated designs, which usually results in higher energy consumption, safety issues related to distractions, and a shorter service life due to the overreliance on screens and complicated interfaces. To address this issue, it is necessary to develop a concept to reduce the number of interior features.

The purpose of this thesis is to investigate whether the VUI (voice user interface) has the ability to replace the physical controls in operating driving task during driving. Three research questions guide this investigation. The literature review employs Desmet's Product Emotion and Experience theories as a foundation for understanding VUI, alongside an analysis of the current state of VUI research. A mixed-method approach was used to explore user perspectives, incorporating qualitative and quantitative data collected through questionnaires, interviews, and statistical analysis. The findings reveal that participants expressed a more positive attitude toward using VUI for secondary and tertiary driving tasks, while showing reluctance for primary driving tasks. Based on these insights, a User Perception Mechanism and a Decision-Making Model were developed, accompanied by design guidelines. A comparative experimental study identified specific physical controls that could be replaced by voice interaction. These findings informed the creation of use scenarios and the development of a concept —a car key fob integrated with VUI, which was visualized with 3D design tools. The results suggest that VUI has significant potential to support controlling driving tasks features, enhance user experience, and contribute to the development of future intelligent car cockpits.

Keywords: Voice User Interface, User Experience, Industrial Design,Interaction Design, Automotive, Driving tasks

Acknowledgments

This report is the culmination of the master's thesis conducted at the Department of Industrial and Materials Science at Chalmers University of Technology. The project was part of the Industrial Design Engineering master's program and was carried out in collaboration with the LYNK&CO Design Studio in Gothenburg. Its completion would not have been possible without the support and guidance of numerous individuals and organizations.

First and foremost, we wish to express our heartfelt gratitude to our supervisors, Dr. Fredrick Ekman, and our examiner, Dr. Bijan Aryana, who provided unwavering support, insightful guidance, and constructive feedback throughout the process, and whose expertise and encouragement were instrumental in helping us complete this thesis.

We are also thankful to LYNK&CO for their support, which provided the foundation for this research. We are grateful to our company contact Stefano Oliva at Lynk & Co, for providing us with an insightful and challenging opportunity. Moreover, we want to direct our appreciation to our opponent William Ivanovski for his insightful comments.

Last but not least, we wish to express our heartfelt gratitude to all the respondents who participated in our project, as well as to everyone who supported and assisted us throughout this journey.

Thank you all for being part of this journey!

Hu Siyuan & Tao Jiawei

Gothenburg, 2025

Acronyms

Below presents the acronyms used in the report.

VUI: Voice User Interface

OEM: Original Equipment Manufacturer

HCI: Human–computer Interaction

PCA: Principal Component Analysis

MLR: Multiple Linear Regression

PDT: Primary Driving Task

SDT: Secondary Driving Task

TDT: Tertiary Driving Task

BEV: Battery Electric Vehicle

PHEV: Plug-in Electric Vehicles

Contents

List of Figures

List of Tables

1 Introduction

The first chapter provides an overview of the project including the background, research questions and objectives. In addition, an overview of the report structure is presented.

1.1 Background

User experience plays an important role in the automotive industry, from the initial research and purchasing process to the driving experience and after-sales service. For car designers, user experience not only includes the physical design of vehicles, but also considers every aspect of interaction between the user and the car.

In the automotive field, driving tasks are defined as different activities conducted by drivers or passengers associated with operating a vehicle and are divided into three classes (Pfleging & Schmidt, 2015). The primary driving task comprises all activities that are required to maneuver the vehicle, including all activities regarding lateral and longitudinal control of the vehicle, as well as "maintaining alertness to traffic and other potential hazards". The secondary task relies on and supports the primary driving task. It refers to functions that increase driving performance or safety, such as activating headlights, cruise control, or windshield wipers. The tertiary tasks refer to all other tasks such as operating comfort, infotainment, and communication systems, or eating and drinking. Features are components or systems that driving tasks rely on. (e.g., the car door is a feature, and opening the door is a tertiary driving task).

The contemporary automotive industry faces a trade-off between consumer demand for numerous features and considerations of safety and sustainability. Various car manufacturers have launched vehicles equipped with numerous interior features with complicated designs, which is a way for them to gain a competitive advantage against competitors. However, these additions usually result in higher energy consumption, safety issues related to distractions, and a shorter lifespan of the design due to over-reliance on screens and complicated interfaces.

VUI, which stands for Voice User Interface, is a technology that enables users to interact with computers, devices, or applications using spoken language (M. Cohen, Giangola, & Balogh, 2004). It emerges as a promising technology that allows for hands-free interaction between users and specific products. Moreover, it has also become a trend to use VUI to complete different driving tasks within vehicles which contributes to minimizing driver distraction and cognitive load due to visual stimuli (Murali, Kaboli, & Dahiya, 2022). For example, Amazon's voice assistant, 'Alexa', has been integrated with tens of car brands, which can play music, adjust cabin temperature, and perform other tasks based on user voice commands (Amazon, 2023).

Apparently, interior features can be optimized if VUI is widely used in the car. However, to what extent VUI can complete driving tasks remains uncertain, as well as what specific interior features can be controlled by VUI. Also, how this transformation will potentially impact user experience remains unclear. This master thesis project was initiated in order to investigate these issues and offer relevant recommendations to the automotive industry.

1.2 Aim

The aim of the thesis is to explore the feasibility of utilizing VUI to conduct different types of driving tasks and to evaluate which tasks is suited to be operated using VUI without setting up physical controls. In here, In-car physical controls refer to the tangible, manually operated devices or interfaces present in the vehicle, such as buttons, knobs, levers, or touchscreen systems, which require direct physical interaction by the user.

1.3 Research questions

Three specific research questions are designed to guide this thesis:

1. What specific driving tasks can be conducted using VUI?

2. Are there any driving tasks that are preferred to be conducted using VUIs? If so, which?

3. How can a concept using VUI be designed and how does user experience affect the design?

1.4 Research objectives

According to research questions, the subdivided research objectives are presented as follows:

1. Identify specific tasks that can be conducted using VUI.

2. Determine the key factors influencing the adoption of VUIs

3. Develop actionable design guidelines and implement a VUI design practice for controlling in-car features.

1.5 Overview of report

The report consists of 8 chapters. Here is the overview of the report excluding chapter *1. Introduction*.

Chapter *2. Literature Review* presents the fundamental theories of the project and focuses on user experience and VUI design and evaluation.

Chapter *3. User Study* presents the research methods used, including quantitative and qualitative approaches. It details the preparation, participant selection, and the design/analysis of survey and interview questions aimed at understanding user experiences and expectations of VUIs in vehicles.

Chapter *4. Influence Model and Factors* introduces a model to illustrate the key factors affecting VUI adoption for in-car feature control without applying physical controls. It includes components such as user attributes, expectations, and product features.

Chapter *5. Comparative Study and Experiment* describes the experiment setup, participant recruitment, and procedures for testing the usability, user experience, and safety of VUIs in feature controls. Results from comparative studies between VUIs and physical controls are presented.

Chapter *6. Design Guidelines and Practice* provides design guidelines for VUIs, emphasizing interaction strategies, feedback mechanisms, and technical requirements. It also showcases a final VUI touchpoint design practice based on these guidelines.

Chapter *7. Discussion and Conclusion* reflects on the study's findings, their implications, and their alignment with existing theories. It also examines the limitations of the research and offers directions for future studies. Key outcomes of the study are summarized, with a focus on the feasibility and design of VUIs as potential replacements for physical controls. The discussion highlights critical insights into user experience and usability, offering design guidelines for practitioners and suggesting areas for future work.

2 Literature review

This chapter aims to identify existing theories related to the project. The theoretical foundation is constructed around three primary components: driving tasks, VUI, and user experience. Each of these elements plays a crucial role in guiding the execution of the project.

2.1 User experience

The International Organization for Standardization (ISO) offers a widely accepted definition of user experience as ”the perceptions and responses resulting from the use and/or anticipated use of a product, system, or service” (International Organization for Standardization, 2018). User experience is an overarching concept which involves the feelings and emotions of users, highlighting the importance of studying the emotional responses users have while interacting with products.

In the first part of this section, we introduced and explained the theoretical foundations of the product emotion model and review how the appraisal model is applied in user interviews. In the second part, we describe the product emotion model developed by Desmet and others based on the appraisal model and discuss how this framework maps the various psychological, social, and behavioral effects pathways that arise from human-product interactions. This aids in our understanding and application of the framework in user research to analyze the various emotions and cognition that users experience when interacting with products. The final part will clarify the generality of the framework and how it will guide our application of this theoretical framework to research and design frameworks suitable for voice user interfaces.

2.1.1 Emotion and appraisal model

Emotions are often understood or defined as ”positive or negative experiences associated with specific physiological activities,” which can lead to various physiological, behavioral, or cognitive changes with varying effects and duration. For example, consumers may feel joy when purchasing a desired product. Frijda notes that emotional states reflect the affective relationship between an individual and a specific object, such as the happiness experienced by consumers when acquiring a desired product. Therefore, when studying product experience and the emotional responses it evokes, the product should be considered as the starting point of research.

In the cognitive tradition of psychology, emotions are considered a signal that indicates whether events are beneficial or detrimental to an individual’s concerns. Appraisal can be understood as the evaluation and interpretation process of the emotion-inducing event, which can then be defined as a rapid assessment of an individual’s sense of well-being (Frijda, 1986; Lazarus, 1991). Although appraisals are mostly automatic and non-verbal, to conceptualize this, appraisal can be broadly understood as answering the question: “What does this situation mean for my well-being?” When the answer is beneficial to personal well-being, it generally results in pleasant emotions; conversely, it may trigger unpleasant emotions. For example, when a computer crashes at a critical moment, one might feel anger or anxiety.

The research literature on appraisal suggests two main methods of differentiation: themes and components. The thematic method describes appraisals using summative statements that reflect the overall personal significance of a situation (Lazarus, 1991). For instance, sadness is generally associated with irrevocable loss, while happiness corresponds to achieving goals. On the other hand, the componential approach does not rely solely on a foundational question but explores various aspects of a situation through multiple questions such as, ”How does this situation affect my motives?” and ”Is this situation as expected?” The answers to these questions are termed appraisal components, and each discrete emotion involves a specific pattern of these components (Roseman, 2001; Scherer, 2001).

Although the thematic method provides more than just a sum of components and offers a holistic understanding of anticipated emotions, Desmet, Demir (2002, 2009), and others believe that the appraisal components approach allows for a more systematic and detailed analysis of the emotions that might arise during the design process and understanding user responses to products and the experiences they deliver across different settings. Despite its limitations and the challenge of connecting abstract concepts like ”irrevocable loss” with product emotional experiences, this method is still valuable. According to research by Demir et al. (2009), the seven appraisal components necessary in the design application process are: motive consistency, intrinsic pleasantness, expectation confirmation, standard conformance, agency, coping potential, and certainty. These components are relevant to user experience in human-product interactions and can effectively assist researchers in designing and refining user study.

- **Motive consistency:** Concerns how a situation aligns with individual motives. Positive emotions arise if the situation is consistent with one’s desires, while conflicting situations evoke negative emotions. Motives range from abstract (universal needs) to specific (immediate goals).
- **Intrinsic pleasantness:** Relates to the direct sensory appeal of an object or situation. This can be viewed as a type of motive consistency related to survival instincts, functioning without conscious motive representation.
- **Expectation confirmation:** Involves assessing whether outcomes meet preconceived expectations. Satisfaction or disappointment occurs when expectations are confirmed or disconfirmed, respectively.
- **Standard conformance:** Assesses how a situation conforms to social norms and personal standards. Violations may trigger negative emotions like anger or guilt, while adherence can elicit pride or admiration.
- **Agency:** Deals with attributions of responsibility for an event. Can lead to self-conscious emotions (pride, shame) if one attributes the cause to oneself, or external emotions (anger, admiration) if others are deemed responsible.
- **Coping potential:** Evaluates one’s capacity to manage or alter undesirable circumstances. High coping potential may lead to proactive behavior to change a situation, whereas low potential might result in avoidance behaviors like fear or anxiety.
- **Certainty:** Pertains to the level of certainty one has about an event. Uncertainty can cause fear or hope depending on the potential outcomes, while certainty is linked to emotions like happiness or sadness when outcomes are clear.

The components and their respective questions are shown in the Table 1 below.

Table 1: Appraisal Components and Their Key Relational Issues

Appraisal Component	**Key Relational Issue**
Motive consistency	Does this event help me in attaining my goals and motives?
Intrinsic pleasantness	Is this event pleasant or enjoyable?
Expectation confirmation	Does this event confirm my expectations?
Agency	Who (or what) is responsible for this event?
Standards conformance	Does this event match with my social norms and standards?
Coping potential	Can I handle or change the event?
Certainty	How certain am I about the event?

2.1.2 Product emotion and experience

According to Ortony's cognitive model (1988), Desmet and Hekkert (2002) built upon the appraisal model to explore how products and their interactive experiences evoke emotions during use, proposing a seminal model that categorizes product emotions. While individual emotional responses to products vary, these responses universally result from the product's impact on personal concerns, such as standards, attitudes, and goals.

This model divides the emotional response process into four key parts: appraisal, concerns, product, and emotion. The first three elements and their interplay determine whether a product triggers an emotion and the type of emotion elicited. The product serves as an emotional stimulus in three ways: the product itself (Objects), the product (or designers) as agents, and the product as a promise of future use or ownership (Events). Each emotion is rooted in a concern—a relatively stable preference for specific states of the world, as outlined by Frijda (1986). Frijda considers these concerns as reference points during the appraisal process and a concept is further elaborated by Desmet (2003). The corresponding concerns are attitudes, standards, and goals, respectively. Firstly, products evoke emotional responses through their aesthetic appeal or functional attributes, directly influencing user emotions based on their appearance or usability. Secondly, the product acts as an agent, where the designers' intentions and the product's functionalities play crucial roles, including innovative elements and potential design flaws, significantly impacting the user experience. Thirdly, the product is seen as a promise of future benefits, involving the emotional impact expected from using or owning the product, such as the status elevation anticipated from acquiring luxury items.

1. **Products as Objects**: Relates to the intrinsic properties of the product (e.g., design, appearance) and how these properties match personal tastes or attitudes.
2. **Products as Agents**: Considers products in terms of their functions or the value brought by its creators(e.g., a car's reliability or a designer's creativity).
3. **Products as Events**: Focuses on the anticipated future use or ownership of the product and its expected consequences.

Building upon the basic model of product emotions, Desmet and others suggest that the interaction process between users and products is as crucial as the product itself. Unlike the direct

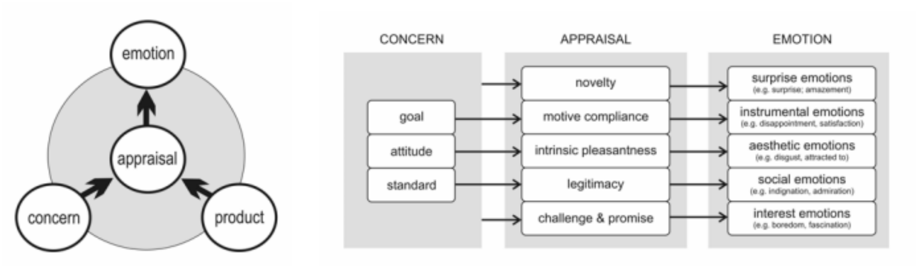


Figure 1: Basic model of product emotions(left) and Classification of product emotions (Right)

emotional stimulation by the product's physical aspects, interactions with the product generate concerns (instrumental interactions, non-instrumental interactions, and non-physical interactions), which foster emotional inclinations towards the product, thereby shaping the overall product experience. Here, the product experience is comprised of three distinct components: aesthetic experience, experience of meaning, and emotional experience (2007). As shown in Figure2. The aesthetic experience involves the sensory delight or offense obtained from a product, from not at all aesthetically pleasing to very aesthetically pleasing (Blijlevens et al., 2017). Experience of meaning involves the cognitive processes of interpretation and association to attribute significance to products. It can be generated from sight, hearing, and touch, and can also be affected by contextual factors like culture and more (Karana et al., 2010). Emotional experience is an affective phenomenon that covers a wide range of feelings and is triggered by the personal evaluation of the product's potential advantages or disadvantages (Desmet, 2002).

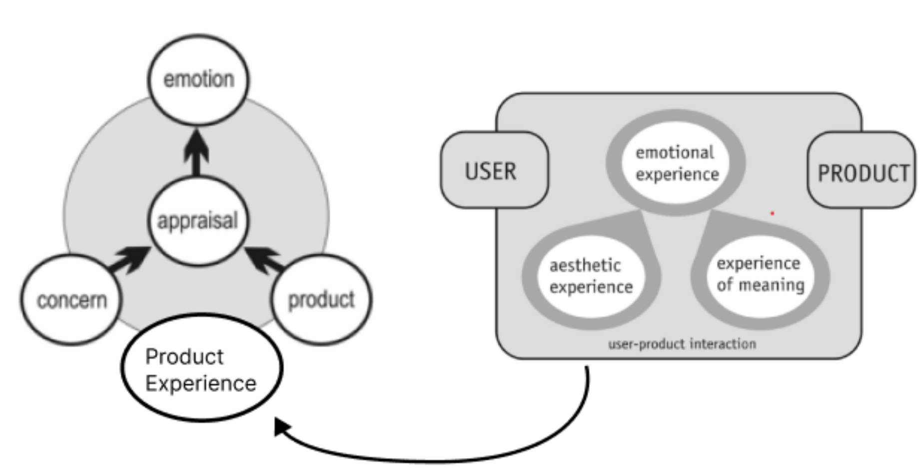


Figure 2: Relationship between product emotion and experience

1. **Aesthetic Experience**: Involves sensory pleasure derived from the product and is influenced by its design, form, and material.

2. **Experience of Meaning**: Engages cognitive processes where products are imbued with personal or symbolic significance, influenced by cultural, social, and individual factors.

3. **Emotional Experience**: Encompasses feelings and emotions that arise from the product's interaction with the user's personal goals and values.

The relationship between aesthetic experience, experience of meaning, and emotional experience is intertwined but can be emphasized as the relationship between the emotional part and the other two. Generally speaking, experience of meaning and emotional experience both generate emotional experience because they are both highly subjective and vary from person to person. For the experience of meaning, a modern stainless kitchen may make people feel efficient while some people may think it is cold and impersonal, which results in disaffection. For aesthetic experience, people may feel emotionally disappointed at something that does not match their aesthetic taste.

Fokkinga, Desmet, and Hekkert (2020) have synthesized previous theories on product emotion and experience to formulate a comprehensive framework for understanding the impact of design on user psychology and behavior. This framework transitions from traditional user-centered design to a broader perspective that includes the effects of design on both individual and societal psychological and behavioral outcomes. This perspective is structured into three levels:

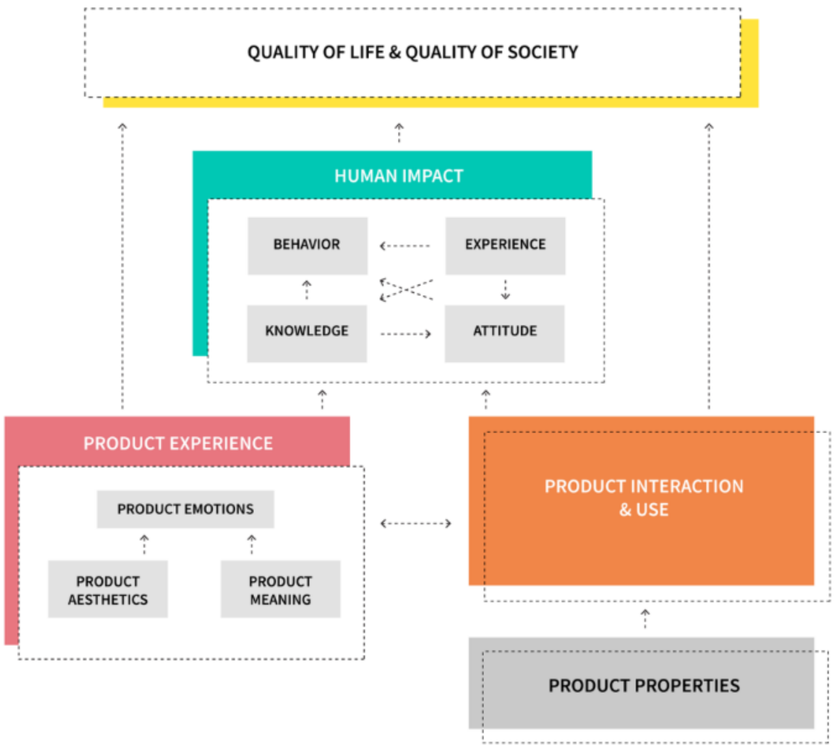


Figure 3: The impact-centered design framework (Fokkinga, Desmet, & Hekkert, 2020).

- **First Level**: Direct product interactions and experiences, including aesthetic, meaningful, and emotional experiences.

- **Second Level**: Indirect and long-term impacts on behaviors, attitudes, knowledge, and general experiences.

- **Third Level**: Effects on quality of life and societal norms.

Table 2: Basic descriptions of the framework elements from the Impact-Centered Design framework (Fokkinga, Desmet, & Hekkert, 2020).

Impact Dimension	Description
Overall Impact	
Quality of Society	The overall impact on the composition and values of a society.
Quality of Life	The overall impact on the well-being of individuals and communities.
Human Impact	
Behavior	Activities and deeds that are enabled and stimulated by using the product.
Knowledge	Insights gained and greater understanding facilitated by using the product.
Experience	Feelings evoked by situations and events that are enabled or supported by the product.
Attitude	Opinions (about people, objects, events, actions, and ideas) that are influenced by using the product.
Product Properties	
Meaning	Meanings, expressions, and associations assigned to the product.
Aesthetics	The extent to which the product gratifies (or offends) the human sensory systems.
Emotions	Positive and negative emotions evoked by the product.
Human-Product Interaction	All events that take place between the individual and the product, such as perceiving it, using it, and storing it.

The first level includes two dimensions: the product itself, which encompasses product properties and the interactions that occur when people use the product. The emotional dimension induced by the product, which incorporates the three types of product experiences introduced by Desmet and Hekkert (2008): Aesthetic, Meaning, and Product Emotion. This level triggers the second layer of personal impact factors, including behavior, attitudes, (overall) experiences, and knowledge. This, in turn, extends to influence the quality of life of users and other stakeholders, as well as long-term impacts on societal dimensions. This is illustrated in the figure 3. Detailed definitions and explanations for each part are displayed in Table2.

When analyzing the impact (anticipated or unanticipated) of products, the framework's logic flows from the bottom to the top: the properties of the product lead to specific product interactions, which in turn trigger product experiences and broader human impacts. This helps us understand the fundamental reasons behind users choosing products and aids researchers in understanding and applying knowledge to meet actual user needs. However, it is noteworthy that in reality, designers find that impacts are complex and challenging to categorize and discuss individually. This means there may be multiple strategic pathways to achieve desired human impacts, possibly due to several reasons: psychological effects do not exist in isolation, designers struggle to directly influence the second layer of human impacts, and any process may involve unintended consequences. This also implies that the framework is subject to modifications based on actual situations. We will further discuss and possibly modify this model based on specific conditions and practical applications in chapter 4.

2.1.3 Integration in design practice and design application

The practical application of these theories in design is advocated by frameworks that guide the creation of products tailored to evoke desired emotional responses. This approach encourages designers to consider the emotional impact of their designs as central to the user experience, extending beyond mere functionality to address the complex emotional landscape of the users. For example, Tonetto and Desmet (2016) studied a new method to enhance the ecological validity of user experience surveys by incorporating natural language into the generation of questionnaire items, using the appraisal model. This approach not only helps capture users' genuine feelings about products but also provides more precise feedback for design practice.

Products are intricately intertwined with various human aspects: they depend on and overlap with each other. For example, attitudes, beliefs, and motivations are influenced by knowledge and feedback, and vice versa. Knowledge and attitudes can affect behaviors, new behaviors can lead to different experiences, and emotions can also influence the equation of behaviors. In principle, since each of the four components can affect all other components (and subsequent ones), designers have many approaches they can take through this framework. It is important to note that designers actually exert influence by manipulating the product itself or its various attributes, or by intervening in human behaviors. Additionally, it is assumed that design actions primarily flow from the bottom to the top of the framework (starting from product functionality to ending at quality of life and social quality), but incidental effects can also occur within the same layer of attributes, and even reverse effects are possible.

2.2 VUI

VUI, which comes in various systems or devices controlled by voice but can also provide voice or vision assistance (Klein et al., 2023), is widely accessible and has been integrated into everyday life via smart devices such as smartphones and smart speakers. People are now showing strong interest in using voice interfaces and often view as common way to perform simple tasks (pyae & Joelsson, 2018).

2.2.1 Arising interest

There are several factors influencing people show the interest in using VUI. The following discussion outlines key aspects influencing user interest: Understanding why people are drawn to using VUIs requires examining several key factors that influence technology acceptance. According to Buteau & Lee (2021), technology acceptance is significantly shaped by social influences, including social and personal norms, and privacy concerns. Mclean & Osei-Frimpong (2019) added that the utilitarian, symbolic, and social benefits of VUIs play a crucial role, with users often valuing performance and practical utility over hedonic qualities. The human-like capabilities of VUIs, which simulate companionship, further incentivize adoption by fulfilling both functional and emotional needs.

Trust is another pivotal element driving interest in VUIs. As users engage in brief social exchanges with these interfaces, they develop trust, which is critical for sustained interaction and influences the extent of user engagement and task completion (Lee et al., 2021). Trust factors include system performance, recognition accuracy (Dietvost et al., 2015), and the perceived personality of the interface, which can be customized to enhance relatability and effectiveness (Braun et al., 2019).

Widespread scenarios and applications have led to greater adoption and acceptance of the technology by users. As noted by Sandra Krüger et al. (2021), these interfaces are increasingly perceived as social entities capable of intricate conversational dynamics akin to human-human interactions. This perception not only improves user experience but also aids in mitigating negative emotions and driving pressures associated with traffic congestion, potentially reducing risky behaviors (Krüger et al., 2020). Moreover, developments in VUI technology envision a future where voice assistants extend beyond mere driving aids to become integral components of social communication (Ringfort-Felner et al., 2022). This aligns with findings by Jung et al. (2020), which suggest that integrating VUIs with other modalities like haptic interactions can enhance operational efficiency without increasing distraction, thereby improving overall safety and interaction quality. Furthermore, targeting specific user groups with tailored VUI applications, as demonstrated by Gordon and Breazeal (2015) with their parental voice interface for car trips, showcases the adaptability of VUIs to meet diverse needs and scenarios. This strategic approach not only addresses immediate user requirements but also fosters a deeper connection and reliance on voice technology, broadening its appeal and integration into daily routines.

Table 4: Description of the various failures categorized by Baughan et al. (2023)

Failure Type	**Failure Source**	**Failure Scenario**
Attention	Missed Trigger	Users say something to trigger the voice assistant, but it fails to respond.
	Spurious Trigger	Users do not say something to trigger the voice assistant, but it activates anyway.
	Delayed Trigger	Similar to system latency, the users say something to trigger the voice assistant, but it replies too late to be useful.
Perception	Noisy Channel	User input is incorrectly captured due to background noise.
	Overcapture	The voice assistant captures more input than intended by either beginning to capture input too early or ending too late, and acting on external data not relevant to the users' request.
	Truncation	System does not fully capture users' speech, by either beginning to capture input too late or ending too early.
	Transcription	System generates a transcription error, often in the form of similar sounding words.
Understanding	Ambiguity	There may be several interpretations of the users' intent, and the system responds in a way that is plausibly accurate but not correct for the users' intent.
	Misunderstanding	The system maps the users' input to an incorrect action, perhaps with some correct inference on the users' intent, but not fully accurate.
	No Understanding	The system fails to map the user's input to any known action or response.
Response	Action Execution: No Action	If the system listens to the full request, but then turns off before giving any type of answer or taking action.
	Action Execution: Incorrect Action	The system gives information that is incorrect.

2.2.2 Common failure

Despite significant progress, unforeseen issues can still arise, leading to unexpected failures. Previous studies have shown that when user expectations exceed the capabilities of voice assistants, it often results in system malfunctions and user frustration (Lahoual & Frejus, 2019) This lead to how to identify and categorize the failures. In practice usage, failure may caused by technical issues such as poor connection, poor voice recognition in certain area like car or household environment, linguistic issues like using homonyms of requests items would affect machines perception Mahmood et al., 2022). In categorization, Hong et al exploited voice user interface failure categorization into four main parts: attention (channel), perception (signal), understanding (intention), and response (conversation) and further Baughan et al. (2023) contributed a dataset of 199 failures and categorized across 12 failure sources in Table 4 for formulating code book description of various failures. In this work, we draw on their fault categorization for existing voice user interfaces to categorize failures during use in user study and to understand and assess user expectations and impacts of voice user interface.

2.2.3 Future improvement in user experience

There are several characters in researching and developing the voices user interface and The potential for improvement in user experience.

Naturalness of conversations and Progressively : Wigdor and Wixon (2021) defined naturalness as the 'mimicry of the real world'. In VUI design, achieving naturalness means providing a user experience that closely resembles natural human interaction. Kim et al. (2021) suggested that to create such an experience, designers must consider various aspects of conversation, including the dependency and contextual nature of interactions. They also highlighted a trade-off between naturalness in transactional versus social agents. At a more granular level, progressivity works as core features in conversations naturalness in conversation analysis, and has been referred as "moving from some element to a hearably-next-one with nothing intervening"(Schegloff, 2007). Fisher et al. (2019) mentioned that employing progressive principles in designing a VUI can lead to a more natural human-computer conversation and this strategy need to draw on how speakers continuously collaborate by leveraging a series of syntactic, prosodic, and pragmatic features to propel the conversation forward, meanwhile considering balance methodology, user reaction and potential repair actions. Also, in accordance with Conversation analysis, developing framework that simulates natural conversation could better guide designers in creating VUIs that more closely mimic human conversation, focusing on the sequence of interaction rather than the visual layout (Moore et al., 2023).

Proactive Behavior: Proactivity in interactive systems refers to "the ability to autonomously initiate anticipatory action based on reasoning, meant to impact people and/or their environments" (Grosinger, 2022). In VUIs, initiation can arise from contextual cues or direct user input from prior interactions. Contextual initiation draws upon factors such as time, location, activity recognition, and system usage (Meurisch et al., 2020). In the context of voice user interface, drivers often prefer proactivity and specific use cases closely related to driving tasks and reckon proactive VUI as more intelligent (Schmidt et al., 2019). Subsequent research by Schmidt et al. (2020) revealed that proactive voice interactions, such as proactive reminders, can effectively reduce cognitive load while driving. However, individuals may deactivate proactivity in certain functionalities, such as making appointments or navigating.

Personality of VUI: Due to the conversational nature of voice user interface, they are more likely to have agents with distinct personalities (Völkel et al., 2020). Enhancing the personality of voice interfaces may also generate positive feedback in the dialogue process and task performance. For instance, research indicates that users particularly enjoy interacting with voice assistant that exhibit human-like traits, which to some extent promotes user engagement and preference (Poushneh, 2021).

In this regard, a key focus of research is on how to specify appropriate personality constructs for voice assistants in various scenarios, as well as providing users with context-based VUI. For example, different voice personalization strategies depending on specific task objectives and contexts could enhance user experience (Braun et al. 2019). Similarly, some studies suggested that drivers can perceive the aggressiveness of a voice assistant based on differences in voice and language, and voice interactions with similar personalities to the vehicle user can result in higher user ratings and trust in the vehicle's performance (He & Burns, 2022). however, studies have shown that beyond use of voice assistants as general conversation agent with personality, users may concerned with the the primary task should be to complete properly in achieving specific goals by providing alerts and information. personalized voice assistants related to driving should consider the safety of the dialogue (Braun et al. 2019).

Furthermore, voice gender may affect people's interpretation of personality for VUI. Previous research has shown that while female voices are perceived as more capable and friendly, male voices are deemed more appropriate in safety-critical environments such as driving (Dong et al., 2020). Additionally, some studies suggest that as user acceptance of voice products increases, their preference for different voice genders tends to converge, indicating a diminishing impact of voice gender on user acceptance (He & Burns, 2022).

Situation-Adaptive methodology Human interlocutors adapt their language style to each other in interpersonal communication to interact efficiently (Pickering & Garrod, 1995), and following the Spoken Dialog Systems, voice user interface – as alternative of human talking– are expected to be more sensitive to simulate people talking style and reaction. Within this framework, Stier et al. (2020) investigated the extent to which situation-adaptive methods impact driving tasks and discovered that the complexity of driving is directly reflected in user language. When Speech production and preception was combined together usability was improved and cognitive load for drivers was reduced. The influence of the interaction context, as shaped by the situation, thus plays a crucial role in voice-based interaction and should be carefully considered in the design of output. Additionally, The importance of language processing in VUIs occurring parallel to primary tasks to avoid distractions is emphasized, highlighting a vital aspect of VUI functionality in multitasking environments (Stier et al., 2020).

2.2.4 Evaluation methods

Numerous studies were focused on enhancing performance in Voice User Interfaces (VUI) and their interaction experiences. It is imperative that designers and researchers concentrate on identifying key determinants and constructing theoretical frameworks that elucidate the adoption and utilization scenarios of VUIs. Efforts should be made to develop a comprehensive set of adaptable design frameworks. Some researchers introduced a design evaluation form that incorporates a tri-level indicator system, derived directly from natural language processing and user requirements. This system posits that five modules—Sensory, Resources, Interaction, Emotional, and Design—significantly influence user experiences in real driving contexts (Pu et al., 2023). Furthermore, Wang et al. (2021) advocated for a systematic modeling and analysis of the entire VUI usage process. They divided the operation into five components: awakening and turn-off, voice recognition, natural language understanding, voice synthesis, and human-machine interaction. Based on this segmentation, they proposed the development of evaluative models and standards tailored to various stages of user interaction. For example, in the human-machine interaction component, designers should consider the interaction rhythms in different command settings.

Usability Evaluation: In the HCI field, as evaluation methodologies expand for VUI usability, various instruments effectively capture user perceptions during technology interactions and can significantly indicate user experience outcomes, such as the SUS (System Usability Scale) and TAM (Technology Acceptance Model). The System Usability Scale (SUS) provides a straightforward method for subjective usability assessments. This scale comprises ten 5-point Likert scales, ranging from ”strongly disagree” to ”strongly agree,” as delineated in Table 5 (Brooke, 1996). Despite SUS was created for evaluate the useability for GUI, Empirical studies have validated the SUS’s applicability for VUI evaluations (Ghosh et al., 2018) and have led to the development of a specialized scale tailored to the nuances of voice-based communication, identified three critical dimensions—Usability, Affective Response, and Recognizability & Visibility—that are essential for standardizing and modeling user needs in VUI evaluation tools. (Zwakman et al., 2021).

Table 5: System Usability Scale (SUS) Questionnaire

No.	Statement
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to use this system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

User Experience Evaluation: With the development of evaluation approaches concerning user experience assessment, various user experience evaluation methods and tools have been refined to capture the experiential dimensions inherent in interactions with voice interfaces. Murad et al. (2023) conducted a comprehensive meta-analysis of 336 VUI design considerations sourced from academic literature. Through thematic analysis, they synthesized these design considerations into a consolidated set of 14 principles, detailed in Figure 4, providing a structured framework for researchers focused on design and evaluation. Additionally, Chen et al. (2023) analyzed consumer reviews of voice assistants and identified four primary user experience dimensions essential for improving measurement scales: utility experience, intelligence experience, personalized experience, and emotional experience. These dimensions are crucial for developing more effective and user-centric voice interfaces.

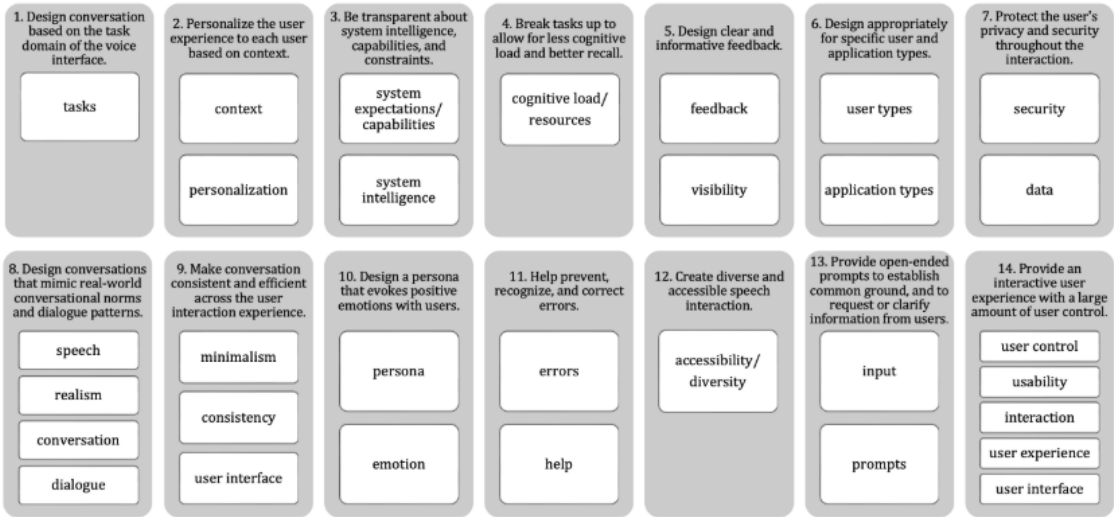


Figure 4: Final Synthesized Guidelines and Associated Axial Codes(Murad et,al. 2020).

Building upon the UEQ (User Experience Questionnaire), Klien et al. (2020) contributed to three specific scales for response behaviors, response quality and comprehensively for evaluating speech interaction to measure and explored the generalizability/applicability of this framework. Although the discussion focuses on specific uses or specific traits of VUI, the findings above offer valuable guidance for the development of comprehensive evaluation methods. These insights can help tailor design and evaluation frameworks to enhance the usability and user experience of Voice User Interfaces across various applications.

2.3 Summary

The literature review primarily explores two key areas: (1) user experience and product emotion research, and (2) VUI-related user behaviors, functionality optimization, and evaluation methods. By integrating user experience methods and analyzing emotional responses through detailed components from appraisal model, this approach allows for a deeper understanding of user interaction and the emotional changes involved in product use. A crucial aspect of this process is recognizing how product characteristics evoke emotions, ultimately shaping the overall user experience. By examining factors such as motive consistency, intrinsic pleasantness, and coping potential, designers can develop more engaging and satisfying products that align closely with user expectations and desires.

The role of VUIs in everyday technology interactions is expanding, necessitating a focus on naturalness in conversation, proactive behavior, and adaptive interaction strategies. The development and refinement of VUIs require careful consideration of these aspects to enhance usability and ensure seamless integration into daily tasks. Furthermore, the adoption of robust frameworks for the evaluation and design of VUIs will be crucial. These frameworks must not only address technical performance but also consider the broader experiential impacts, including emotional and psychological dimensions. By doing so, VUIs can become more than tools; they can be empathetic interfaces that enhance the quality of interactions in a technologically driven world.

This study specifically focuses on the automotive domain, where VUI parameters should adapt to the complex demands of driving. As automotive interfaces continue to evolve, VUI integration should prioritize intuitive, user-friendly systems that enhance both safety and functionality. Additionally, insights from designers and researchers play a crucial role in this development—not only in ensuring technological efficiency but also in aligning VUIs with user expectations and preferences to foster broader acceptance. For instance, designers are instrumental in making voice interactions more natural and in balancing task efficiency with engaging user experiences (Kim, Reza, McGrenere, & Yoon, 2021).

Despite these advancements, there remains a noticeable gap in the literature concerning specific research and design guidance on the user experience dimensions of VUIs. Building on previous research, this study seeks to address this gap by exploring the comprehensive user experience. By examining users' perceptions, emotions, and interactions with VUI in task completion, this study deepen our understanding of user experience and provide valuable insights for the design and development of intelligent assistant technologies. This foundational knowledge sets the stage for subsequent empirical investigations into the practical application and user acceptance of VUIs as replacements for physical controls.

3 User Studies

This chapter outlines the user study steps taken to carry out the project, from designing and evaluating concepts to conducting the studies, collecting and analyzing data.

Given that our study involves sensitive personal data, we ensured that participation was entirely voluntary and all responses would be treated with the utmost confidentiality. All participants provided informed consent and received appropriate compensation. The user research content was critically reviewed and revised by experts to ensure the professionalism and integrity of the study.

3.1 Research preparation

Before conducting user experience research, it is necessary to make preparations based on the content of the research. In this research, we were supposed to apply the qualitative study (interviews) and quantitative study (like surveys) to collect valid data for analysis, primarily focusing on a specific scenario where users engage in driving activities and controlling the car features.

Our exploration of user experience is grounded in the theoretical framework developed by Desmet and his colleagues, which integrates elements from the Appraisal Model and User-Centered Theory, collectively referred to as the Product Experience and Impact Design Theory. Questionnaires baselines designed to assess experience draw upon evaluative theory that corresponds to specific ”relationship questions” about stimulus events (Roseman, 2001; Scherer, 2001) and questionnaire questions would be identified by seven key evaluative components concluded by Demir and his colleagues (2009). The definitions and explanations of these components have been thoroughly examined in the section literature review. Through this preparation, we could investigate the different dimensions of user experience including the users' individual experience, preferences, encountered issues, and potential improvement suggestions, which could perform as factors in influencing their attitude and behaviors in using voice user interface to facilitate in-car features interaction instead of employing physical controls like buttons.

3.1.1 Research objects selection

Prior to discussing control methods and driving activities, an important question arises: Which car features should be selected, and what standards should guide this selection, particularly when considering whether these features are already installed? Moreover, the diversity in vehicle types, brands, models, designs, ergonomics, and powertrain variations can result in different sets of car features and interior layouts. Additionally, variations in terminology and definitions of car features may lead to inconsistencies in comparisons and differences in user understanding of functionality.

In this context, we could not select participants' personal vehicles or familiar vehicle models to examine the layout and usage frequency of car features and control methods. Instead, we focused on analyzing users' familiarity with and usage of common features with physical controls in vehicle interiors, as well as investigating their experience and frequency of using voice controls during driving activities. Furthermore, the data indicates that new electric car registrations accounted for 22.7% of total new car registrations in Europe, alongside a 4% increase in Plug-in Electric Vehicle registrations in 2023 (Agency, 2025). Our research would focus primarily on the application of this study in Battery Electric Vehicles (BEVs) and Plug-in Electric Vehicles (PEVs). In order to ensure these chosen features reflect the key features widely found in Europe's mainstream electric vehicles, ten BEV and PEV car models from different brands that are sold in the European market and have achieved certain sales volumes have been selected for research (Alternative Fuels Observatory, 2023). The well-selling electric vehicles in Sweden are also included. Additionally, with the support of Lynk&Co Sweden, we were able to conduct research on in-car voice user interfaces, along with insights into their vehicle models. The brands and models chosen for the survey are as shown in the Table 6:

Table 6: Overview of the best-selling BEV and PHEV models in the European market, including their drivetrain specifications and the presence of voice assistants.

Rank	**Model**	**Powertrain**	**Voice Assistant**
1	Tesla Model 3	Electric	Yes
2	Volvo XC40 (BEV+PHEV)	BEV+PHEV	Yes
3	MG4	Electric	Yes
4	Skoda Enyaq	Electric	Yes
5	Audi Q4 e-tron	Electric	Yes
6	Volvo EX30	Electric	Yes
7	BMW i4	Electric	Yes
8	Mercedes GLC PHEV	PHEV	Yes
9	Kia Niro (BEV+PHEV)	BEV+PHEV	Yes
10	VW ID.3	Electric	Yes
11	LYNK&CO 01	PHEV	Yes
12	Polestar 2	Electric	Yes
13	BYD Atto 3	Electric	Yes
14	KIA EV6	Electric	Yes
15	Toyota bZ4X	Electric	Yes

All the models listed in the table are equipped with voice assistants or voice interaction interfaces. The majority of these models enable voice control of in-car features through their voice interaction systems. Some models, such as the Tesla Model 3, are equipped with advanced voice interaction technologies. However, for many others, the functionality of the voice user interface is often limited to basic tasks, such as controlling the entertainment system.

Upon surveying the car models, it was observed that the majority of them have physical buttons or controls to operate and adjust specific features. Drawing upon the definitions of various features and their functionalities defined in the SAE International Dictionary of Automation Engineering (Kershaw, 2023) , alongside Pfleging's classification of features within the context of non-autonomous vehicles, a total of 16 features were identified and grouped into three categories for analysis: features directly related to driving, features indirectly related to driving tasks, and other features. The specific classification, corresponding features, and their definitions are presented in Table 7.

Table 7: Definition of Features (Kershaw, 2023) and Classification by Driving Task Categories

Feature	Feature Functionality Definition
Related to primary driving tasks:	
ACC system control	Activate or deactivate ACC and maintain desired safe distance.
Cruise Control Switch	Activate or deactivate and Set(decrease) and Resume(increase) work for speed control.
Intelligent/automatic parking assist system controls	Activate or deactivate automatic parking assist system.
Gear stick (lever/shift)	Manually change gears to control the vehicle's speed and torque output.
Related to secondary driving tasks:	
Drive mode switch	Switch for controlling different drive modes.
Electric Parking Brake (EPB) switch	Activate or deactivate Electric Parking Brake.
Side-view (wing) mirror control	Adjust vertically and horizontally to optimize visibility of the rear and side views of the vehicle.
Windshield washer and wiper controls	Activate or deactivate Windshield washer and wiper to clean and adjust the speed of wiper.
Exterior Headlights controls	Turn on or off and adjust light beam of headlights.
Interior Door Handles	Open or close the door from the inside.
Exterior Door Handles	Open or close the door from the outside.
Related to tertiary driving task:	
Steering wheel Adjustment	Adjusts the position of the steering wheel.
Glove compartment Control	Open and close car glove-box.
Air-recirculation Buttons	Activate or deactivate air recirculation system.
Entertainment system	Manage media, phone, navigation, and car status.
Sunroof Manual controls	Open and close panoramic sunroof.
Climate control system Operation	Activate or deactivate climate control system and adjust desired temperature.
Trunk/Boot Controls	Open and close truck.

3.1.2 Planning user research

This section elaborates in detail on the specific design on content of the survey and interview questions that are crafted based on the theories discussed previously, which aims to capture a snapshot of the current user experience, users' expectations and the potential improvements of in-vehicle voice interaction technology.

3.1.2.1 Questionnaire questions design

The questionnaire is primarily designed for a comprehensive assessment that includes collecting demographic data, evaluating familiarity/usage of VUI and physical controls, and satisfaction with in-vehicle voice interaction systems, and quantifying expectations for replacing physical controls with voice interaction. Several components were structured, including:

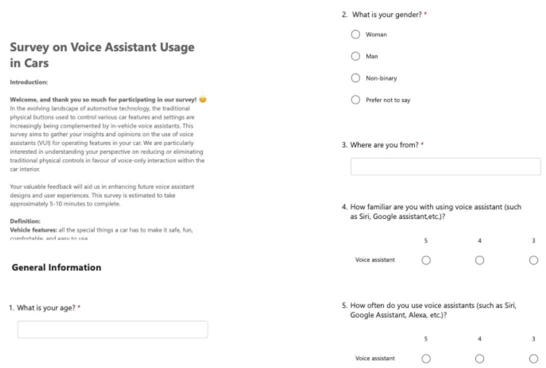


Figure 5: User study questionnaire

- **Demographic Information:** This part gathers basic details such as age, gender, participants' driving behaviors (such as driving experience and frequency) and basic knowledge on voice user interface. which serve to understand the basic user profiles.
- **Usage Frequency and Proficiency of both control methods:** This part of the survey evaluates how frequently participants use the two control methods while driving and examines their understanding and proficiency in operating these features.
- **User experience and Satisfaction with Voice Interfaces:** A series of questions was designed to evaluate user experiences regarding the voice user interface's effectiveness in completing intended tasks, user preferences, and concomitant emotional changes. Example questions include: "Do you think the voice user interface helped you complete the intended tasks?", "Do you enjoy using the voice assistant?", and "Did the functionalities meet your expectations?" According to the research objective, we selected five key components from the adjusted appraisal model.

Likert scale, which was devised and tailored, had been applied in collecting behavioral and user experience questions responses data for measuring. Furthermore, an increasing number of battery electric vehicles (BEVs) are reducing physical controls and integrating functionalities into touchscreens. It is essential to understand whether this emerging trend influences participants' expectations regarding alternative control interfaces, such as voice interaction interface. We propose the following question: *What are users' attitudes toward replacing physical controls with in-car voice interfaces, and are they willing to use voice interfaces instead of physical buttons for controlling in-car functions?*

3.1.2.2 Interview questions design

The formulation of the interview questions targeted several critical dimensions:

- **Motivation and contextual Usage:** These questions probe the user circumstances and reasons behind using the voice user interface during driving, such as, "What tasks do you typically accomplish using the voice assistant while driving? and What motivates these choices?"
- **Users experience:** This section of the interview questions investigate user experience, VUI performance and problems, and some instances of satisfied or unsatisfied usage. For instance, participants might be asked: 'Which aspects of the voice user interface met your expectations, and which fell short?'
- **Substitutability and Future Expectations:** The discussion focused on the feasibility of replacing physical controls with voice user interfaces during driving and users' aspirations for advancements in voice interaction. Key questions included: 'Which physical controls, if any, do you think could be fully replaced by a voice assistant?' and 'What improvements or new features would you like to see in future in-car voice user interface?'

3.2 Quantitative study

3.2.1 Data collection via questionnaire

The questionnaire was developed using *Microsoft Forms* and offered in two language versions: English and Chinese. Both versions were professionally reviewed and refined by researchers and two experts to ensure consistency in wording and translation accuracy.

The questionnaire was distributed using both online and onsite recruitment methods. The Chinese version was distributed via social media platforms, such as *WeChat*, and Chinese survey platforms like *Wenjuanxing*. The English version was shared through channels like *Facebook* and data collection services such as *Prolific*. Invalid responses were eliminated, Chinese results were translated into english, the remaining data were consolidated to facilitate a streamlined and comprehensive analysis.

3.2.2 Hypothesis development

In the quantitative study and analysis, it is crucial to identify the independent variables that prominently influence people's preferences for control features. Based on these independent variables, hypotheses for statistical analysis will be defined. Specifically, to understand the factors influencing users' preferences for control features and their acceptance of the potential substitution of physical controls with voice interaction, the following hypothesis is proposed to explore the relationship between the independent variables listed in Table 8 and the dependent variable: *Users' attitudes toward relying exclusively on VUI for controlling in-car features, without utilizing physical controls.*

Null Hypothesis (H_0): These factors defined in the **Table** 8 (e.g., understanding of physical controls, preference for in-car VUI) have no significant effect on whether users prefer to only set up voice user interface to control features.

Alternative Hypothesis (H_1): Some Independent variable (as showed in Table 8) has a significant effect on whether users prefer to only set up voice user interface to control features.

Table 8: Explanation of Each Item Related to In-Car Functions and Voice Assistant Usage

Item	**Explanation**
Understanding of in-car features	This refers to the degree of users' familiarity and understanding of in-car features, as previously defined. It reflects how well users have mastered these features and whether they are familiar with operating through different control methods.
In-car Feature Usage Frequency	This refers to how often users employ features during daily driving activities. A high frequency of use might indicate a higher dependency on a particular function.
Future expectation on Voice controls	This item refers to the degree of users' expectation for using the in-car voice interaction to perform operations.
User Experience with Current VUI Use	This refers to the assessment of user experience when utilizing voice user interfaces (VUI) to perform feature operation tasks.
Awareness of Features Controllable via Voice interaction	This refers to users' knowledge of in-car features that can be operated through voice interaction, including their understanding of available functionalities and tasks supported by voice control.
Preference for control methods	This item refers to users' preferences for control methods when manipulating features. It reflects their inclinations and favored approaches to interacting with in-car features.

3.2.3 Statistical analysis

In the first stage of analysis, Principal Component Analysis (PCA) was employed to reduce the dimensionality and extract key information from the five metrics derived from assessments of in-vehicle Voice User Interface (VUI) experiences. Prior to the PCA, essential pre-processing steps were taken to ensure that all variables contributed equally to the analysis. The suitability of the data for PCA was evaluated initially by examining the correlations among variables, followed by the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's Test of Sphericity to assess the appropriateness of the sample data. Once the data was deemed suitable, the covariance matrix was computed, from which eigenvalues and corresponding eigenvectors were extracted. These eigenvalues and eigenvectors form the core of PCA, assisting in the identification of principal components that account for the maximum variance. Principal components were selected based on a cumulative contribution rate threshold of 85% or more (Jolliffe, 2002), with retained components considered key indicators of user assessments of VUI experiences, feeding into subsequent analyses.

In the second phase, firstly the Scores for preference, understanding, and awareness of features within each driving task category were treated as homogeneous and suitable for averaging. This step simplified the data structure by consolidating data within the same group, thereby facilitating subsequent computations and reducing random data fluctuations to enhance the stability and efficacy of statistical analyses. Afterwards, Multivariate Linear Regression (MLR) was applied to analyze the impact of predetermined factors on user preferences for VUI controlling features across three types of driving tasks: primary driving tasks (PDT), secondary driving tasks (SDT), and tertiary driving tasks (TDT). The analysis incorporated fixed effects such as user perception of features, preferences for control (physical buttons or voice assistants), as well as user experience of the different interface (emotions like pleasure, safety, convenience). Models were independently established for each category of driving task to evaluate how these factors influenced user choices both individually and collectively. The covariance structure of residuals was modeled as a first-order autoregressive matrix with homoscedastic variances to address temporal correlation and potential autocorrelation in measurements. Additionally, each model included random intercepts and slopes for individuals across different contexts, employing a scaled identity matrix for the covariance structure.

All these process was conducted in the SPSS 29.0.2.0. Through this comprehensive analysis, regression analysis not only revealed the specific impacts of different user interface features on driving task performance but also addressed dependencies from repeated measures and imbalances due to missing data.

3.2.4 Statistical model

3.2.4.1 Principle component analysis

Principal Component Analysis could be acted as a simple, non-parametric method for extracting relevant information from confusing data sets (Shlens, 2014). This statistical technique perform a linear transformation to reduce the dimensionality of dateset. This process involves calculating the covariance matrix Σ of the data, then extracting its eigenvectors and eigenvalues. The eigenvectors define new orthogonal axes for the data, while the magnitude of the eigenvalues indicates the importance of each principal component. By retaining the eigenvectors associated with the largest eigenvalues, PCA captures the major variations within the data, thus achieving dimensionality reduction. The transformation projects the data onto a new subspace defined by the principal components, which are the directions of maximum variance in the data. In this case, different dimensionality of user experience data has been reduced into one main dataset to represent.

3.2.4.2 Multiple linear regression analysis

Regression analysis is performed to determine the correlations between different variables having cause-effect relations, and to make predictions for the topic by using the relations(Uyanık & Güler, 2013). In the explanatory case, regression models could be functioned to study how the customers make decisions or form impressions and attitude. Multiple Linear Regression (MLR) is a statistical method used to model the relationship between a dependent (criterion) variable and valuable multiple independent (predictor) variables which be weighted by regression analysis procedure to ensure maximal prediction (Black & Babin, 2019). The MLR model is represented by the following equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

where:

- Y: Dependent variable
- $X_1, X_2, \ldots, X_k$: Independent variables
- $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$: Coefficients of the independent variables
- ε: Error term, accounting for deviations from the model

The goal of multiple linear regression is to estimate Y based on the independent variables $X_1, X_2, \ldots, X_k$ and determine the impact of each variable. The method typically employs the least squares approach to minimize the sum of squared residuals, ensuring the best possible fit of the model. In this study, regression analysis was applied to determine whether the identified independent variables significantly influence individuals' behaviors and attitudes.

3.2.5 Quantitative study results

A total of 408 questionnaires were collected, of which 398 were completed, and 395 had all sections fully filled out. Incomplete questionnaires were excluded from the analysis, resulting in a final dataset of 390 valid responses. To ensure the accuracy and reliability of subsequent analyses, all invalid questionnaires were removed, and the data from the retained sample were used for further examination.

Among the valid sample data, there were a total of 174 female respondents, accounting for 44.6%, and 207 male respondents, making up 53.1% of the sample. A total of 237 respondents were from China, 78 from Sweden, and 60 from the Netherlands and other European countries. The average driving experience among the respondents was 5.9 years, with 43.1% of them using or driving a car at least once a week. Regarding familiarity with or usage of voice interaction system in daily life, more than 345 respondents indicated they were familiar with such technology, suggesting that the majority of respondents were either aware of or had hands-on experience. Furthermore, 60.8% of participants had used car voice interaction system (Voice assistant) at least once.

3.2.5.1 User experience on in-car VUI

Participants were asked to evaluate their experience with Voice User Interfaces (particular in-car voice user interface which include both vehicle-integrated systems and other devices installed for in-car use). The majority of respondents indicated that they had previously or currently used VUIs to complete tasks within the vehicle. Specifically, 129 respondents reported that they needed to use a voice assistant at least once a week to assist them in completing tasks.

Across the five dimensions of user experience evaluation, the average score was $M = 2.78$, indicating a moderate level of satisfaction or agreement with the VUI experience. When respondents were asked whether the VUI was pleasant/enjoyable to use, 23.1% indicated that they found the VUI to provide a pleasant experience, while 31.5% expressed disappointment with the VUI's performance during use. This suggests that respondents have reservations about the maturity of current VUIs, potentially due to negative experiences during use that prevent them from fully endorsing the VUI as capable of controlling the features correctly. 30.8% of respondents stated that the VUI either completely or largely failed to meet their expectations, while only 20.7% indicated that the VUI met their anticipated standards for functionality. 25.4% of respondents felt that the existing features adequately met their needs, whereas 117 users expressed disappointment that the VUI failed to align with their standards. Regarding the experience of using the VUI to complete tasks, 52.5% of respondents believed that the VUI performed well in executing their commands, but the remaining respondents expressed concerns about the VUI's ability to execute more complex functions effectively. 54.6% of respondents reported that they were clear or fairly knowledgeable about the operations and outcomes when using the VUI system, while 10.8% indicated a lack of certainty regarding the results of their interactions with the VUI.

Principal Component Analysis (PCA) was conducted to identify positive correlations among the user experience dimension variables. Data reliability was tested using Cronbach's alpha, yielding a value of 0.904, indicating excellent internal consistency. To assess the suitability of the data for PCA, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was calculated, resulting in a value of 0.862, which is considered excellent and suggests that the data is highly suitable for factor analysis (Kaiser, 1974). Additionally, Bartlett's test of sphericity was conducted to determine whether the correlation matrix is an identity matrix, which would suggest that the factor model is inappropriate. The results were significant at a level less than 0.001, confirming that the correlations among (Tabachnick, Fidell, & Ullman, 2013). In Table 9 Commonalities showed an average value of 0.724, indicating that each variable shared a substantial amount of variance with other measured variables. All items in the scale had a communality of 0.65 or higher.

If two or more variables exhibit a high positive correlation, this may suggest that they mea-

Table 9: Commonalities for User experience Components in PCA

Appraisal Component	Question	Extraction
Motive consistency	Does this help me in attaining your goals and motives?	0.658
Intrinsic pleasantness	Is this pleasant or enjoyable?	0.788
Expectation confirmation	Does this confirm your expectations?	0.705
Standards conformance	Does this match with your social norms and standards?	0.777
Certainty	How certain are you about this voice user interface?	0.691

Extraction Method: Principal Component Analysis.

sure similar or identical underlying constructs, which could be considered for consolidation into a single indicator. Based on the correlation matrix, the variable “Is this pleasant or enjoyable?” showed a strong positive correlation with all other variables, with Pearson correlation coefficients ranging from 0.661 to 0.733. These correlations were significant at the 0.01 level ($p < 0.01$), indicating a less than 1% probability that these correlations were due to random variation. Showed in Table 10, The analysis revealed that the first principal component explained 72.382% of the total variance, with an eigenvalue of 3.619, a substantial proportion. The component matrix showed high loadings for all variables on this principal component (all above 0.8), indicating strong positive correlations among the variables. This suggests that these variables may be measuring similar constructs.

Table 10: Variance Explained by Principal Components (PCA Results)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.619	72.382	72.382	3.619	72.382	72.382
2	0.542	10.838	83.220			
3	0.346	6.924	90.144			
4	0.274	5.484	95.628			
5	0.219	4.372	100.000			

The analysis suggests that overall enjoyment is a central aspect of the user experience and could potentially serve as a representative measure of the emotions and attitudes users hold toward the VUI across various dimensions of user experience. Although the item ”Is this pleasant or enjoyable?” showed strong correlations with other variables, none of the correlations were perfect (i.e., none reached 1.0).

3.2.5.2 Factors on user preference in primary driving tasks features

Regarding primary driving tasks, respondents demonstrated varying levels of familiarity and usage frequency with different functions. Among the driving control functions, over 86.1% of respondents reported understanding the operational logic and usage steps of major driving-related functions and how to activate or deactivate these functions using physical controls. Specifically, 96.9% of respondents were familiar with the working principles of gear shift paddles and used them after becoming acquainted with the different gear positions. Additionally, 47.7% and 55.4% of respondents were familiar with how to control automatic parking and cruise control via physical buttons, with 20.8% and 22.3% of respondents reporting that they frequently used these features. However, for Adaptive Cruise Control (ACC), 33.8% of respondents stated that they were familiar with the functionality, but only 14.6% reported having used or had relevant experience with it.

When asked whether they were aware of using a Voice User Interface (VUI) to control these functions, only 7.7% of respondents indicated being highly knowledgeable about automatic parking via VUI, respectively. Due to current technological limitations and safety requirements, neither ACC nor gear shift can be controlled by voice commands, and no such cases have been documented.

For ACC and gear shift paddles, 10.5% and 5.4% of respondents indicated moderate support for using voice control alongside physical buttons. Additionally, 51.5% of respondents indicated no objection to using a hybrid control scheme for automatic parking. For the cruise control function, 26.2% of respondents expressed support for voice control but insisted on retaining the physical button controls.

A regression analysis was conducted on the data that were filtered accordingly. We set users' behavior, familiarity with features (including their physical controls) and voice interfaces, User experience on both control way, preferences, and expectations as independent variables, and the respondents' tendency to control driving functions using only the VUI as the dependent variable in a multiple linear regression analysis. To check for multicollinearity among the variables, independent variables condition indexes and Variance Inflation Factors (VIF) were examined. The results, through Pearson Correlation Coefficient analysis, showed that no correlation coefficient exceeded 0.80, indicating no multicollinearity between variables. A VIF value equal to or greater than 10 would suggest multicollinearity (James, Witten, Hastie, Tibshirani, et al., 2013). As seen in Table 11b, all VIF values were below 10, leading to the conclusion that multicollinearity was not present. The independent variables could be considered there are no multiple relations between variables if their CIs are no bigger than 30 (Uyanik, 2013). The results, showed in Table 11a, indicated that there are no multiple relations between variables.

The model test results showed that the model was statistically significant, indicating that the specified factors significantly influenced respondents' choices [$F(21, 369) = 17.491$, $p < 0.001$, $R^2 = 0.460$, Adjusted $R^2 = 0.434$, Durbin-Watson = 1.867]. The multiple linear regression analysis confirmed that the interaction between user behavior, knowledge of features/voice interaction, user preferences, and expectations, and the final choice of control method was statistically significant ($p < 0.05$). However, it is important to note that the independent variable *User Experience* was not statistically significant, indicating that it has a limited impact on people's choices.

Table 11: Combined Analysis of Collinearity Diagnostics and Multiple Linear Regression results for Primary driving task related features Model

(a) Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions						
				(Constant)	Understanding	Feature Usage	VUI controls	Multi-Controls	Preference	User Experience
Primary Driving Task Related Features	1	6.393	1.000	.00	.00	.00	.00	.00	.00	.00
	2	.218	5.421	.00	.00	.05	.02	.00	.25	.05
	3	.134	6.907	.01	.01	.17	.18	.00	.00	.43
	4	.090	8.445	.00	.00	.24	.64	.01	.33	.03
	5	.084	8.736	.08	.06	.04	.14	.28	.13	.33
	6	.054	10.880	.04	.33	.41	.01	.37	.27	.08
	7	.028	15.112	.87	.55	.09	.01	.28	.00	.08

[a] Dependent Variable: Only control in VUI

(b) Multiple linear Regression Analysis related to People's solely controlling by VUI

Model Components	Unstandardized B	Std. Error	Standardized Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
(Constant)	0.484	0.395	-	1.226	0.223	-	-	-	-	-
Feature Understanding	-0.222	0.094	-0.197	-2.370	0.019	-0.033	-0.209	-0.157	0.635	1.576
Feature Usage	0.269	0.095	0.240	2.814	0.006	0.303	0.246	0.186	0.603	1.659
Familiarity of VUI on feature controls	0.351	0.099	0.287	3.564	0.001	0.512	0.306	0.236	0.676	1.479
Control Preference	0.301	0.086	0.286	3.511	0.001	0.534	0.302	0.233	0.661	1.514
Expectation on Voice Controls	0.076	0.038	0.167	2.005	0.047	0.496	0.178	0.133	0.636	1.572
User experience of VUI	-0.026	0.075	-0.024	-0.348	0.728	0.044	-0.031	-0.023	0.940	1.064

The model's goodness of fit, including the R^2 and adjusted R^2 values, standardized regression coefficients (Beta), and significance levels for each variable, is presented in Table 11b . Results also illustrated the regression coefficients and statistical significance results for each factor, where the degree of user preference for using VUI is the dependent variable.

Based on the regression analysis, the following regression equation was derived:

$$Y = 0.484 - 0.222 \cdot X_1 + 0.269 \cdot X_2 + 0.351 \cdot X_3 + 0.301 \cdot X_4 + 0.076 \cdot X_5 - 0.026 \cdot X_6$$

Where:

1. Y: **Dependent variable**
2. X_1: **Feature Understanding**
3. X_2: **Feature Usage**
4. X_3: **Awareness of Features Controllable via Voice interaction**
5. X_4: **Preference for control methods**
6. X_5: **Future Expectation on Voice controls**
7. X_6: **User Experience on VUI**

The absolute values of the standardized regression coefficients (Beta) indicate the relative importance of each independent variable. According to the results, users' awareness of VUI-related features had the greatest impact on the dependent variable (Primary Driving Task related Features) (Beta = 0.287), followed by the frequency of feature usage (Beta = 0.240) and user preferences (Beta = 0.167). Awareness of the feature (Beta = -0.197) showed a negative correlation, suggesting that higher awareness scores may decrease the likelihood of choosing to use the VUI.

3.2.5.3 Secondary and other driving tasks related features factors

Regarding the features related to secondary driving tasks, respondents were generally familiar with the usage of each function and its corresponding physical control buttons (*mean* = 3.901). However, familiarity with the Driving Mode switch function and the Electronic Parking Brake (EPB) function did not exceed 85%, with percentages of 66.9% and 63.8%, respectively. Similarly, for the functions associated with other driving tasks, 90% of respondents reported familiarity with the Climate System Control due to its frequent use in daily life.

In terms of usage frequency, respondents indicated that the door handle and exterior light controller were more frequently used among secondary driving task functions, while the Climate System Control and Entertainment System for other driving tasks had higher usage rates, at 75.4% and 76.9%, respectively. When asked about their familiarity with using the Voice User Interface (VUI) to control secondary driving activities, the percentages of users who indicated being fairly knowledgeable or above were 23.1% for mirror adjustment, 18.5% for steering wheel adjustment, 33.8% for wiper control, 33.1% for headlight control, and 24.6% for door control. For other driving tasks, only the Entertainment System Control and Climate System Control showed a relatively high level of familiarity, at 51.5% and 41.5%, respectively.

For secondary driving functions, users showed a stronger preference for hybrid control schemes (mean = 3.389), particularly for windshield wiper and automatic cleaning, exterior light control, and side mirror control, with 56.2%, 63.6%, and 54.6% of respondents, respectively, indicating fairly supportive or higher levels of preference. In contrast, for door switch control, 36.3% and 35.4% of respondents expressed disagreement or strong disagreement. A similar pattern was observed for other driving task functions, where users preferred a hybrid control approach when using VUI (mean = 3.6253), with 65.4%, 70%, 66.9%, and 63.1% of respondents accepting VUI for circulation system, media control, climate system control, and sunroof control, while retaining physical controls.

A multiple linear regression analysis was conducted for both secondary and tertiary driving task related features, with user behavior, familiarity with the functions/voice interface, user experience, and user preferences set as independent variables, and respondents' inclination to use only the voice user interface (VUI) for driving function control as the dependent variable. In both models, the correlation coefficients between independent variables were all less than 0.80, and the variance inflation factors (VIF) were below 10, indicating no multicollinearity among the variables.

The model testing results provide statistically significant evidence that, for the secondary driving task function model, the predetermined factors significantly influenced respondents' choices [$F(21,369) = 15.265$, $p < 0.001$, $R^2 = 0.418$, Adjusted $R^2 = 0.389$, Durbin-Watson = 1.678], suggesting that these factors can explain a significant portion of the variance in the dependent variable. The findings,showed in Figure12b indicate that user expectations for VUI ($\beta = 0.286$, $p = 0.001$), familiarity with voice-controlled functions ($\beta = 0.276$, $p = 0.001$), and preferences for control schemes ($\beta = 0.227$, $p = 0.010$) have a significant positive impact on the choice to use VUI exclusively for function control. These results suggest that the higher the user's expectations for voice assistant control functions and the greater their familiarity with voice interfaces and functions (physical controls), the stronger their preference for using voice control functions. The level of functional understanding ($\beta = -0.164$, $p = 0.084$), although not significant at the traditional level ($p < 0.05$), approaches significance, indicating that the degree of functional understanding may negatively influence choice preferences.

Table 12: Combined Analysis of Collinearity Diagnostics and Multiple Linear Regression results for Secondary and Tertiary driving task related features Model

(a) Collinearity Diagnostics for independent variables in analyzing Secondary and Tertiary driving task related features

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions						
				(Constant)	Understanding	Feature Usage	VUI controls	Multi-Controls	Preference	User Experience
Secondary Driving Task related Features	1	6.484	1.000	.00	.00	.00	.00	.00	.00	.00
	2	.206	5.615	.00	.01	.02	.03	.03	.28	.09
	3	.104	7.896	.00	.00	.01	.64	.00	.21	.20
	4	.094	8.318	.03	.02	.03	.23	.07	.01	.58
	5	.066	9.934	.02	.01	.17	.01	.40	.44	.05
	6	.030	14.788	.38	.07	.45	.07	.44	.05	.07
	7	.017	19.641	.56	.88	.30	.02	.06	.01	.00
Tertiary Driving Task related Features	1	6.528	1.000	.00	.00	.00	.00	.00	.00	.00
	2	.176	6.083	.00	.02	.01	.00	.01	.51	.10
	3	.101	8.048	.00	.02	.05	.17	.01	.05	.61
	4	.086	8.691	.08	.01	.01	.24	.22	.05	.21
	5	.059	10.486	.00	.04	.15	.44	.25	.29	.01
	6	.028	15.139	.48	.04	.43	.13	.48	.03	.03
	7	.021	17.747	.44	.86	.35	.01	.02	.06	.02

(b) Merged Multiple Linear Regression Analysis related to People's solely controlling by VUI with Model Category

Model Category	Model Components	Unstandardized B	Std. Error	Standardized Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
SDT Model	(Constant)	1.022	0.462	-	2.210	0.29	-	-	-	-	-
	Feature Understanding	-2.18	0.125	-0.164	-1.741	0.084	-0.035	-0.155	-0.120	0.535	1.868
	Feature Usage	0.132	0.113	0.111	1.170	0.244	0.103	0.105	0.080	0.527	1.898
	Familiarity of VUI on feature controls	0.281	0.086	0.276	3.252	0.001	0.472	0.281	0.224	0.656	1.525
	Control Preference	0.241	0.092	0.227	2.627	0.01	0.509	0.231	0.181	0.632	1.581
	Expectation on Voice Controls	0.126	0.038	0.286	3.320	0.001	0.533	0.287	0.228	0.637	1.569
	User experience of VUI	-0.091	0.078	-0.085	-1.169	0.245	-0.020	-0.105	-0.080	0.886	1.129
TDT Model	(Constant)	0.669	0.422	-	1.586	0.115	-	-	-	-	-
	Feature Understanding	-2.22	0.094	-0.197	-2.370	0.019	-0.033	-0.209	-0.157	0.635	1.576
	Feature Usage	0.269	0.095	0.240	2.814	0.006	0.303	0.246	0.186	0.603	1.659
	Familiarity of VUI on feature controls	0.351	0.099	0.287	3.564	0.001	0.512	0.306	0.236	0.676	1.479
	Control Preference	0.301	0.086	0.286	3.511	0.001	0.534	0.302	0.233	0.661	1.514
	Expectation on Voice Controls	0.154	0.036	0.353	4.281	¡0.001	0.566	0.360	0.292	0.684	1.463
	User experience of VUI	-0.017	0.076	-0.016	-0.224	0.823	0.097	-0.020	-0.015	0.896	1.116

In the tertiary driving task function model, statistically significant evidence also supports the hypothesis [F (21, 369) = 17.491, $p < 0.001$, $R^2 = 0.460$, Adjusted $R^2 = 0.434$, Durbin-Watson = 1.867). Similar to the findings for secondary driving task functions, user expectations for VUI ($\beta = 0.286$, $p = 0.001$), familiarity with voice-controlled functions ($\beta = 0.276$, $p = 0.001$), and preferences for control schemes ($\beta = 0.227$, $p = 0.010$) exhibited a significant positive impact. The model fit indices in the regression analyses (e.g., R^2 and Adjusted R^2), standardized regression coefficients (Beta), and significance levels of the independent variables are presented in Table 6. Table 7 illustrates the regression coefficients and statistical significance results for each factor as an independent variable against the degree of user preference for VUI as a dependent variable.

Based on the regression analysis, the following regression equations for Secondary and tertiary driving task related features were derived:

SDT related Features:

$$Y_{\text{SDT}} = 1.022 - 0.218 \cdot X_{1,\text{SDT}} + 0.132 \cdot X_{2,\text{SDT}} + 0.281 \cdot X_{3,\text{SDT}} + 0.241 \cdot X_{4,\text{SDT}} + 0.126 \cdot X_{5,\text{SDT}} - 0.091 \cdot X_{6,\text{SDT}}$$

TDT Related Features:

$$Y_{\text{TDT}} = 0.669 - 0.079 \cdot X_{1,\text{TDT}} + 0.058 \cdot X_{2,\text{TDT}} + 0.242 \cdot X_{3,\text{TDT}} + 0.264 \cdot X_{4,\text{TDT}} + 0.154 \cdot X_{5,\text{TDT}} - 0.017 \cdot X_{6,\text{TDT}}$$

Where:

- Y_{SDT}: **Dependent variable for SDT Model**
- Y_{TDT}: **Dependent variable for TDT Model**
- $X_{1,\text{SDT/TDT}}$: **Feature Understanding (SDT/TDT)**
- $X_{2,\text{SDT/TDT}}$: **Feature Usage (SDT/TDT)**
- $X_{3,\text{SDT/TDT}}$: **Awareness of Features Controllable via Voice interaction (SDT/TDT)**
- $X_{4,\text{SDT/TDT}}$: **Preference for control methods (SDT/TDT)**
- $X_{5,\text{SDT/TDT}}$: **Future Expectation on Voice controls (SDT/TDT)**
- $X_{6,\text{SDT/TDT}}$: **User Experience on VUI (SDT/TDT)**

In both multiple linear regression models, the regression coefficients for the variable related to user experience did not reach statistical significance. In the first model, the standardized regression coefficient for user experience was -0.085 ($p = 0.169$); in the second model, it was -0.017 ($p = 0.823$). This indicates that, at this stage, the User experience with in-car VUI is not considered a significant factor influencing people decision-making on selecting VUI as the only way to control the features .

3.2.6 Conclusion

The quantitative study reveals that while users acknowledge the potential of Voice User Interfaces (VUIs) to simplify operations and enhance driving safety, several critical gaps limit their adoption. Familiarity with and understanding of both physical controls and VUI systems significantly shape user preferences. For primary driving tasks, physical buttons remain the preferred control method due to their reliability and user trust. However, users expressed moderate openness to VUI integration for secondary and tertiary tasks, provided these systems improve recognition accuracy and contextual adaptability.

3.3 Qualitative study

Before conducting the formal interviews,the interview questions and procedures were reviewed with user experience experts, and a pilot study was conducted to assess the feasibility of the process and identify potential issues. For the formal interviews, twelve participants from diverse backgrounds were recruited. They had varying levels of driving experience and differing degrees of familiarity with and usage of in-car voice interaction interfaces.

The interviews were conducted both onsite and online, with online sessions facilitated via Zoom, and all sessions were audio-recorded. During the interviews, questions could be adjusted based on the participants' knowledge and experience, or supplemented by follow-up inquiries where appropriate. At the end of the interviews, the recording ceased and participants had the opportunity to ask questions and share their comments on this project. The duration of the interviews, excluding preliminary and concluding conversations, ranged from 30 to 50 minutes. The average duration of the interviews was 43.8 minutes. After the interviews concluded, the audio recordings were transcribed into text using professional transcription software.

3.3.1 Demographic

In this qualitative study, 16 participants were recruited, including 9 males (56.25%) and 7 females (43.75%), with an age range of 18 to 53 years (Age *Mean* = 32.13, Age *SD* = 9.94). Specific participant details are provided in Table 13. Although the sample size is relatively

Table 13: Detailed information of the individuals interviewed

Participants	Gender	Age	Driving years	Participants	Gender	Age	Driving years
1	Female	20	1	9	Male	45	3
2	Female	24	3	10	Male	19	1
3	Male	25	6	11	Male	53	20
4	Male	28	5	12	Female	37	7
5	Male	34	10	13	Male	29	4
6	Female	21	1	14	Female	24	2
7	Female	32	2	15	Male	35	16
8	Female	46	20	16	Male	42	20

small, it falls within the recommended range and is sufficient to obtain in-depth insights into the research questions (Saunders, Townsend, et al., 2018). To ensure diversity within the sample, participants were drawn from a variety of occupational backgrounds, including students, corporate employees, and freelancers. The majority of participants (75%) had received higher education (i.e., bachelor's degree or above), and 87.5% reported having used products with voice-interactive interfaces (such as Siri). Among those who had experience with voice user interface products, 71.4% of participants had used voice to control features within a vehicle (such as Navigation).

3.3.2 Processing

The audio recordings and verbatim transcriptions of the interviews were conducted by the author, with all data undergoing de-identification to protect participants' privacy. Subsequently, the author utilized Atlas.ti software (version 29.0) to analyze the transcriptions using a hybrid approach combining inductive and deductive thematic analysis methods (V. Braun & Clarke, 2006). This approach provides a practical and effective procedure for identifying, analyzing, and reporting themes (or patterns) within qualitative data (Attride-Stirling, 2001; V. Braun & Clarke, 2006).

In thematic analysis, themes can be identified using inductive, deductive, or a combination of both methods. Inductive analysis focuses on extracting themes directly from the raw data, while deductive analysis is based on theoretical or analytical frameworks, where coding elements often originate from pre-existing conceptual frameworks. The combination of inductive and deductive approaches enhances the rigor and comprehensiveness of contextual analysis.Therefore, we applied hybrid inductive-deductive analysis method to conduct two phases of coding and analysis on the factors influencing user interactions with VUIs (voice user interfaces) and driving features control, ensuring that all relevant content was captured and considered during the analysis process. In the first phase, a preliminary reading of the transcribed text

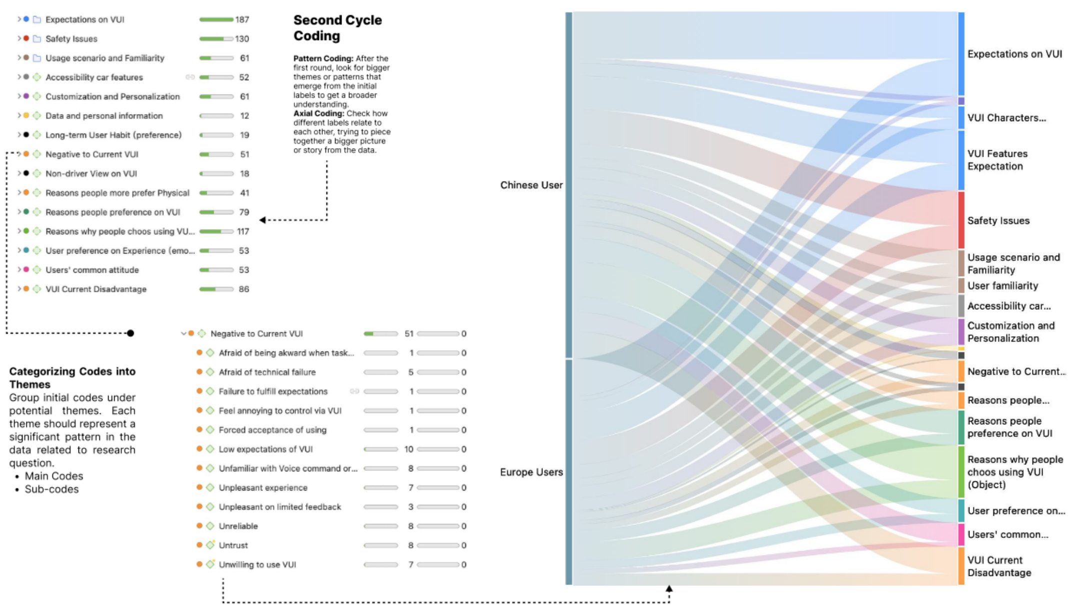


Figure 6: The workflow of thematic analysis

was performed, and the data was segmented based on the participants' opinions and narratives. Initial codes (sub-nodes) were derived inductively from the raw data, with corresponding labels created for each data fragment, forming an initial coding set. In the second phase, relationships between codes were further identified, and the data was organized according to the attribute classification methods outlined in the literature. Key core codes were extracted to form a core coding set. At this stage, codes that did not form coherent patterns were removed from the initial coding set. Figure 6 illustrates the workflow of the process. Themes (parent nodes) were then established based on the refined codes, and the candidate themes were evaluated to determine whether they 'adequately capture the contours of the coded data (V. Braun & Clarke, 2006).

Additionally, an external researcher independently reviewed the codes and themes at this stage to ensure their validity and consistency.

3.3.3 Results

This section outlines the qualitative analysis results which revealed key themes and insights into user experiences and preferences. These findings were derived from semi-structured interviews and observational data, focusing on user interactions with the system and their subjective feedback.

3.3.3.1 Perception of VUI user experience

In this case, the user's expectation standard was defined as the benchmark or reference point set by users based on the product or service's performance during its usage. Among the participants, most of them stated that the current voice interaction interfaces and systems basically meet users' needs and influence their preference for using it .

A majority of participants expressed a preference for using voice commands to manage secondary and tertiary driving functions, particularly for information systems (93.75%), climate control (37.5%), and navigation (31.25%). For example, one user mentioned:

> *"I definitely prefer using voice commands for things like navigation or air conditioning. Over time, I've gotten so used to voice control—it's second nature now. I just say the destination, and it finds the place for me, which is especially helpful when I'm in a rush to find a parking spot near where I'm going."*

Participants highlighted several positive aspects of using voice interaction systems, emphasizing their ability to reduce cognitive load and psychological pressure in various ways. First, Using voice to control could minimize the number of steps required, thereby decreasing the attention needed for operation. One participant explained, *"When playing a saved song, I just said, 'Hey, I want to play my saved songs', and it responded right away"*, while another shared *"It helps me park... sure, I still have to double-check the spot myself... but I don't have to stress about messing up the parking."* Secondly, it reduces interference with feature control, especially when the user's hands were occupied. For instance, a user noted, *"While driving, I just say the destination to set the navigation."* This functionality prevents secondary driving tasks from disrupting the driving, alleviating cognitive pressure. Additionally, Participants noted that utilizing voice interaction can streamline the learning process, enabling users to become familiar with its functionality more quickly and efficiently. As one user remarked, *"When borrowing* *your friend's car, it's impossible to know all the buttons and layout, but if it has voice control, you can jump right in and start using it by just talking through voice user interface."*

However, negative experiences during use could significantly diminish the perceived comfort of voice user interface and negatively impact users' perceptions on the system's reliability. Current technological limitations may erode users' confidence in using and trusting these technologies. Some participants expressed,

> *"If the accuracy of task execution decreases, I feel like this technology isn't very good... it gives me the impression that it's clumsy and childish."*

Participants frequently reported challenges with voice interaction systems including failure to trigger dialogues, low recognition efficiency and Unrecognized, slow recognition speed, misunderstanding, and ambiguity during everyday use. For example, one user mentioned, *"I often have to call out the wake word multiple times before it responds, or it takes several attempts to understand what I need."* Similarly, when performing specific tasks, users noted issues with command comprehension, as one participant explained, *"It doesn't fully grasp the content of my commands. For example, when I try to use the 'play music' command, it doesn't understand my actual intent."* The system also struggles to recognize synonymous or similar commands, such as identifying "dark mode" but failing to interpret "night mode" when a user says, "Quickly turn the screen to dark for nighttime."

External environmental noise further impacts the performance of voice user interfaces (VUIs), leading to inconsistent results. As one user pointed out,*"When environment inside the car is noisy, or there are many people talking, it can't recognize my commands."* Additionally, participants highlighted the limited structure of voice commands, which restricts users to completing one task at a time. A participant observed, *"... You have to say one function at a time... there's no way to cover all tasks in one sentence; it can't handle complex recognition."* Moreover, The lack of contextual awareness also detracts from the system's usability, making interactions feel mechanical and overly scripted. This limitation prevents continuous conversation or multi-step operations, resulting in what participants described as a "non-human-like" experience. For instance, a participant reported, *"After giving several commands, it just freezes and doesn't respond at all, which feels unnatural.* Such frustrations force users to repeat commands frequently, leaving them "very frustrated and reluctant to use it."

These negative experiences undermine users' confidence in the reliability of voice interaction systems. Some participants described the interface as

> *"A fun gadget, but not a practical interactive way."*

In interviews, five participants noted that frequent malfunctions would likely lead them to abandon using voice commands for controlling functions altogether.

3.3.3.2 Factors influencing choices

Many factors have impact on the people choice on utilizing different controls way. For instance, the user behavior preference influence their decision-making when choosing the control medium.

Reasons for choosing physical controls: Compared to the Voice user interface, The 'tangible and visible' nature of physical controls is a key reason users often prefer them over voice control. Unlike the intangible nature of voice interfaces, physical buttons provide a sense of reliability and enhance user trust. For instance, one respondent mentioned,

> *"Buttons and layout are quite consistent and reliable compared to software updates... I still prefer pressing the function button myself because I can see it, and it's faster."*

Participants highlighted that long-established usage habits are difficult to break, and adopting new interaction methods or designs can be challenging. One interviewee explained: *"I'm used to driving old cars that have many physical buttons, so I'm accustomed to using them... I find it hard to switch to voice control directly."* These habits, formed through years of repetition, create resistance to change. When updates or modifications deviate from established patterns, users may feel discomfort or confusion, with some expressing that it is *"impossible to reverse long-standing habits overnight."* Additionally, Users with conservative attitudes toward technology and prioritizing execution efficiency tend to prefer familiar and stable control methods. As one participant stated, they would rather *"maintain the status quo."*

However, current design trends toward integrating in-car control systems into single control units, such as touchscreens, are disrupting long-standing habits. One participant commented: *"My car is already quite integrated... so there are few things I can physically control. I either use the screen or voice."* This forced adaptation is accelerating behavioral changes, compelling users to adjust to newer systems despite their initial preferences.

Choosing voice control: preferences for simplified and streamlined operations significantly influenced the choice of voice interaction. In specific contexts, such as tasks requiring longer response times, many participants preferred voice commands for their ease and speed. Two key factors significantly influence users' decisions when selecting voice user interface: Knowledge of features and their control methods and Confidence in their operational abilities.

Differences in users' familiarity with features and their control methods play a critical role in shaping preferences for voice interaction. Some users struggle with the position and procedure of buttons and physical controls, making such operations challenging. As one participant noted, *"Sometimes, I don't know which button controls what function, and I don't want to divert my attention while driving to find the right button."* For these users, voice control offers a more natural and intuitive alternative, as it aligns with cognitive patterns and removes the need to navigate complex button layouts or operational logic. By allowing users to express their intent using simple natural language to eliminate the burden of learning and memorizing intricate control processes. This perception drives a preference for VUI as a means to simplify and streamline vehicle operations.

Second, Lack of confidence in one's control ability significantly influences the preference for voice control. Many users expressed concerns about being distracted by complicated interfaces, potentially posing safety risks while driving. One user noted,

> *"As a driver who is not very skilled, I get nervous and need to focus intensely on*

> *the road and the car... I don't have the mental bandwidth to talk to people around me or perform other tasks."*

For such users, voice control reduces the dependency on physical interfaces, helping to avoid distractions and the insecurity associated with frequent manual operations. By handling routine tasks through voice commands, these systems enable drivers to focus on core activities like driving, ultimately boosting their confidence, sense of control, and overall driving experience.

Beyond explicit factors influencing users' preference for voice control, several implicit factors also contribute to their inclination toward using voice functionality. These include: 1) Infrequent usage and Low automation of features, 2) Reducing physical controls enhances the aesthetic appeal of interior design, and 3) The ergonomic optimization of the interior layout.

> *Participant 1: "...Adjusting the mirrors and steering wheel? I don't think those buttons need to stay... you set them once after buying the car, and then hardly touch them again. I haven't used them much over the years."*
>
> *Participant 2: "...You might press some of those buttons once or twice a year, but they take up space where you could add something more useful, like an extra air vent. It would make the car feel less cluttered and more open."*

3.3.3.3 Difference in use scenarios and stakeholders when utilizing VUI

This part explores user preferences for voice control across different scenarios and the usage patterns of different user types (such as drivers, passengers, and children).

Differences in use scenarios: During the interviews, participants described various scenarios in which they might use voice user interfaces or voice interactions, including situations involving other stakeholders, such as passengers. These scenarios included entertainment purposes, driving-related tasks (at both high and low speeds), pre-trip preparation, and use during traffic congestion. The performance and utility of voice control varied across these scenarios. For entertainment purposes, participants often viewed the voice user interface as a fun feature to demonstrate to others. Before starting a journey, users frequently employed voice control to streamline multitasking and improve efficiency. While driving, participants preferred using voice control for tasks unrelated to driving, as it helps minimize manual operations and maintain focus. In particular, during slow-moving traffic or when stopped, such as in congestion, drivers used voice interaction to manage features more thoroughly, helping to alleviate the stress and pressure associated with driving.

> Entertainment purpose: *"Whenever I have friends who have never been in my car, I like to show them how I use voice control for things like 'air conditioning' or other functions... it's kind of fun..."*
>
> Pre-Trip preparation: *"...Before heading out, I could use the voice assistant to set navigation, adjust the AC, and take care of other stuff while getting ready... it saves so much time...."*

> Low-speed scenario:*"...Traffic jams can be so draining with all the stop-and-go... but using the voice makes it easier to handle simple tasks and helps me relax a bit..."*

Different stakeholders: Different stakeholders, such as passengers, influence habits and preferences related to control methods. Passenger behavior varies depending on their seating position. Front-seat passengers tend to rely more on physical controls, while back-seat passengers, who often face physical barriers to accessing controls (e.g., buttons), are more inclined to use voice control. As one respondent noted, *"When I'm in the back seat and want to adjust the air conditioning or control some other functions, I might have to ask the person in front to help me, which is a hassle—voice control would be much better then."*

Participants also highlighted children as passengers. They mentioned that children often display greater curiosity and interest in voice user interfaces compared to adults, finding them engaging and entertaining. However, participants also pointed out the potential risks, such as accidental activation of critical control functions by children, which could pose safety concerns.

3.3.3.4 Changes in social dynamics and connection

Social changes in voice user interface (VUI) usage manifest in two primary dimensions: age-related differences and the evolving emotional connection between humans and machines facilitated by the modality and mediators of control methods (VUI or physical controls).

Age-Related differences: Age plays a significant role in influencing VUI usage tendencies. Older drivers (aged 45 and above) with over a decade of driving experience tend to approach VUI conservatively, prioritizing reliability, accuracy, and simplicity. Conversely, younger users display a greater willingness to experiment with new technologies, embracing the innovative features of VUI with enthusiasm. As one 25-year-old respondent stated,

> *"I wish more features could be controlled by VUI... if you're going to do it, do it thoroughly."*

Emotional connection and carrier: In the interviews, some participants reckoned that VUI could be perceived not only as a functional tool but also as a conversational partner, fostering emotional connections. One participant remarked,

> *"...If it could be a bit smarter, like keeping me company while I drive alone, that would be great."*

This shift positions VUI as a source of emotional interactive objects, enabling users to form deeper bonds with the system. Ultimately, users may develop emotional dependencies on VUI, reacting emotionally to its responses and performance while attributing human-like characteristics to it. For example, participants pointed out the potential for personalization, saying, *"...If It could learn your habits through customization and understand your emotions from the tone of your voice, I'd definitely want to use it."*and

> *"...I think personalized experiences are about it understanding you—your mood, your habits—that would be great. It shouldn't feel like a cold, mechanical tool that just answers questions and solves problems; it could have some emotions to make it better."*

This emotional need becomes more evident when users drive alone, with two respondents noting, *"...Especially when driving alone, using VUI could make me feel like there's someone helping me operate the car."*

Furthermore, Some participants expressed a preference for a tangible product to complement the voice user interface. For instance, one participant noted that they would "happily accept" a physical embodiment, like "Nomi" (an in-car voice assistant product designed by NIO), compared to a purely virtual interface. These findings indicate that users' emotional dependence and expectations of VUI are not limited to functional needs but also encompass demands for social interaction and emotional support.

3.3.3.5 Perception of safety

Safety, a critical factor in driving and user interactions, significantly shapes decisions regarding the adoption of voice user interfaces (VUI). Participants consistently raised safety concerns, either explicitly or implicitly, during discussions about using VUI while driving. This section examines user perceptions of safety issues and concerns, focusing on two main areas: data privacy and the safety of voice control, as well as the prerequisites for safe use of VUI.

Concerns about data privacy: Privacy-sensitive participants expressed unease about using voice interaction interfaces, which they perceive as invasive and potentially compromising their private space, particularly within the confined environment of a car. There is widespread concern that companies or malicious actors could gain unauthorized access to their data. As one interviewee remarked, *"I don't want 'they' to be able to hear conversations inside the car."*

Utility-focused users, on the other hand, were more accepting of data collection if it served to improve services, provided the data remained stored locally for system training or upgrades. They emphasized the need for *"a return that matches their input."* However, nearly all participants (15 in total) stressed that while they might *"compromise for the sake of use,"* they firmly believe that data generated while driving—such as the frequency and type of voice-controlled functions—should not leave their control. As one interviewee noted, *"I don't want what I say or the functions I control while driving to be recorded."*

Users expressed particular discomfort with their voice data being analyzed for targeted recommendations, especially during moments that demand high concentration, such as driving. One participant explained their refusal to use VUI due to fears that usage habits could be tracked, potentially leading to accidental activations or disoperation, which they described as especially concerning when concentrating on driving.

Concerns about the safety of voice control features While the potential convenience of voice commands was acknowledged, participants raised concerns about the inherent safety of VUI, including its reliability and control allocation.

Reliability of voice technology: Many users doubted the reliability of current voice technology, citing poor recognition capabilities. The need to repeat commands multiple times before achieving the desired outcome introduces instability and increases cognitive load for drivers. One interviewee remarked, *"I'm afraid of saying something randomly in the car and accidentally activating the voice assistant to adjust the side mirror—I'd be really scared."*

Unlike physical buttons, VUI lacks visual feedback, making it harder to quickly identify and correct errors when users "say the wrong thing" or give an inaccurate command. This challenge is amplified during complex driving situations, where repeating or correcting operations can compromise safety. Novice and elderly drivers, in particular, confront with greater difficulties in ensuring operational reliability. In case of emergencies, the limitations of voice control become more pronounced. Noise interference, variations in speech speed, or technical errors can prevent the system from executing critical commands. One participant shared their concern:

> *"In an emergency, I'm afraid it won't understand my needs in time, which could lead to bigger problems."*

Control allocation issues: Participants highlighted that ambiguities in control allocation between humans and machines could create an unstable driving experience. Excessive machine intervention or delays in voice response could confuse decision-making, impacting safety. As one participant noted,

> *"...If there's a delay in the voice interaction, I need to decide whether to take over... this could affect my operation sequence... and my decision-making."*

Additionally, conflicts arise when VUI processes commands from multiple passengers. For instance, one interviewee expressed concern about unintended activations, saying, *"If a child in the backseat accidentally says something related to specific keywords, like those associated with the door handle, the voice assistant might mistakenly execute the command to open the car door, which could be dangerous."* This type of misoperation can distract the driver and compromise safety.

Safety prerequisites before using voice control: Participants emphasized the importance of establishing safety prerequisites to mitigate risks associated with VUI. Key suggestions included:

Disabling certain features: Users recommended disabling remote control functions or critical commands to prevent unauthorized operations or accidental triggers.

Training and guidance: Providing clear instructions, through manuals or interactive exercises, was seen as essential for helping users understand how to interact effectively with VUI. One participant suggested, *"Through manuals, prompts, or some interactive exercises, users could better understand how the voice assistant works and what commands are available... becoming more familiar with the operation."*

3.4 Summary

The combined results of quantitative and qualitative research provide a comprehensive understanding of the user experience associated with using Voice User Interfaces (VUI) for controlling in-car functions in current driving scenarios.

By employing a multiple linear regression approach to investigate how various factors predict final user choices, we found that among the predefined independent variables, those reflecting objective attributes—such as users' familiarity with feature, and actual usage frequency—had a more pronounced influence on user choices, particularly in both primary driving task and secondary task scenarios. This finding further indicates that users' current assessments of voice user interfaces (VUIs) for in-car feature control remain focused on whether the required capabilities are fully realized and how effectively they perform, thereby hindering broader acceptance of this interaction method. Moreover, because there is a significant gap between advanced voice interaction technologies and mass-produced VUIs regarding in-car control capabilities—and users' experiential assessments are often based on these less advanced products—users' evaluations may include more negative sentiments. Consequently, user experience carries less weight in decision-making (failing to reach a statistically significant difference), making it difficult for user experience to serve as the primary evaluative criterion. Although user preference is generally considered a reflection of usage habits and needs, it also implies users' expectations for future VUI control capabilities, and reveals users' attitudes and their potential for future adoption, suggesting that user preference may serve as a more critical reference factor in designing and predicting user choices.

In the qualitative study primarily employing semi-structured interviews, direct feedback from participants provided deeper insights into the logic and rationale behind users' choices of different control methods. In particular, the interviews shed light on how users' willingness, decision-making, and past usage experiences influence their choice of using a VUI as the medium for controlling in-car functions. These qualitative findings helped refine the set of influential factors identified in the quantitative study. For instance, negative user experiences with current VUI technology—such as low recognition accuracy, lack of contextual awareness, and delayed responses—pose objective technical issues that significantly undermine user confidence. Furthermore, previously overlooked factors, including privacy and security concerns (especially regarding data misuse and operational reliability), became more apparent, further affecting user decisions. Addressing these issues through improved system design and robust privacy safeguards is critical for building trust and promoting acceptance. Additionally, statements made during the interviews offered insights into the capabilities, features, and user experience requirements regarding VUI-based control. For example, personalization and emotional intelligence within VUI systems were perceived as drivers that could increase user demand for such products. Participants expressed a desire for systems that not only execute tasks effectively but also adapt to their preferences and provide more humanized interactions. This could include recognizing individual habits, offering contextualized feedback, and even serving as an emotional companion during solo drives.

This mixed-methodology approach offers a comprehensive understanding of user needs and concerns, forming a strong foundation for further exploration. In the next section, we will develop a factor model framework to analyze the key factors influencing VUI adoption for controlling in-car functions. Subsequently, in Section 6, we will apply these findings in two ways: 1) By analyzing interaction needs, including VUI-specific requirements and those related to its design and functionality. 2) By proposing design guidelines and evaluation standards based on the study’s insights, optimizing the user experience evaluation framework. These efforts aim to enhance sensitivity to user experience and predictive accuracy, facilitating the broader adoption of VUI technology for in-car function control.

4 Model and Influence Factors

In this section, we introduced a descriptive model aimed at understanding the cognitive and behavioral patterns of users when utilizing voice interaction to manipulate car features, without engaging with other tangible interaction (e.g., physical buttons/stick). The model is structured based on the Impact-Centered Design framework and Product Experience Theory, which delineate and synthesize the product attributions, psychological and behavioral impacts triggered by human-product interaction.

In model development, commonalities were identified, conceptualized, and summarized from the results of both qualitative and quantitative analysis conducted in the prior user study. Following the procedure of establishing a descriptive model in cases where the phenomenon under investigation is not yet fully understood, two primary research questions guided this process:

- *What types of factors constitute the dimensions affecting users' interactions?*
- *How can these factors be grouped and interact?*

Inspired by Impact-Design framework, we aimed to differentiate the objective properties of the product from the subjective psychological impacts and user perceptions of the interaction experience. However, unlike Fokkinga's model—which takes a general perspective on human impact—the proposed model seeks to incorporate aspects that are tailored to the considerations when controlling vehicle features. These metrics reflect specific cognitive and behavioral considerations relevant to the driving context, such as trust, driving confidence, and safety.

The subsequent part presents an overarching hierarchy of the model, outlining the interrelations between its various components. Additionally, each element would be explained in detail.

4.1 Overall model

The architecture of the model, as illustrated in Figure 7, focuses on the interplay between multiple interrelated variables and themes that collectively influence users' decisions to adopt a Voice User Interface (VUI) for controlling product features. The model integrates individual user characteristics (e.g., knowledge, behaviors, habits, and preferences) with user attitudes (trust and confidence) and safety perceptions as the primary components of "Human Impact." Simultaneously, the model considers the intrinsic characteristics of the product (such as its functionality, features, security, and reliability) and the user experience (including product aesthetics, meaning, and emotional experience) as factors at the product level.

The model defines the product level through a two-tiered structure: VUI product attributes, which include functional capabilities as well as aspects related to security and reliability. These attributes determine how users interact with the product through voice (VUI Interaction & Usage). Variations in product attributes and modes of interaction directly affect the Product Experience, which encompasses elements like product aesthetics, meaning, and emotional experience.

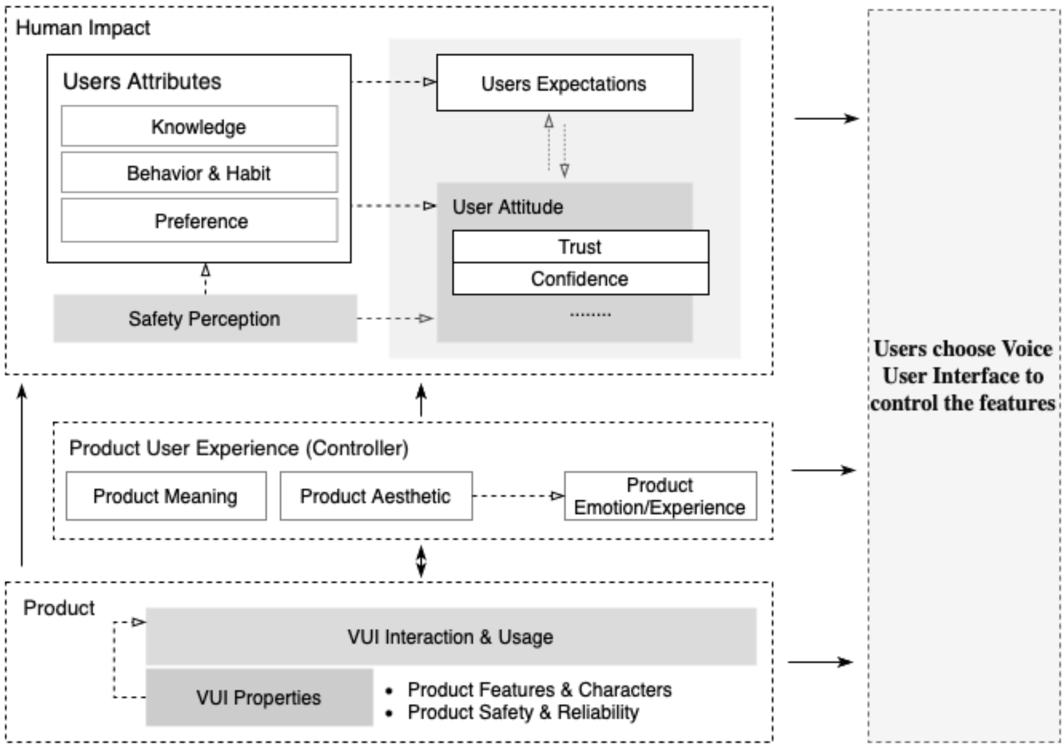


Figure 7: The overall hierarchy of the influencing model for user adopting VUI for feature control

In the "Human Impact" module, individual user characteristics (such as knowledge, behaviors, habits, and preferences) interact with user expectations and attitudes. The user's knowledge and habitual behavior shape their initial attitudes and expectations regarding product use. When the product experience aligns with or exceeds these expectations, the user's trust and confidence in the product are likely to increase; conversely, negative experiences may diminish their willingness to use and prefer the product. Safety Perception acts as a moderating variable, further influencing the user's assessment of VUI security and their subsequent decisions regarding its use.

Arrows within the model indicate how the critical component of product experience influences user behaviors and selection, and this process illustrates a bottom-up pathway: from product attributes and interactions to user experience, and finally to the user's overall attitudes and behavioral choices. Ultimately, these pathways converge to influence the user's behavioral choice—whether to adopt VUI for controlling product features.

Component	Sub-component	Definition / Examples
User Attributes	Knowledge	User's understanding of product features, functions, and usage, impacting decision-making.
	Behavior & Habit	Patterns and habits of Car features and VUI use in daily life, including usage frequency and context.
	Preference	User's inclinations toward specific product features or interaction modes, influenced by personality, context, and past experiences (Sjöberg, 2000).
Users Expectations	Anticipations regarding product performance, usability, reliability. Expectations shape initial attitudes and intentions, with increased satisfaction if actual experience aligns (Oliver, 1980).	
User Attitude	Trust	Overall perception of product integrity, including trust in safety, reliability.
	Confidence	Self-assessment of one's ability to use the product effectively(Bandura, 1977); Related to greater exploration and frequent use .
Safety Perception	User's subjective sense of security when using a product, including physical and data safety.	
Product Experience	Product Aesthetic	Sensory enjoyment from visual, auditory, tactile aspects, which influences emotional response and product attitudes (Hekkert, 2006).
	Product Meaning	Functional and emotional significance assigned to a product by users. Products can symbolize identity or serve as social tools (Desmet & Hekkert, 2010).
	Product Emotion	User's emotional responses to product use, such as pleasure or frustration. Emotions are subjective and influenced by individual needs, goals, values (Smith & Lazarus, 1990).

Component	Sub-component	Definition / Examples
Product Features		Product features refer to the inherent attributes and characteristics of a product in terms of functionality, technology, and design. In this context, they include the characteristics of VUI products (interaction methods, product security, and reliability) as well as the attributes of the actual products used by users in real-life situations. Product features directly impact user experience, emotional response, and usage behavior. Specific product features can satisfy both the functional and emotional needs of users, thereby influencing their attitudes and purchasing decisions (Keller, 1993).
VUI Interaction		Interaction and usage encompass all non-emotional or pre-emotional activities and events that occur between users and the product, including behaviors during product perception and operation, such as checking, initiating, and terminating use, interface interactions, and product updates and maintenance (Schifferstein & Hekkert, 2008). In this study, interaction and usage specifically refer to all interactions and operations through which users trigger or use functionalities via different control methods (e.g., VUI), such as initiating and terminating voice input, recognizing and providing feedback on commands, and interacting with physical control buttons.

4.2 Model components

4.2.1 User attributes

User attributes encompass the personal characteristics and behavioral patterns of individuals in their interactions with products. These attributes serve as moderating factors that influence users' attitudes, expectations, and, ultimately, their decision-making processes. User attributes can be categorized into the following key aspects:

Knowledge: This encompasses the depth of a user's understanding of VUI and physical controls, including their functionalities, features, usage methods, and limitations. According to consumer behavior theory, knowledge level is a critical determinant of users' information processing abilities and decision-making behaviors (Alba & Hutchinson, 1987). In the context of technological product usage, a user's knowledge level directly impacts their ability to comprehend product functionalities and proficiency in utilizing them (Park & Lessig, 1981). Generally, higher levels of knowledge are associated with increased acceptance of technology and a greater willingness to use it, as knowledgeable users are better equipped to understand and leverage product features to fulfill their needs.

Behavior and Habit: This refers to users' behavioral patterns and habits in utilizing in-car functions in their daily lives, such as using voice commands while driving or in non-driving contexts. It encompasses aspects like usage frequency, contexts of use, and interaction modes. According to habit formation theory (Wood & Neal, 2007), repetitive behaviors become automated habits through conditioning and reinforcement mechanisms. Established habits significantly influence future behavioral decisions, as users who frequently engage with a particular product are less likely to switch to alternative products or control methods.

Preference: This refers to a user's personal inclinations toward specific car features or interaction methods (e.g., physical controls or voice controls). Preferences are shaped by various factors, including personality traits (e.g., openness, conservatism), usage contexts, and past experiences (Sjöberg, 2000). Preferences play a pivotal role in the decision-making process, as users tend to select interaction methods that align with their preferences, thereby enhancing their subjective satisfaction and perceived utility.

4.2.2 Users' expectations

Users' expectations refer to their predictions about a product's performance, reliability, ease of use, and other attributes. In this study, these expectations specifically relate to using a Voice User Interface (VUI) instead of physical buttons for controlling functions. When experiences meet or exceed expectations, user satisfaction and willingness to continue using the product increase. Conversely, unmet expectations can lead to negative attitudes and reduced usage intentions (Oliver, 1980). This thesis specifically focuses on users' expectations regarding controlling through VUI and their future usage tendencies.

4.2.3 User attitude

User attitude refers to the overall perception or emotional inclination of users toward a product or service, encompassing cognitive, affective, and behavioral components (Ajzen, 1991). In this model, user attitude primarily includes the following aspects:

Trust and confidence: When users self-assess their ability and skill in controlling features through a VUI, trust and confidence reflect their belief in both the product's functionality and their own capability to use it effectively. Trust encompasses perceptions of the product's security, reliability, and integrity, while confidence reflects a user's evaluation of their capacity to perform specific tasks (Bandura, 1977). Higher levels of trust and confidence in a particular control method increase the likelihood of users adopting it continuously and recommending it to others.

4.2.4 Safety perception

Safety perception refers to a user's subjective sense of security when using a VUI, encompassing both physical safety (e.g., risks associated with VUI operation) and data security (e.g., protection of personal information). In user experience research, safety perception is recognized as a critical factor shaping user trust and willingness to use (Cox & Rich, 1964). When users perceive high levels of risk, they are more likely to reduce or abandon product usage (Bauer, 1960). Safety perception is influenced by a combination of the product's objective safety attributes and the user's knowledge, past experiences, and psychological state.

4.2.5 Product experience

Product experience encompasses the overall sensory and emotional impressions a user derives from interacting with a product. This includes aspects such as the product's appearance, aesthetics, tactile qualities, and the emotional responses elicited during use. Drawing on Desmet and Hekkert's (2007) experience framework, we define product experience as comprising the following dimensions:

> **Product aesthetic**: This part refers to the sensory enjoyment users derive from a product's visual, auditory, and tactile qualities. Aesthetic experience is characterized as disinterested, focusing on perceptions of order, harmony, symmetry, and related attributes (Hekkert, 2006).
>
> **Product meaning**: This dimension refers to the functional and emotional significance users attribute to a product. For instance, a product may be seen as a symbol of personal identity, an expression of lifestyle, or a medium for facilitating social interactions (Desmet & Hekkert, 2010).
>
> **Product emotion/experience**: Product emotion refers to the emotional responses users experience when interacting with a product or VUI, such as pleasure, satisfaction, or frustration. These emotions are inherently subjective, influenced not only by the product's attributes but also by the user's individual needs, goals, values, and capabilities (Smith

& Lazarus, 1990). In this study, emotion “it is pleasant or enjoyable?” serves as direct indicators of users’ evaluations of their experience with the VUI.

4.2.6 Product features and properties

Product features comprise the inherent attributes and characteristics of a product. In the context of this study, they include the characteristics of VUI products (e.g., interaction methods, security, and reliability) as well as the attributes of the features and control methods used by users in real-world scenarios. By meeting both functional and emotional needs, specific features can significantly influence user attitudes and decisions (Keller, 1993).

4.2.7 VUI interaction and usage

Interaction refer to all activities and communication that occur between users and product. These include behaviors related to product perception and operation, such as checking, initiating, and terminating use, as well as interface interactions, updates, and maintenance (Schifferstein & Hekkert, 2008). In this study, interaction specifically pertain to the actions and operations users perform to trigger or utilize functionalities via various control methods (e.g., VUI).

4.3 Summary

This section explains how users decide to adopt voice user interfaces (VUI) for controlling functions instead of using physical controls. It highlights the role of user characteristics—such as knowledge, habits, and preferences—and how they interact with trust, confidence, and safety perceptions to influence decisions. It also considers how product attributes like functionality, reliability, and overall user experience impact adoption.

The model combines these factors to provide a clear framework for understanding user behavior. It offers practical guidance for designers to improve VUI systems and reduce the need for physical controls while maintaining a positive user experience.

5 Comparative Study and Experiment

While Chapter 4 explores how users understand and ultimately choose to employ VUIs for function control, an important question remains: which types of functions are users currently willing to control through voice, without the need for a corresponding physical interface? To address this, we designed an experiment to gain deeper insights into user preferences and determine which categories of physical controls could be effectively replaced by voice interaction.

This experiment employed a within-subjects design, where the independent variable was the type of control method used by participants (physical buttons versus voice control) to perform identical operational tasks. The experimental environment featured a detailed simulated driving scenario using a driving simulator. Participants were invited individually to the experiment for evaluation and were asked not to discuss the content of the experiment with other participants afterwards. The data were collected and analyzed from the answers of the questionnaires and interviews.

5.1 Experiment equipment and VUI program setup

The experimental tasks were predefined. The experimental setup included a driving area which was equipped with a driving simulator, a display screen, physical button simulation devices, and a computer system operating a voice interaction interface, as shown in Figure 8a. The display screen, positioned in front of the simulator, played driving videos sourced from the internet to create a realistic driving environment. The physical control device was designed to replicate the functionality of real vehicle physical control.

(a) The setup for experimental test

Primary Driving Task:

"Open acc"
"set adaptive cruise control to 65 mph."
"activate cruise control."
"maintain speed with cruise."
"open parking assistant"
"help me park the car"
"activate parking assist."
"help me find a parking spot."
"start the parking assistance."
"switch to eco mode."
"switch to sport mode"
"put the car in comfort mode."
"enable sport driving mode."
"activate eco mode"
"activate brake."
"activate parking brake"
"release the parking brake."
"set the brake."
"brake"

Secondary/Other driving Task:

"Adjust right mirror in."
"move side mirrors to default position."
"tilt the left mirror down."
"adjust right mirror"
"adjust left mirror"
"start windshield cleaning."
"activate wipers."
"open wipers"
"open cleaning"
"adjust wiper speed."
"adjust wiper speed quicker"
"turn off the headlights."
"turn on the headlights"
"switch headlights to high beam."
"activate fog lights."
"open the light"
"turn on the light"
"turn off the light"
"unlock all doors."
"lock the exterior doors."
"open my door."
"disable door locks."
"unlock passenger door."
"open the door"
"close the door"
"lower the steering wheel."
"raise steering column."
"Adjust steering to my profile."
"turn up the volume."
"play the next song."
"switch to radio."
"lock the trunk."
"open the trunk"
"open the front box"
"open the back"
"release trunk latch."
"close the trunk."
"increase ac temperature."
"turn on the heater."
"turn off ac"
"open ac"
"adjust airflow direction."
"close the sunroof."
"tilt the sunroof."
"fully open the sunroof."
"open it"
"start interior air circulation."
"start circulation"
"stop air recirculation."
"close the circulation"
"open glove box."
"close the box."

(b) Recommended voice commands for controlling features

Instructions:
Please respond to the following questions based on your interaction experience with the voice assistant interface. Use the following 5-point rating scale for your answers:

- 1 = Strongly Disagree
- 2 = Disagree
- 3 = Neutral
- 4 = Agree
- 5 = Strongly Agree

Dialogue Evaluation

Question	Rating Scale (1-5)
How do you rate the quality of this feedback?	

Integrated Usability and User Experience Evaluation Survey (Part 1)

Questions	Rating Scale (1-5)
I think that I would like to use this interface frequently.	
I found the interface unnecessarily complex.	
I thought the interface was easy to use.	
I think that I would need the support of a technical person to be able to use this interface.	
I found the various functions in the interface were well integrated.	
I thought there was too much inconsistency in this interface.	
I imagine that most people would learn to use this interface very quickly.	
I found the interface very cumbersome/awkward to use.	
I felt very confident using the interface.	
I needed to learn a lot of things before I could get going with this interface.	
I felt afraid when it was unable to work properly.	
I think I would concern and evaluate its safety when using.	

Integrated Usability and User Experience Evaluation Survey (Part 2)

Questions	Rating Scale (1-5)
Does this VUI help you in attaining your goals and motives?	
Is this whole process(event) pleasant or enjoyable?	
Does this interface confirm your expectations?	
Does this event match with your standards?	
Can I handle or change the event?	
How certain are you about the event?	
After using it, Do you prefer to rely on voice for controlling functions rather than physical buttons?	

1

(c) Evaluation scale for after-experiment evaluation

Figure 8: Experiment equipment and VUI program Setup

The voice user interface (VUI) followed a structured control logic. Users issued commands in natural language (Designed commands shown in Figure 8b), after which the system acknowledged the request with a "processing" response one second later. After three seconds, as a simulation of the operation time, the system notified the user with "function activated." Subsequently, the system paused for one second to ask if further operations were required. If no response was received within ten seconds, the interaction process automatically terminated.

The response time of the voice interaction interface is designed based on the current research and evaluation on voice interaction response speed and functional execution capabilities. The average response time is approximately 1 second, with an overall task completion time of about 3 seconds (Isyanto et al., 2020).

5.2 Participants

A total of 12 participants took part in the experiment, including six who had previously participated in interviews. The order of participant selection was determined through a random draw.

5.3 Experiment procedure

Before the start of the experiment, all 12 participants received training on the experiment equipment, including physical controls and the voice user interface (voice control), as well as the experimental procedure. Participants were instructed by researchers to ensure they fully understood the experimental procedures and techniques, including the ”Wizard of Oz” method. They were also briefed on the purpose of the study and the privacy protection measures in place before signing an informed consent form.

During the experiment, each participant individually entered the testing environment and performed a series of standardized operational tasks, such as adjusting air conditioning temperature, selecting a music, and setting navigation. Each participant interacted with four primary driving-related features, five secondary driving features, and five features related to tertiary driving features. Due to experimental constraints, door handle controls were excluded. In the first round, participants used physical controls (for example, buttons) to complete the feature task. Adequate breaks were provided between control method switches to minimize fatigue and learning effects. In the second round, participants used the voice user interface (voice control). This design enabled a direct comparison of the efficiency and accuracy of the two control methods. Task completion times and accuracy were meticulously recorded for all participants.

Following the experiment, participants completed a questionnaire evaluating their interaction experiences. The entire experimental procedure lasted approximately 30 minutes per participant. Participants were appropriately compensated for their time and effort.

5.4 Measurement

Participants responded to a series of questions using a 5-point Likert scale, ranging from “strongly disagree” (1) to “strongly agree” (5) and the scale shown in Figure 8c. System usability was evaluated using the System Usability Scale (SUS), while user experience was measured utilizing the user experience evaluation framework employed in the previous survey. In response to user research findings highlighting concerns about the safety of voice interactions, a voice interaction safety assessment was incorporated into the scale. The final version of the scale is included in the appendix of the research report. Detailed data analysis results will be presented in the results section.

In this study, the 12 participants had a mean age of 31.4 years ($SD = 10.7$), spanning a wide age range from 18 to 53 years. Participants exhibited significant variation in driving experience, with an average driving experience of 7.3 years ($SD = 7.8$). The majority of participants (8 individuals) reported prior experience with in-vehicle navigation systems or voice assistants, and 5 indicated frequent use of such technologies. Control functions for both primary and secondary driving tasks were grouped into five categories, and all participants tested each function comprehensively. For other driving-related tasks, the most frequently selected features included media control (12 participants), climate control system (12 participants), recirculation system (12 participants), trunk switch (10 participants), and sunroof controls (9 participants).

5.5 Results

5.5.1 Usability study results

The SUS scores for different task level features across different control methods are shown in Figure 9. In the study examining usability for controlling functions during primary driving activities, a significant difference was observed between physical control methods (e.g., buttons) and voice assistants. Results of the t-test revealed a statistically significant difference in usability scores between the physical control group and the voice assistant group ($t = 5.914$, $p < 0.005$). This finding indicates that for primary driving tasks, users tend to prefer physical controls for setting functions, while expressing concerns about the usability of voice controls.

In contrast, when evaluating usability during non-driving tasks (secondary driving tasks [SDT] and tertiary driving tasks [TDT]), no statistically significant differences were found between traditional physical control methods (Group 1) and voice user interfaces (Group 2). The paired-sample t-test showed no significant difference in usability scores for either task category (SDT: $t_1 = 0.34$, $p_1 = 0.743$; TDT: $t_2 = -0.302$, $p_2 = 0.768$). For SDT features, the physical control group had a slightly higher mean usability score ($\bar{x}_1 = 67.92$, SD $\sigma_1 = 9.46$) compared to the voice user interface group ($\bar{x}_2 = 66.46$, SD $\sigma_2 = 8.42$), but the difference was not statistically significant ($p > 0.05$). This suggests that for controlling secondary tasks, both interaction methods performed similarly in terms of usability.

For TDT features, the voice user interface group scored slightly higher in usability ($\bar{x}_2 = 67.08$, SD $\sigma_2 = 5.31$) compared to the physical control group ($\bar{x}_1 = 66.25$, SD $\sigma_1 = 5.78$). Although the difference was minimal and not statistically significant, these results indicate a slight user preference for voice controls over physical ones for tertiary tasks. Overall, the findings suggest that while physical controls are favored for primary driving tasks, voice user interfaces perform comparably to physical controls for non-driving tasks.

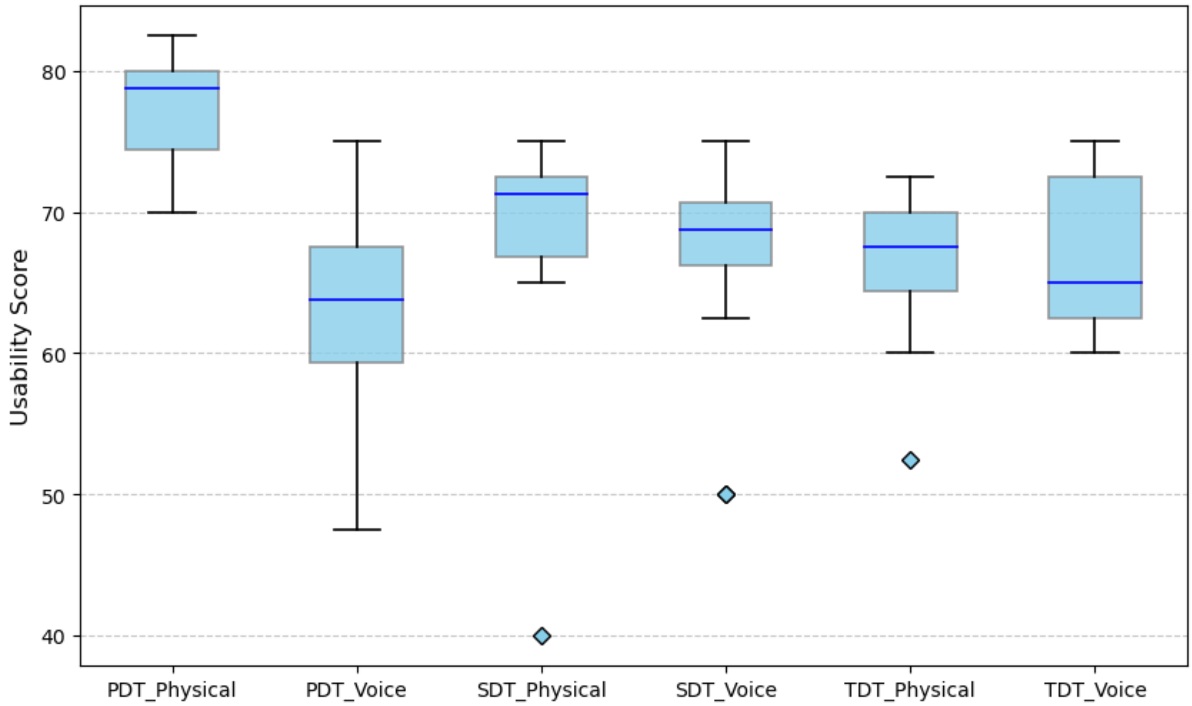
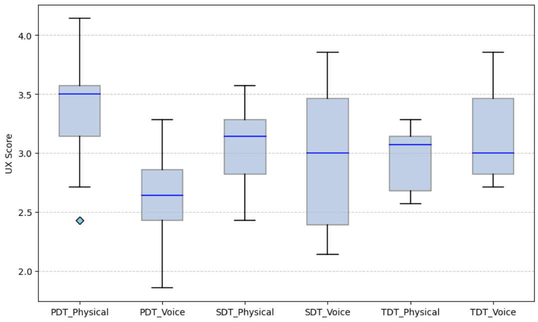


Figure 9: Usability Test Scores for PDT, SDT, and TDT Features Across Different Control Methods

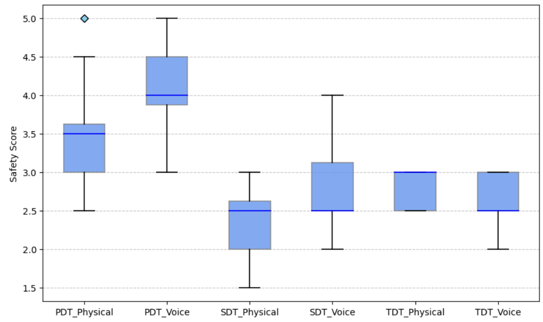
5.5.2 User experience results

The UX scores for different tasks level features across different control methods are shown in Figure 10a. For primary driving activities, a significant difference was observed in user experience (UX) scores between physical controls (e.g., buttons) and voice assistants. A paired-sample t-test revealed a statistically significant difference between the physical control group and the voice assistant group ($t = 4.29$, $p < 0.001$). Specifically, the physical control group (UX Group 1) had a mean UX score of $\bar{x}_1 = 3.37$ (standard deviation $\sigma_1 = 0.47$, variance $\sigma_1^2 = 0.22$), indicating higher user experience ratings with relatively consistent data distribution. In contrast, the voice assistant group had a mean UX score of $\bar{x}_2 = 2.62$ (standard deviation $\sigma_2 = 0.38$, variance $\sigma_2^2 = 0.15$), reflecting lower UX ratings. These results suggest that for primary driving tasks, users exhibit a clear preference for physical control methods, perceiving them as offering a better overall experience.

For secondary driving activities (SDT) and tertiary driving task-related feature controls (TDT), the paired-sample t-test found no statistically significant differences in UX scores between physical controls and voice interfaces (SDT: $t_1 = 0.39$, $p_1 = 0.704$; TDT: $t_2 = -1.152$, $p_2 = 0.266$). For secondary tasks, the physical control group had a mean UX score of $\bar{x}_1 = 3.05$ (standard deviation $\sigma_1 = 0.37$, variance $\sigma_1^2 = 0.13$), compared to $\bar{x}_3 = 2.96$ (standard deviation $\sigma_3 = 0.58$, variance $\sigma_3^2 = 0.33$) for the voice interface group. For tertiary tasks, the physical control group had a mean UX score of $\bar{x}_2 = 2.95$ (standard deviation $\sigma_2 = 0.28$, variance $\sigma_2^2 = 0.08$), while the voice interface group scored $\bar{x}_4 = 3.14$ (standard deviation $\sigma_4 = 0.42$, variance $\sigma_4^2 = 0.17$). These findings indicate that for secondary driving tasks, users perceive physical controls as performing slightly better than voice interfaces. However, for tertiary driving tasks, users are more willing to employ voice interfaces, suggesting greater acceptance and utility for such tasks. Overall, in terms of user experience, both interaction methods perform comparably, with no significant differences for secondary and tertiary tasks.



(a) User Experience Scores for different features and control ways



(b) Safety perception Scores for different features and control ways

Figure 10: User Experience and Safety perception Scores for PDT, SDT, and TDT Features Across Different Control Methods

5.5.3 Safety study results

During driving, different control methods elicited varying levels of safety concerns among users, particularly in the case of VUIs, which provoked greater apprehension. The safety perception for features of different task levels across different control methods is shown in Figure 10b. When analyzing safety scores for primary driving features, we found a statistically significant difference in the t-test($t = -2.244$, $p = 0.0351$), indicating that participants perceived VUIs as posing higher safety risks compared to physical controls in these situations.

In contrast, regarding SDT- and TDT-related features, the t-test results (SDT group: $t_1 = -1.836$, $p_1 > 0.05$; TDT group:$t_2 = 1.172$, $p_2 > 0.05$)did not reach statistical significance, suggesting that users perceive the safety risks of voice-controlled features to be on par with those of physical controls. Overall, these findings indicate that users are highly concerned about safety—particularly with voice interaction technologies—remaining wary of potential uncertainties and risks associated with VUIs.

5.6 Summary

This study investigated the usability, user experience, and safety of various control methods for managing features across different driving task categories, including primary, secondary, and tertiary tasks. Under the experimental conditions, the results indicated that, apart from functions closely tied to driving or highly frequently used secondary tasks (such as door handles), users perceived the usability and experience of voice control alone to be comparable to or better than physical controls. This suggests that, provided voice controls function seamlessly, users may be open to managing all other functions exclusively through voice interaction, eliminating the need for physical controls.

For primary driving tasks, physical controls (e.g., buttons) for selected features significantly outperform voice interfaces in terms of both usability and user experience. The high-risk nature of primary driving tasks and their related features amplifies users' apprehension about VUI (voice interaction), as reflected in higher perceived safety risks for VUIs. Consequently, it is crucial to incorporate well-designed, ergonomically placed physical controls for these core features instead of replacing them. Regarding secondary and tertiary driving tasks examined in this research, no statistically significant differences were identified between physical controls and voice controls in terms of usability or user experience. While users remain somewhat cautious about voice interactions, the perceived safety risks associated with voice user interfaces did not differ significantly from those of physical controls, indicating comparable performance for non-critical tasks. Interestingly, a slight, albeit non-significant, preference for voice controls emerged for tertiary tasks, reflecting a growing acceptance and potential viability of VUIs in low-risk scenarios. This suggests that, provided voice controls align with user needs, features related to non-primary driving activities can be effectively managed through voice interaction alone, eliminating the necessity for corresponding physical controls such as buttons. Furthermore, in instances where physical controls require more complex interactions, users are not only more likely but also more inclined to prefer voice controls as their primary method.

When functions are no longer constrained by control mediators, with only primary driving task-related features retained, the removal of physical control buttons will drive changes in the vehicle's interior design and layout. This evolution will, in turn, reshape users' control behaviors and needs. The insights presented in this chapter will serve as a pivotal input for the next chapter, informing our discussions and guiding the design practice of VUIs and associated work procedures.

6 Design Guidelines and Practice

In this chapter, the design guidelines are introduced, and the design practice of VUI is described. Detail is given to the design preparation, modeling and rendering.

6.1 Design guidelines

Design guidelines refer to a set of standardized rules and recommendations that assist designers in creating human-computer interaction systems, ensuring interfaces are both user-friendly and efficient (Shneiderman, 2000). These guidelines are typically rooted in user research, theoretical frameworks, and practical experience, offering clear design directions that help products meet the needs of specific user groups while improving usability, acceptability, and user satisfaction (Norman, 1988). In this context, the guidelines were developed using user research insights and comparative studies, where the combination of direct user feedback and empirical evidence contributed to shaping practical design principles.

6.1.1 Clarification of the scope and capabilities of use

In the context of voice interaction control and VUI, "clarifying the scope of functionality" typically refers to the user understanding or clearly defining the boundaries or range of the current task or operation. This means that the user identifies which functions can be controlled via voice and to what extent these functions can be controlled. For example, clarify which functions can and cannot be controlled. This ensures consistency between the interaction system and the user in terms of functionality, preventing misunderstandings or operational errors. Users expect consistent and positive feedback during interactions, and if the system meets or exceeds expectations, user acceptance and satisfaction are likely to increase (Oliver, 1980).

- **Clarification of functional capabilities** Users, particularly drivers, need to understand the existing VUI capabilities, including the range of features that can be controlled by voice and the limitations of those functionalities. In user studies and evaluations, participants expressed the desire to understand what features are controllable and to what extent those features can be controlled.
- **Task confirmation prompts** Through VUI or Multi-modal control method (like tactical and voice), the system should indicate whether a specific feature is currently feasible or could be set according to the user's understanding.
- **Task confirmation example** User research indicates that when users cannot issue recognizable commands through voice interaction, they often feel frustrated, negatively impacting the overall user experience. Therefore, in a safe driving environment, the system can provide usage examples to guide users. This approach effectively reduces confusion during initial use and increases user satisfaction.

6.1.2 Confirmation strategies and command-control modes

6.1.2.1 Functional limitations based on driving status and user permissions

The authority level of the control should vary according to the driving conditions. For example, during high-speed driving, complex tasks that require great attention should be limited, while in safer environments, they should not. The authority level should also vary depending on the user. For example, drivers usually have the highest authority level, while passengers are restricted from any driving-related functions because granting control to all passengers in the vehicle can cause anxiety for drivers. *Design example*: When driving at high speeds, the system only allows simple control commands such as “adjust volume” or “turn on air conditioning,” while more complex inputs like detailed navigation settings are restricted.

6.1.2.2 Confirmation strategies

The system should suggest functions based on the user’s driving status and habits, or support personalized settings that allow users to customize frequently used commands and controls. It should also enable quick access to these settings through voice commands. For example, a user can define a driving mode that includes their preferred settings like air conditioning, music volume, and driving mode configurations, and switch to this mode with voice commands.

Design example: A user says “Activate my driving mode,” the air conditioning, music volume, and adaptive cruise control are adjusted according to pre-defined settings.

6.1.2.3 Reducing confirmation frequency based on familiarity

As the system becomes more familiar with the user’s preferences, it can gradually reduce the frequency of confirmations. When the system can clearly infer the user’s intent (e.g., based on context, environmental information, or historical interaction data), it executes tasks itself and only confirms with the user when necessary.

Example: The system may frequently ask for confirmation during initial use, but as the user repeatedly performs similar actions, the system can skip the confirmation steps.

6.1.2.4 Avoiding over-confirmation during use

For tasks with lower risks, unnecessary confirmations can be avoided. For example, when adjusting the volume, the system can set the volume to a default level without asking for confirmation.

6.1.2.5 Emotionally-Aware dialogue recognition during confirmation

Emotionally-aware dialogue recognition refers to the system's ability to detect the user's emotion (e.g., anger, anxiety, frustration, or joy) during confirmation and to avoid inappropriate and emotional responds.

- *Avoiding user frustration*: When users receive multiple unnecessary confirmation prompts, they may become frustrated or impatient. If the system can detect emotions such as urgency or dissatisfaction in the user's voice, it will skip redundant confirmation steps and expedite task execution.
- *Dynamic confirmation adjustments in Response to negative feedback*: If the system detects user dissatisfaction with its confirmation prompts, it can simplify the process or adjust its responses to reduce frustration and provide quicker execution options.

6.1.3 Dialogue management

6.1.3.1 Handling negative responses

When users issue a negative or corrective command (e.g., changing a task), the system should quickly adjust the task to ensure that the user's intention is accurately executed.

- *Continuous confirmation*: The system should accept consecutive correction commands. For example, if a user says, "Not increase volume, but decrease it," the system should interrupt the current task and promptly adjust.
- *Multi-step clarification*: If the user's correction involves multiple steps or requires further clarification, the system should ask follow-up questions, such as "Would you like to decrease the volume to a specific level?"

6.1.3.2 Wake word settings

- *Adjustable wake words*: The system should allow users to customize wake words according to their preferences. For instance, users could choose phrases like "Hello, Assistant" or "Activate Assistant" to initiate the system.
- *Wake-free scenarios*: In certain contexts, the system can listen to conversations in the background without being woken up. For example, when detecting user queries unrelated to driving, the system could be activated automatically.

6.1.3.3 Multi-tasking capabilities

The system should be able to handle multiple user commands or tasks at the same time, and prioritize them based on their relevance to driving safety to avoid conflicts or interference.

- *Task priority management*: The system should prioritize tasks related to driving safety (e.g., navigation, speed alerts), followed by secondary tasks like entertainment system controls. When users issue multiple commands at once, the system should use a priority algorithm to determine the execution order.
- *Parallel task execution*: The system should be able to execute different tasks in parallel, and it should be possible to interrupt them. For instance, users can adjust the car's audio or air conditioning while navigating.

Voice interruption handling Voice interruption handling allows the system to pause the current task when interrupted by the user and resume it.

- *Pause and resume*: The system should handle user interruptions, such as pausing a navigation setting to perform another task, and then resuming the previous task when permitted by the user.
- *Multi-task interruption*: Users should be able to interrupt one task to address a more critical one. The system should automatically manage the resumption order based on priority. For example, a user might interrupt navigation to adjust the volume, after which the system resumes navigation.

Open-Ended dialogues During non-critical driving activities or the use of secondary driving-related functions, open-ended dialogues allow users to ask unstructured questions or requests freely. The system should interpret and respond to these open-ended dialogues based on the command content and context.

Endpoint detection Endpoint detection refers to the system recognizing the end of a user's speech, determining when to stop listening and proceed with command execution.

- *Pause detection*: By detecting natural pauses in speech, the system can determine when the user has finished speaking. When a sufficiently long pause is detected, the system will assume the user has completed their input.
- *Contextual judgment*: The system should use context to avoid terminating the conversation prematurely. For instance, when a user pauses to think while answering a complex question, the system should give the user more time and allow the user to continue speaking.

6.1.4 Feedback

6.1.4.1 Exception handling

The system should be able to handle unexpected situations and inform the user how to address them through voice or visual prompts. Clear exception feedback should be provided, such as "I don't understand, please repeat," with pre-designed content.

6.1.4.2 Delay management

When the system experiences delays in responding, it should notify the user with "Processing your request" to prevent the user from assuming the system is crashed. Timely feedback is essential to reducing user stress, especially during driving, where delayed responses can cause uncertainty and tension (Wickens, Helton, Hollands, & Banbury, 2021).

6.1.4.3 Disambiguation

When the system cannot clearly understand the user's command, it should clarify or ask follow-up questions to ensure the correct task is executed.

Clarification prompt: When a command has multiple interpretations, the system should ask for a brief clarification. For example, if a user says "Play music," the system might clarify, "Would you like to play from the radio or the music library?"

Dynamic feedback: The system should adapt to various contexts, providing concise clarifications. For example, when a user says "Increase temperature" while driving, the system should quickly clarify, "Would you like to raise the air conditioning temperature to a specific degree?"

6.1.5 Non-verbal and visual confirmation

Non-verbal confirmation refers to the acknowledgment of a user's command through actions or feedback without using verbal communication. It can be done by executing the corresponding task or through visual feedback. For example, during an operation, instead of using voice prompts, the system can confirm task completion by directly performing the action. Visual confirmation, on the other hand, uses display screens or other visual elements (such as projection or lighting) to confirm the user's actions and choices. This approach reduces the cognitive load of voice interaction for the user, making it intuitive when handling complex or multiple tasks.

- *Delay management*: In non-verbal confirmation, the system must manage response delays, particularly in situations requiring quick feedback. The design should ensure timely feedback to avoid users perceiving that the command has not been executed.
- *Multi-model feedback:* When designing the system, it is important to minimize information overload, such as by displaying information step by step or simplifying interface elements to help users quickly confirm their actions.

6.2 Technical requirements for VUI

From user study, the demand for VUI can be discussed from several perspectives. Based on the interview findings, it is categorized into three areas: Technical Requirements, Operational Requirements, and Expectations for Characteristics. The tables below shows the details.

6.2.1 Technical need

Technical needs refer to the system capabilities and performance that ensure VUI functions efficiently and effectively. These needs focus on how the technology is designed to meet user expectations.

Specific Need	Definition	Rate
Adaptive and Prediction	The ability of the VUI to adapt to user preferences and predict user needs for a more personalized experience.	5
Feature Expansion and Flexibility	Expansion of VUI functionalities to accommodate more diverse and flexible user requirements.	4
Humanized Dialogue	Creating more natural and engaging conversations that mimic human interaction.	5
Optimized Performance (no latency)	Improving system performance to eliminate delays and enhance responsiveness.	5
Recognition and Understanding Enhancement	Enhancing VUI's ability to accurately recognize and understand diverse user inputs, including accents and speech variations.	4
Specialized Functional Adjustments	Adjustments to the VUI to accommodate specialized tasks and specific user needs.	3

6.2.2 Operational need

Operational needs focus on how users interact with the VUI in their everyday activities. This category emphasizes the system's usability, interface customization, and how well it supports the user's daily operational tasks.

Specific Need	Definition	Rate
Customization and Personalization	Allowing users to customize the interface and responses according to their preferences.	5
Multimodal (Hybrid Controls)	Supporting multiple modes of input, including voice, touch, and hybrid controls for flexibility.	4
Tutorial Instructions of Using	Providing clear and comprehensive instructions for using the VUI effectively.	4
Visualization and Design	Designing the VUI interface in a visually appealing and functional way.	3

6.2.3 Character expectation

Character expectations relate to the persona or behavior of the VUI as perceived by the user. These needs focus on the emotional and psychological aspects, aiming to build trust and comfort.

Specific Need	Definition	Rate
Careness	Ensuring the VUI shows empathy and understanding towards user emotions and needs.	5
Different-characters	Offering varied personalities or styles for the VUI to cater to different user preferences.	4
Exploratory	Encouraging exploration and discovery through interactive and engaging VUI features.	4
Fast and Correct	Ensuring the VUI provides quick and accurate responses to user commands.	5
Intelligent	Developing a VUI that can learn and understand user behaviors and adapt accordingly.	5
Naturally	Making the VUI interactions feel natural and seamless, similar to human conversations.	4
Proactive	Designing the VUI to be proactive in suggesting actions or options to the user.	4
Reliable	Ensuring the VUI operates reliably under various conditions and provides consistent performance.	5
Trust	Building trust through secure, transparent, and dependable VUI interactions.	5
Unique	Creating a unique and distinctive VUI personality that stands out.	4
User-friendly and Convenient	Designing the VUI to be easy to use and navigate, providing a comfortable user experience.	5

6.3 Use case and scenario

This part details how to configure VUI for controlling and managing various car features by organizing usage scenarios into three stages — before use, during use, and after use.

6.3.1 Before use

To begin with, driving tasks and use scenarios are categorized for the various features. Then the user sets up the controllable features. It is important to define the range of each function, especially for those requiring precision, where users are prompted to configure basic settings and their default values.

The system also allows customization of command inputs and shortcuts. Furthermore, the interface can recommend or automatically set defaults for frequently used functions, using data from user research to predict usage patterns. Finally, voice interaction touchpoint can be mounted in the certain slot in the car interior like Figure 11, and configured to act as control points, with options for enabling or disabling controls, as well as integrating other physical or touch-based control switches.

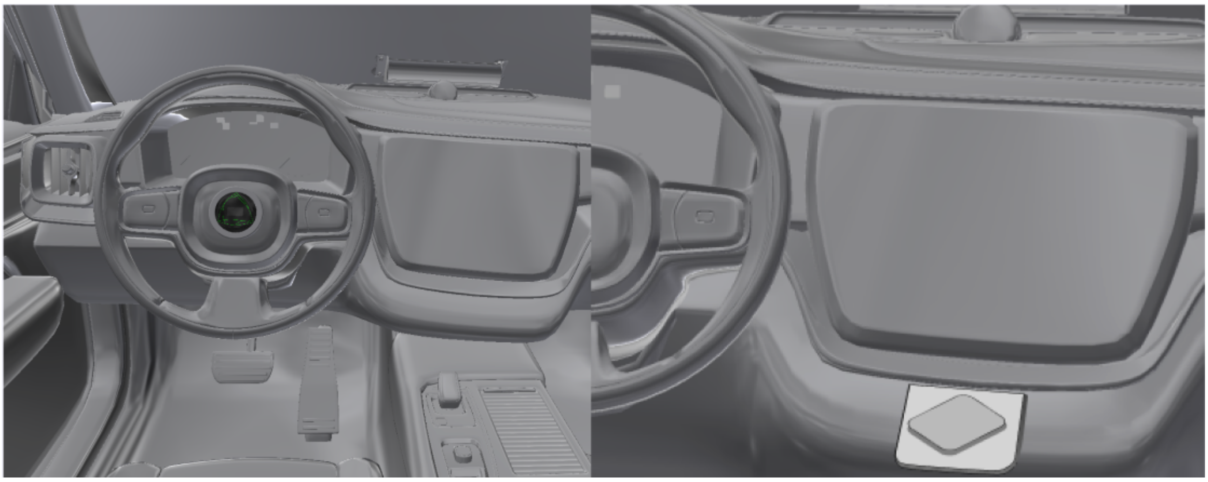


Figure 11: VUI hardware and its position in the interior

The Before Use process is as follows:

1. Classify the driving task categorization for the features.
2. Configure settings within the car's HMI, as detailed in Table 18. Examples of settings include:
    - Set up the features that can be controlled(Secondary and Tertiary driving tasks) in Car HMI and hardware concept (Figure 12). The default value is set to "open."
    - Set the range within which the functions can be controlled (For example, for functions requiring precise control, ask the user to perform basic settings, including setting default values.)
        - Recommending features or set defaults for frequently used features.
3. Customize command input/shortcut input.
4. Set endpoint for controlling the voice interaction interface (Evaluating whether it can be used as a control endpoint).

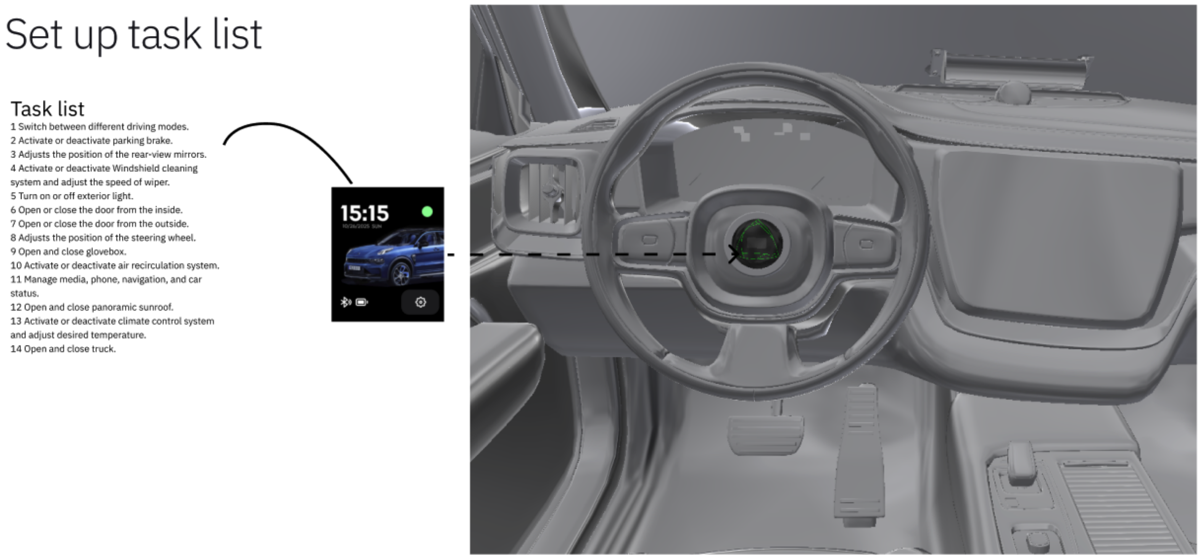


Figure 12: Setting up the features through the VUI hardware concept

Table 18: Overview of Categories and Descriptions for Car HMI System, Touchpoint GUI, and VUI Design

Category	Description
Car HMI System and Touchpoint GUI	**Function Activation Interface/Input Settings** **Usage Scenario Selection:** Design activation or deactivation shortcuts for different functions (e.g., navigation, entertainment, driving assistance) **Touchpoint GUI:** Interactive touchpoint device GUI showed in Figure 13.
VUI	**Voice Prompt Input:**Option to enable or disable voice prompts, allowing users to control system feedback, such as voice guidance or reminders. **Custom Command Input Confirmation:** Enables users to input custom commands or shortcuts, with the system providing confirmation and responses to enhance personalization.

- Displaying in Car HMI
- Other physical touchpoint control.

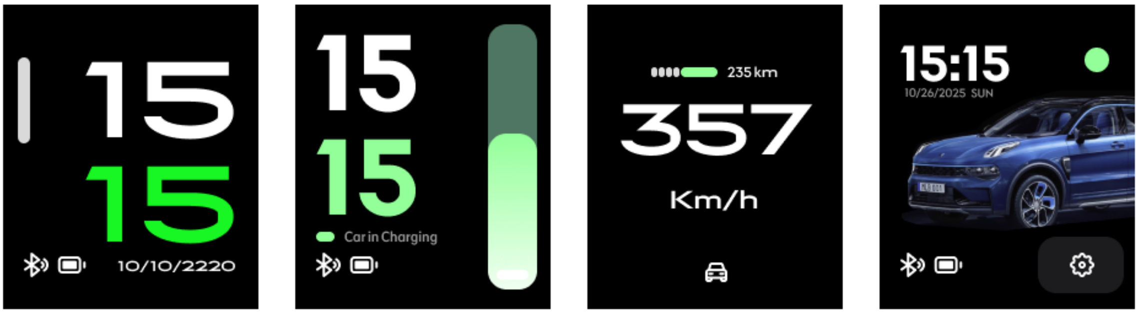


Figure 13: Touchpoint GUI preview

6.3.2 During use

This process describes the steps how voice interaction systems work in car features control during usage. It begins with detecting both internal and external conditions, such as the driving environment, the status of passengers, and the car's settings. The system then activates the voice interaction interface, ensuring that it is functioning properly and can handle user input. Once commands are detected, the system analyzes them to determine whether they involve simple actions like on/off toggling or more complex adjustments that may require default settings or further clarification. If the system is unsure of a command, it seeks confirmation from the user, offering guidance for multi-step functions when necessary. Throughout this process, the system maintains a continuous dialogue with the user, offering feedback and context-specific prompts to create a seamless and intuitive interaction. The overall goal is to enhance both safety and responsiveness for features controlling by ensuring that the system efficiently adapts to user needs and the driving environment.

The During Use process is as follows:

1. Detect driving status and the passengers' conditions.
2. Detect the external environment (environmental monitoring and perception).
3. Detect if an interactive physical touchpoint is being used.
4. Detect VUI User.
5. Detect current car settings and car status:
    - No specific feedback.
    - Prompt the user (via voice) to configure functions that may need a state change.

- Task distribution guidance (under specific conditions, such as when parked) for multi-level functions.

6. Wake-up (activate/feedback from VUI) — check if the features are working properly and if they are controllable.

7. Detect and analyze user input commands and Visualize the process in the hardware (Figure 14):

   - Check whether the function is restricted from voice control.
   - Check whether the function requires fine adjustment settings.
     (a) If it's a simple control (like on/off), the command should consist of the function name and a simple toggle (on/off).
     (b) If the function requires fine-tuning:
       – When the system detects a simple command for switching, it triggers the default settings.
       – If the user provides specific commands, the system won't trigger the default settings.
   - Check whether multitasking is involved.
   - Check for specific/custom commands.

8. Command confirmation and processing:

   (a) If the system understands the user's command, it will directly adjust without confirmation.
   (b) If the system is uncertain about the user's command, it will ask for confirmation.
   (c) If the command is ambiguous or incomplete, the system can further ask the user for clarification.
   (d) Immediate command suggestions.

9. Function execution/system operation:

   (a) Normal execution.
   (b) Interrupt execution and return the function to the previous state or default value.
   (c) When the user enables or disables a function, the system returns it to the default value.

10. If confirmation or processing issues arise:

    (a) Notify the user promptly, "Processing your request."
    (b) Clarify or further inquire to ensure the correct task is executed, allowing the user to return to the previous step.

11. Continuous conversation:

    - User confirmation status (simple feedback, including silence or unrelated dialogues).
    - Continuous dialogue control loops (context-related prompts).

12. Check for status notifications (visual or auditory),and check whether smart dialogue/guidance is needed.

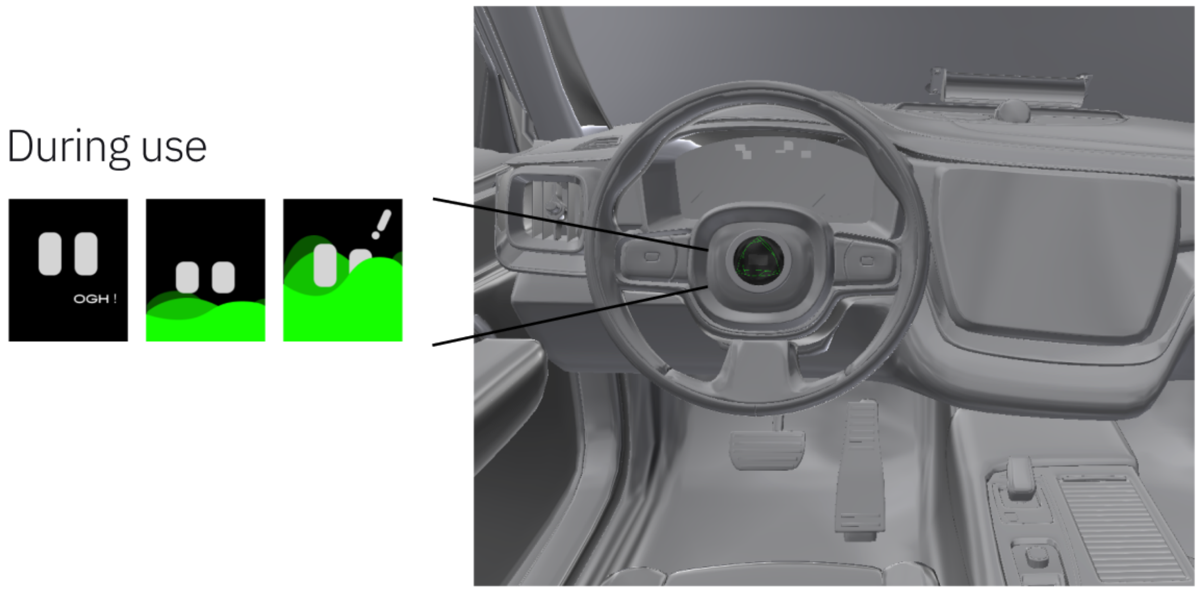


Figure 14: Interaction and Visualization feedback for the process

6.3.2.1 Interaction

We designed a human-like facial expression system to indicate the state of interaction in Figure 15. This approach uses facial expressions to provide feedback, helping users understand whether the system is processing, ready for input, or responding to commands. For example, a neutral or focused expression might indicate that the system is actively listening, while a smile or nod means successful completion of a task. This human-like feedback makes the system more responsive and emotionally satisfies users' needs.

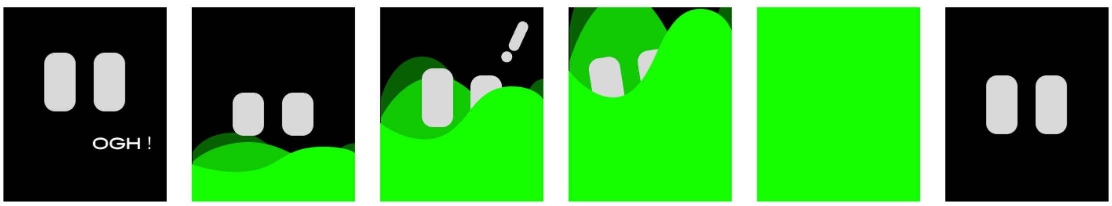


Figure 15: Human-like facial expression system for VUI

6.3.3 After use

Once the function is set or the task is complete, the system offers multiple feedback methods to confirm successful operation. These include voice confirmation that informs the user, as well as multi-modal feedback that enhances the interaction experience through visual and haptic feedback from VUI hardware design. For example, visual feedback may involve changes in the display, while haptic feedback could involve vibrations or other tactile responses. Additionally, other feedback mechanisms, such as conversational responses about the current system status, can further enhance interaction. Through generative dialogue, the system can offer more detailed information about functions when it is safe to do so, although this feature may require further testing in real-world conditions.

1. When the function is completed, various feedback methods are provided:
   (a) Voice confirmation feedback.
   (b) Multimodal status feedback:
      - Visual feedback in the VUI Design (Figure 16).
      - Haptic feedback.
   (c) Other feedback (such as dialogue about the current state) to enhance interaction through advanced AI model for chatting, providing some degree of function-related content when it is safe to do so *(pending real-world verification)*.

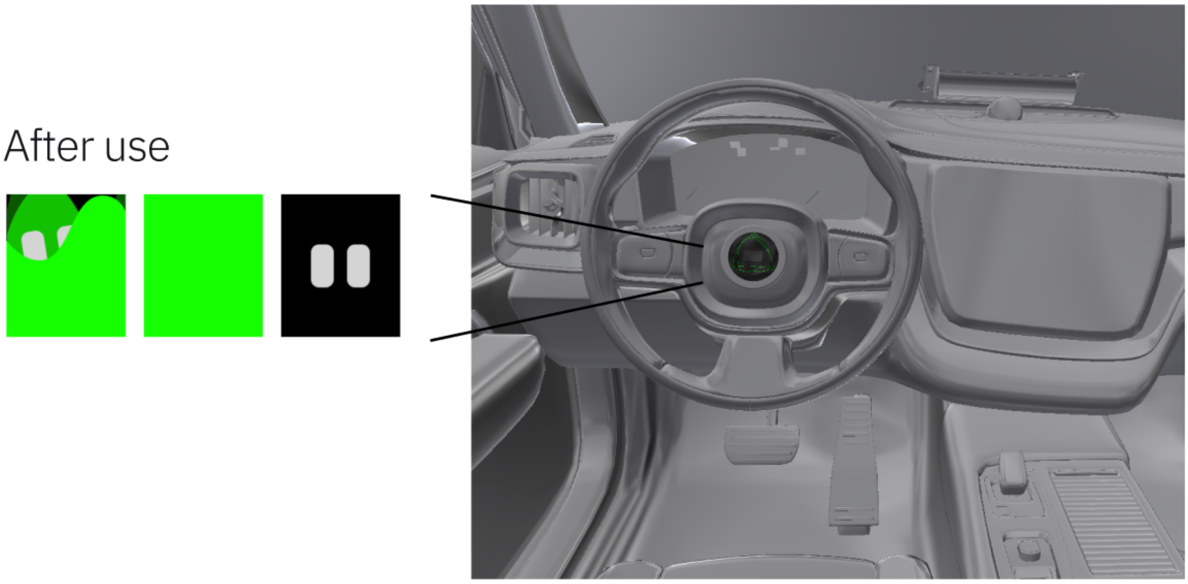


Figure 16: Visual Feedback and Interaction on the VUI hardware concept

6.4 Concept development

In the design practice stage, it is essential to determine the carrier of VUI, which can make VUI easier to use and contribute to a better user experience. As a result, the following three aspects were evaluated.

- User needs and preferences from user study
- Current VUI products in the market
- Data collection and analysis

During the user study, the majority preferred interacting with VUI through a physical product. Some interviewees noted that, compared to virtual interfaces or avatars on car displays, they preferred a physical product due to ”a more emotional connection”.

In the current market, there are two different kinds of mainstream in-car VUI products. The first one exists virtually in the car system without an independent user interface or system, which is currently the most common way, such as Mercedes-Benz’s MBUX Virtual Assistant. The other kind has its own user interface which is integrated into physical devices, such as NIO’s product NOMI. Additionally, some tech companies, like Google and Xiaomi, integrate VUI into speakers. As we can see in table 19, voice assistants from different car manufacturers differ in product image, form, customization and proactive interaction. Compared to the VUI existing in the car system as a sub-interface, the one integrated into physical devices is fewer in the market. Also, there is more room to design physical products such as a car key which belongs to car accessories.

Table 19: Some Car OEM VUI products on market

Brand	**VUI Products**	**Product Image**	**Anthropomorphic Emoji**	**Customization**	**Proactive Interaction**
XPeng	Junior P	Circular screen	Yes	Support	-
NIO	NOMI	Physical robot head	Yes	Support	Yes
Lynk&Co	JOJO	3D avatar	Yes	Support	Yes

The data for the study comes from two sources: a questionnaire and an offline test. The questionnaire was sent to the interviewees of the quantitative study regarding their preferences for different product images. The test focuses on different product designs and was carried out offline. Four VUI carriers were presented to the interviewees, including a few product models and some potential VUI physical carriers. The participants need to vote for two out of the four options, including display, phone, driving recorder, and portable accessories like car key or band. Their choices were collected and analyzed to obtain their preferences for product designs.

In the questionnaire, while 24 out of 86 respondents chose VUI that exists virtually in the car system, most preferred other types of interaction. For example, 27 respondents chose smart phones and 29 chose portable devices like car key or band. In the test, 18 out 25 participants voted for portable devices, followed by car screen and smart phone, which both received 15 votes. Participants expressed a preference for new product forms or specialized VUI devices over traditional standard voice interactions in cars or phones.

Overall, it is a suitable and effective design decision to choose portable hardware device as a VUI carrier. To meet sustainability goals, the design will focus on existing in-car accessories. The car key or a portable device that functions as a key will be the focus of our design.

6.4.1 Modeling and rendering

A mood board was created on the web application *Figma* to brainstorm and combine different design elements as it shows in Figure 17. To create the digital model, computer-aided design was conducted using the software *SolidWorks*. Different elements and ideas are shown in the mood board and some of the ideas became concepts with different development levels.

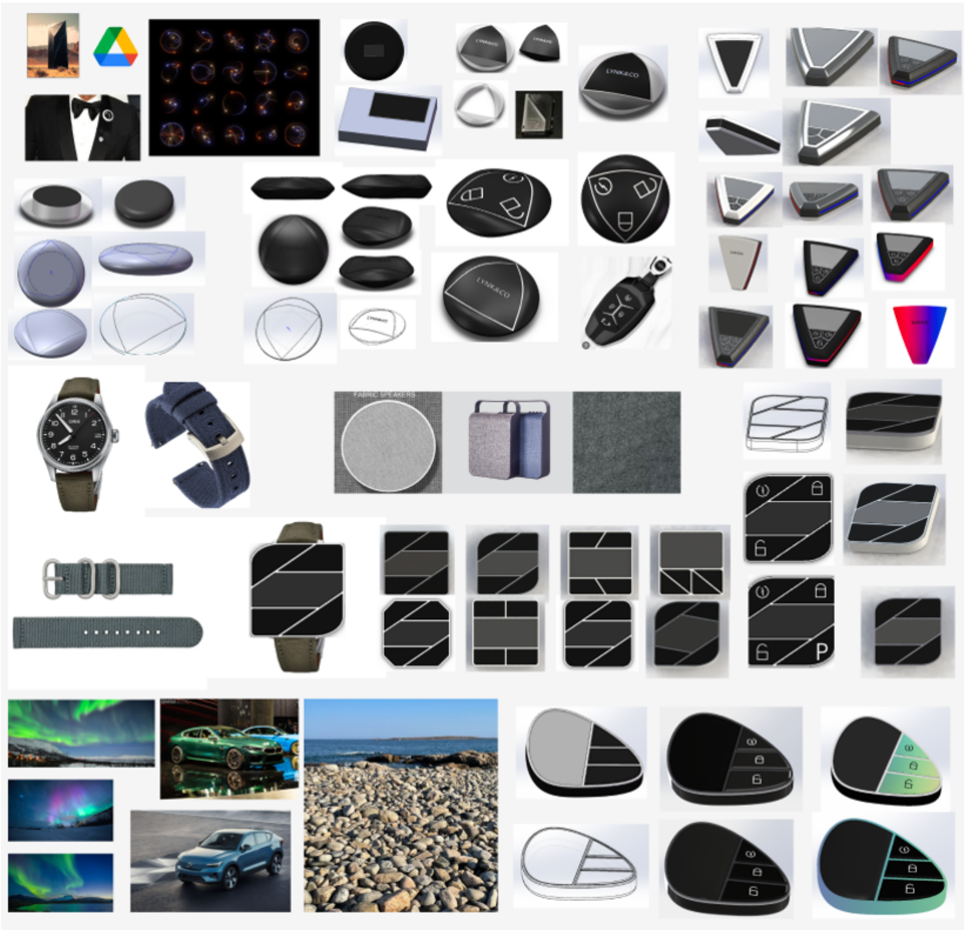


Figure 17: Moodboard for Design

A car keyfob was selected as the carrier of the voice assistant. The early design varied between different basic shapes and the round shape was selected to move on to the next stage, as it shows in Figure 18. Some of the alternatives in the early stage are presented in Figure 19.

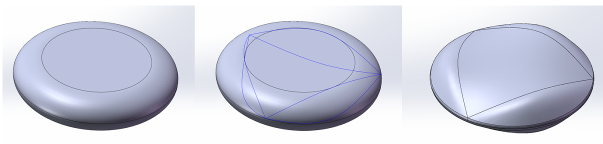


Figure 18: Basic shape of the design

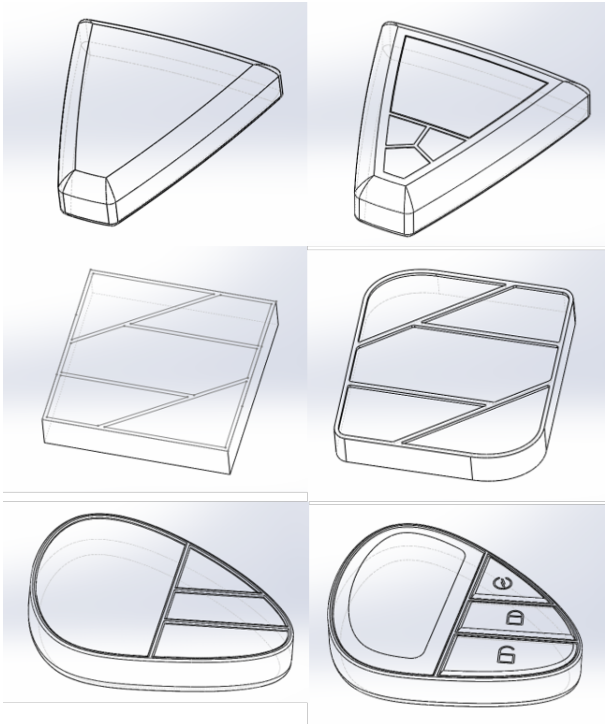


Figure 19: Early alternatives design concepts for VUI touchpoint

Then the logo of the company is added to the design. The color black and white are selected as the basic theme to match the color of the logo. Some extra lines are added as it shows in Figure 20 and one of them is selected for further design. The final design is shown in 21 where a digital display and three buttons related to ”Lock”, ”Unlock” and ”Charge” are added to the design.

Figure 20: Early stage of the design

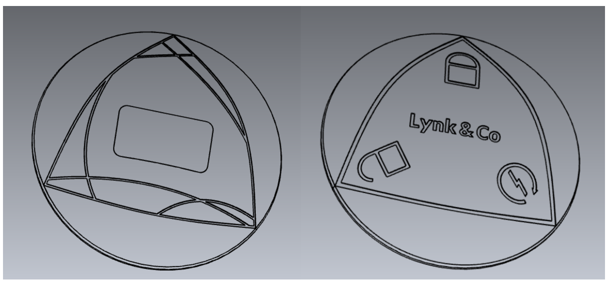


Figure 21: Final design

Then design was divided into three different contexts including city, family and nature as it shows in Figure 22, 23 and 24. The characteristics of the three contexts are summarized as follows:

- **City:** Fashionable, with strong contrasts and modern elegance; lighting serves as a strong brand element/expression; smart richness, without opulence.
- **Family:** Comfortable, with soft materials; durable and resistant to heavy use; washable and made from recycled materials.
- **Nature:** Easy, achievable adventures, without extremes; comfortable, glamping, and relaxing time.

Figure 22: Variants in the context of "City"

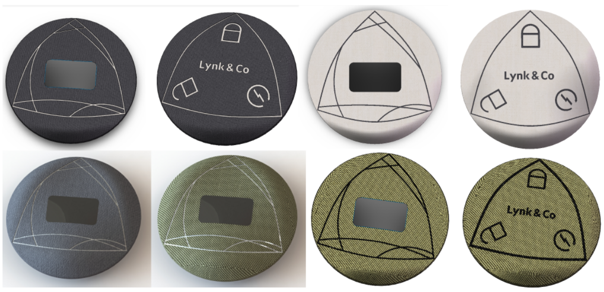


Figure 23: Variants in the context of "Family"

Figure 24: Variants in the context of "Nature"

Other than car key fob itself, we also tried a combination of car key and watch, inspired by the survey result in the preparation. The design process and variants are shown in Figure 25.

Figure 25: Combination of car key and watch

6.5 Summary

This section presents a set of design guidelines and principles for developing effective VUI tailored to the control of features across various driving scenarios. It emphasizes the importance of designing systems that can be operated solely through voice commands, ensuring both functionality and user satisfaction. The guidelines are centered on three core aspects: clarifying system functionality, adapting to diverse driving conditions, and enhancing user satisfaction through intuitive and personalized interaction designs. Key recommendations include minimizing confirmation steps based on the user's familiarity with the system, dynamically managing tasks with a priority on safety-critical functions, and utilizing multi-modal feedback mechanisms to enhance overall usability.

From a technical perspective, the requirements focus on improving the adaptability of VUI components, such as advancing Automatic Speech Recognition capabilities, optimizing system performance, and ensuring consistent output quality. On the operational side, the guidelines highlight the need for customizable interfaces and intuitive user tutorials to facilitate ease of use. Additionally, the semantic design considerations emphasize the integration of emotional feedback, enabling the VUI to respond contextually to the user's emotional state and situational needs.

This section also explores the design of touchpoint concepts, providing an illustrative example in the form of portable devices such as car keys, which offer convenience and portability. By integrating emotionally intelligent and contextually adaptive GUI elements, the VUI system is designed to deliver seamless, responsive, and trust-enhancing interactions. This comprehensive approach not only tackles technical and operational challenges but also places a strong emphasis on the emotional and experiential aspects of user interaction. As a result, it establishes a solid foundation for future innovations in VUI design.

7 Discussion and Conclusion

This chapter presents the key findings of this study. Additionally, we discuss the challenges of VUI implementation, the study's limitations, and responses to the research questions. Finally, future work in this direction is suggested.

7.1 VUI for controlling secondary and tertiary driving related features

The integration of VUI technology into automotive systems represents a transformative shift in how drivers interact with their vehicles, with an emphasis on enhancing both safety and convenience. This transformation is primarily driven by the need to reduce cognitive load and minimize distractions—two critical factors that significantly impact driver performance and safety. Research consistently shows that users generally prefer employing VUI to manage secondary and tertiary driving tasks, such as controlling navigation, entertainment systems, and other non-driving settings. Compared to traditional physical controls like buttons or dials, VUI systems for managing these features have been shown to offer comparable or even enhanced usability. For example, some participants in user studies noted that tasks such as adjusting the air-recirculation system could be easily handled through voice commands, eliminating the need for dedicated physical buttons. Moreover, when VUI systems provided intuitive and reliable interaction, user willingness to rely on voice control for these tasks increased significantly. Additionally, VUI systems demonstrated distinct advantages in multitasking and managing multi-level operations, outperforming physical buttons in terms of efficiency and user satisfaction. In the user study, several participants expressed concerns about the complexity of driving tasks and the need for heightened focus. In these situations, a lack of confidence in driving, combined with the fear of making operational errors, caused them to hesitate and avoid relying on manual physical controls. In contrast, VUI provides a more natural interaction, reducing psychological pressure and minimizing distractions while driving. Wickens' Attention Model highlights the distinction between cognitive processes related to spatial and linguistic information (Wickens, 2008). Unlike physical interactions, which require simultaneous processing of both spatial and linguistic cognition, voice interaction primarily engages linguistic processing. This separation allows for more efficient concurrent processing of spatial and linguistic information, thereby reducing cognitive complexity.

Nevertheless, participants in the study noted that the shift from physical buttons to voice control, particularly for secondary driving tasks features, should not be immediate. Several participants highlighted that a complete removal of physical controls could hinder adaptation, especially for those who rely on tactile feedback. Therefore, further research is essential to identify effective strategies for minimizing physical controls in design and engineering, ensuring smooth transitions for users in real-world driving contexts.

7.2 VUI for primary driving tasks and related features

In terms of primary driving tasks, users exhibited a clear reluctance towards utilizing VUI for features related to primary driving operations. This resistance reflects a significant concern and distrust regarding the substitution of human intervention with voice control in such critical tasks. The study found that users were particularly apprehensive about the reaction time, accuracy, and stability of voice systems in complex, fast-paced driving environments, fearing that these systems might not provide the level of safety required. Some researchers emphasized that system reliability and response time are crucial factors influencing driver trust in voice interaction for high-stakes tasks (Young & Regan, 2007). Drivers, in particular, are hesitant to rely on VUI for core tasks such as speed control or emergency braking, due to a lack of confidence in the system's performance during critical moments. According to Fitts and Posner's three-stage learning theory (1967), users remain in the "cognitive stage" when interacting with new technologies in dynamic environments like driving. At this stage, they have yet to build a reliable mental model of the technology, and they hold high expectations for its reliability and accuracy. As a result, when confronted with high-risk tasks, users are more likely to depend on familiar, trusted physical controls like buttons.

From a psychological perspective, Norman's emotional design theory (2013) suggests that users experience stronger emotional responses during high-risk tasks, particularly when safety is at stake. They tend to gravitate toward intuitive, physical control mechanisms that offer a sense of security and control. In contrast, the virtual feedback from voice interaction often fails to provide the immediate emotional satisfaction and reassurance that physical interfaces can deliver.

Thus, while VUI offers distinct advantages for controlling auxiliary functions, caution is warranted when considering its potential to fully replace physical controls for primary driving tasks. Further, more detailed research is necessary to explore this issue and better understand the boundaries of VUI integration.

7.3 Contextual and environment awareness

Participants in the study emphasized that in-vehicle VUI design should accommodate the needs of various stakeholders and use scenarios. To achieve this, the VUI should be capable of perceiving, interpreting, and responding to both internal and external environmental factors across different contexts. Contextual awareness is key—this involves designing the system to adjust based on different driving phases, such as differentiating between dynamic and static vehicle conditions, or customizing interaction modes for pre-driving, driving, and post-driving stages. Furthermore, the range of voice-controlled features should be customized for different users. For example, the driver might have access to all vehicle settings, while passengers could be restricted to controlling specific, non-critical features.

7.4 Enhancement in user experience

In the qualitative study, we found that designing interactive entities for VUI can significantly enhance the user experience during operation. Participants noted that while physical buttons could be replaced by voice interfaces, an interactive entity or object remains essential for visualizing processing progress. When designing such entities, the visual appeal—including interface and virtual character design—plays a crucial role in user satisfaction, consistent with the findings of Mishra et al. (2022) on voice assistants. For interface design, a well-structured layout that minimizes cognitive load is key to improving usability. Regarding virtual character design, several participants indicated that these designs elicited positive emotional responses during interactions. As Frijda (1986) observed, emotions help individuals position themselves in relation to their environment, drawing them toward certain people, objects, actions, and ideas, while pushing them away from others. Positive emotions, therefore, play a vital role in establishing a strong emotional connection with desirable products (P. Desmet & Hekkert, 2007).

Furthermore, the ability to customize the system significantly influences user acceptance during interaction. VUI that is tailored to a user's individual characteristics aids in adapting the interaction process to suit specific needs. For instance, a voice interface could be configured to control a range of features customized to the user's preferences and driving habits, leveraging user data (such as driving patterns) to provide personalized presets and semantic understanding.

7.5 Challenges in VUI implementation

While participants expressed optimism about using VUI for controlling in-car functions during user research, several challenges persist, particularly in the complex driving environment and high cognitive load situations. One major challenge is the VUI's technical ability to accurately understand and process natural language in varying conditions, which is especially crucial in the context of driving. Previous studies highlight that the accuracy of voice interactions significantly influences user trust and reliance on the system (Young & Regan, 2007). To address these challenges, comprehensive design guidelines are important. These guidelines will help designers identify and resolve potential issues, thereby maximizing the efficiency benefits of voice as an interaction medium. Key focus areas should include improving voice recognition capability, providing clear system feedback, and implementing confirmation mechanisms to ensure users are aware of the status of their interactions.

Additionally, given the complexity of designing and implementing VUI to control vehicle features, designers may encounter difficulties in fully understanding the technical requirements and capabilities needed for effective integration. As a result, collaboration with multidisciplinary teams during the development process is critical to ensure both the technical integrity and practical feasibility of the system. Developed design guidelines will aid designers in planning and evaluating crucial design elements more effectively, allowing for quicker validation of design feasibility.

7.6 Research limitation

In both quantitative and qualitative user research, increasing the sample size and incorporating a more diverse group of participants will greatly improve the generalizability and reliability of the findings. This study had a relatively small sample, and most participants were experienced drivers. While their insights were valuable, their familiarity with vehicles and related technologies may have introduced an "idealization bias," where they reported their expectations of the system rather than their actual experiences.

Quantitative research examines a wide range of influencing factors, while qualitative research provides deeper insights into users' motivations and behavioral patterns. However, the complexity of VUI systems, with their specialized features, made it challenging for this study to fully explore how these features interact. The absence of a detailed analysis of these specific interactions limits our understanding of user preferences and potential needs in different driving scenarios.

In the design and execution of the comparative study, while this research provided valuable insights into the use of VUI in driving environments, several limitations need to be addressed. First, the data was mainly collected from simulated driving environments, which affects how well the findings apply to real-world scenarios. Although simulations help control variables, they don't capture the full complexity of real driving, where factors like traffic, weather, and road conditions can significantly impact VUI performance. Additionally, the VUI technology used in the experiment may not fully reflect the current state of the technology or how people use it. Experimental setups are often in controlled environments, which don't account for everyday challenges like background noise, different ways of speaking, or varying language habits that could affect VUI performance. Features like interaction frequency and voice recognition accuracy are often tested under ideal conditions, while real-world use might be affected by factors like device performance and network connectivity. This creates a gap between how the technology works in the lab and how it performs in real-life situations, so the results may not fully reflect users' actual experiences. Furthermore, some participants may have given "idealized" responses based on expectations rather than experience, which can skew the findings and make it harder to accurately understand users' real needs.

The participant group in this study had a similar professional background, which limits how broadly the results can be applied. Most participants were highly educated, which may have influenced their comfort with technology and their problem-solving abilities. To draw more accurate and comprehensive conclusions, future research should include a wider range of participants from various professions, education levels, and levels of comfort with technology. Expanding the sample size and making it more diverse will help balance the results and improve how well they represent the general population.

Additionally, the study did not fully account for how different driving cultures across countries and regions might affect user behavior. Different traffic laws, cultural habits, and road conditions can strongly influence how people accept and use VUI technology. For example, driving conditions in China are quite different from those in Europe, which directly affects driving habits and technology needs.

Finally, as a concept design, the design of the VUI is still in the early stage of a product design by exploring ”what could be” without diving into the detailed engineering part of how the product will actually be made. The VUI concept mainly focused on the idea generation, sketching, modeling, and rendering for the product. There is still a long way to go before putting it into practical use.

7.7 Answering the research questions

This thesis investigates the feasibility of using VUI to replace physical controls, such as buttons, for managing driving features. Specifically, it examines whether VUI can effectively replace physical controls in executing primary, secondary, and tertiary driving functions. Additionally, the thesis presents a model that illustrates the factors influencing user choices and, through experimental analysis, identifies which driving features are most suitable for VUI control.

To provide a comprehensive understanding of this area, the study adopts a mixed-method approach, combining user research with experimental testing. Both quantitative and qualitative data were gathered via surveys and user interviews. The qualitative data were analyzed through thematic analysis, while the quantitative data were processed using various statistical methods. The results reveal that participants had a positive attitude toward using VUI for managing secondary and tertiary driving tasks. However, for primary driving functions, users expressed strong opposition, and experimental data indicated that VUI could not match physical controls in these scenarios. The study also identifies several key human factors—such as user experience, expectations, attitudes toward the product, safety concerns, and product functionality—that significantly influence users' willingness to adopt VUI for in-car control.

Moreover, the study highlights that users generally perceive in-car VUI as a promising tool for enhancing safety and convenience while driving. Despite this, qualitative analysis shows that users still require an interactive entity as a touchpoint for engaging with the VUI. To improve the overall experience, the study strongly recommends designing a more flexible, personalized, and user-centric interaction model.

In light of the findings obtained through the research, three research questions mentioned in the beginning could be fulfilled:

1. What specific driving tasks can be conducted using VUI?

First, several driving tasks of different kinds were collected. A multiple linear regression analysis on user experience proved the tasks in table 20 can be conducted using VUI, which are secondary and tertiary tasks. Primary driving tasks cannot be conducted using VUI due to safety reasons, which users prefer using physical controls.

Table 20: Tasks that can be conducted using VUI

No.	Tasks
1	Switch between different driving modes.
2	Activate or deactivate the parking brake.
3	Adjust the position of the rear-view mirrors.
4	Activate or deactivate the Windshield cleaning system and adjust the speed of wiper.
5	Turn on or off the exterior light.
6	Open or close the door from the inside.
7	Open or close the door from the outside.
8	Adjust the position of the steering wheel.
9	Open and close the glove box.
10	Activate or deactivate air recirculation system.
11	Manage media, phone, navigation, and car status.
12	Open and close panoramic sunroof.
13	Activate or deactivate climate control system and adjust the desired temperature.
14	Open and close the truck.

2. Are there any driving tasks that are preferred to be conducted using VUIs? If so, which?

In terms of preference in conducting the above tasks, VUI and physical control perform similarly, showing little significant difference. In other words, there is no strong evidence showing that the features corresponding to these tasks are preferred to be controlled through VUI, considering usability, user experience, and safety. Further, due to this similarity in preference, it is proved that physical control of the features can be replaced by VUI to reduce the need for extensive physical control.

3. How can a concept using VUI be designed and how does user experience affect the design?

The car key fob was selected as the carrier of the VUI according to customer needs, and the VUI concepts were developed based on three different contexts that cover a certain market. In addition, mood board brainstorming, 3D design, and rendering were used to present the ideas and the steps of how the ideas were developed into specific product concepts.

7.8 Future work

In terms of user study, future research should consider how technology familiarity affects user behavior and preferences when analyzing user behavior and technology acceptance. It is suggested to recruit ordinary drivers who are less familiar with VUI, particularly those in the 30 to 60 age group. These users, with their extensive driving experience, can offer more representative and broadly applicable feedback on vehicle usage. Also, expanding the sample size and enhancing its diversity will help reduce the influence of individual participants on the results and improve the generalizability of the study's conclusions. The research findings suggest that the variability in user experience may be related to demographic factors such as age and technological proficiency, which can influence users' interaction with and perception of new technologies (Kumar & Smith, 2018). Therefore, cross-cultural comparisons should also be included to explore how geographic and cultural factors affect user behavior and technology acceptance.

For the VUI concept, the future work should put more focus on the engineering part, starting with the selection of essential hardware components like microphones, speakers, displays, batteries, processors, to the integration and connectivity between these components. It is also essential to do the internal layout design considering the high integration of the hardware components. During prototyping, the material selection and manufacturability have to be assessed. The prototype also needs to be tested in a real car and can be refined by doing user testing based on real-world driving conditions.

References

Agency, E. E. (2025). *New registrations of electric vehicles — european environment agency.* Retrieved from https://www.eea.europa.eu/en/analysis/indicators/new-registrations-of-electric-vehicles (Accessed: 2025-01-03)

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*.

Alba, J. W., & Hutchinson, J. W. (1987). Dimensions of consumer expertise. *Journal of consumer research*, 411–454.

Alternative Fuels Observatory. (2023). *Tesla the best-selling brand in europe, volkswagen group #1 oem.* Retrieved 2025-01-02, from https://alternative-fuels-observatory.ec.europa.eu/general-information/news/tesla-best-selling-brand-europe-volkswagen-group-1-oem

Amazon. (2023). *Learn more about alexa skills.* Retrieved from https://www.amazon.com/b?node=21576558011&ref_=alxcom_lrnmore_btn_23 (Accessed: January 5, 2025)

Attride-Stirling, J. (2001). Thematic networks: an analytic tool for qualitative research. *Qualitative research*, *1*(3), 385–405.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*.

Bauer, R. A. (1967). Consumer behavior as risk taking. *Marketing: Critical perspectives on business and management*, *593*, 13–21.

Baughan, A., Wang, X., Liu, A., Mercurio, A., Chen, J., & Ma, X. (2023). A mixed-methods approach to understanding user trust after voice assistant failures. In *Proceedings of the 2023 chi conference on human factors in computing systems* (pp. 1–16).

Black, W., & Babin, B. J. (2019). Multivariate data analysis: Its approach, evolution, and impact. In *The great facilitator: Reflections on the contributions of joseph f. hair, jr. to marketing and business research* (pp. 121–130). Springer.

Blijlevens, J., Thurgood, C., Hekkert, P., Chen, L.-L., Leder, H., & Whitfield, T. (2017). The aesthetic pleasure in design scale: The development of a scale to measure aesthetic pleasure for designed artifacts. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(1), 86.

Braun, M., Mainz, A., Chadowitz, R., Pfleging, B., & Alt, F. (2019). At your service: Designing voice assistant personalities to improve automotive user interfaces. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–11).

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77–101.

Brooke, J. (1996). Sus: A quick and dirty usability scale. *Usability Evaluation in Industry*.

Buteau, E., & Lee, J. (2021). Hey alexa, why do we use voice assistants? the driving factors of voice assistant technology use. *Communication Research Reports*, *38*(5), 336–345.

Chen, Q., Gong, Y., & Lu, Y. (2023). User experience of digital voice assistant: Conceptualization and measurement. *ACM Transactions on Computer-Human Interaction*, *31*(1), 1–35.

Cohen, M., Giangola, J. P., & Balogh, J. (2004). Voice user interface design.. Retrieved from https://api.semanticscholar.org/CorpusID:109101615

Cohen, M. H., Giangola, J. P., & Balogh, J. (2004). *Voice user interface design*. USA: Addison Wesley Longman Publishing Co., Inc.

Demir, E., Desmet, P. M., & Hekkert, P. (2009). Appraisal patterns of emotions in human-product interaction. *International journal of design*, *3*(2).

Desmet, P. (2002). Designing emotions. *Delft University of Technology*.

Desmet, P. (2003). A multilayered model of product emotions. *The design journal*, *6*(2), 4–13.
Desmet, P., & Hekkert, P. (2007). Framework of product experience. *International journal of design*, *1*(1), 57–66.
Desmet, P. M., & Hekkert, P. (2002). The basis of product emotions. *Pleasure with products, beyond usability*, 60–68.
Dong, J., Lawson, E., Olsen, J., & Jeon, M. (2020). Female voice agents in fully autonomous vehicles are not only more likeable and comfortable, but also more competent. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 64, pp. 1033–1037).
Fischer, J. E., Reeves, S., Porcheron, M., & Sikveland, R. O. (2019). Progressivity for voice interface design. In *Proceedings of the 1st international conference on conversational user interfaces* (pp. 1–8).
Fitts, P. (1967). Human performance. *Brooks/Cole*.
Fokkinga, S., Desmet, P., & Hekkert, P. (2020). Impact-centered design: Introducing an integrated framework of the psychological and behavioral effects of design. *International Journal of Design*, *14*(3), 97.
Frijda, N. H. (1986). The emotions. *Studies in Emotion and Social Interaction*.
Ghosh, D., Foong, P. S., Zhang, S., & Zhao, S. (2018). Assessing the utility of the system usability scale for evaluating voice-based user interfaces. In *Proceedings of the sixth international symposium of chinese chi* (pp. 11–15).
Gordon, M., & Breazeal, C. (2015). Designing a virtual assistant for in-car child entertainment. In *Proceedings of the 14th international conference on interaction design and children* (pp. 359–362).
Grosinger, J. (2022). On proactive human-ai systems. In *Aic* (pp. 140–146).
He, F., & Burns, C. M. (2022). A battle of voices: A study of the relationship between driving experience, driving style, and in-vehicle voice assistant character. In *Proceedings of the 14th international conference on automotive user interfaces and interactive vehicular applications* (pp. 236–242).
Hekkert, P. (2006). Design aesthetics: principles of pleasure in design. *Psychology science*, *48*(2), 157.
Heritage, J. (2015). Well-prefaced turns in english conversation: A conversation analytic perspective. *Journal of Pragmatics*, *88*, 88–104.
International Organization for Standardization. (2018). Ergonomics of human-system interaction – part 11: Usability: Definitions and concepts (Computer software manual No. ISO 9241-11:2018). Geneva, Switzerland. (Available at `https://www.iso.org/standard/63500.html`)
Isyanto, H., Arifin, A. S., & Suryanegara, M. (2020). Performance of smart personal assistant applications based on speech recognition technology using iot-based voice commands. In *2020 international conference on information and communication technology convergence (ictc)* (pp. 640–645). Jeju, Korea (South). doi: 10.1109/ICTC49870.2020.9289160
James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). *An introduction to statistical learning* (Vol. 112). Springer.
Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer.
Jung, J., Lee, S., Hong, J., Youn, E., & Lee, G. (2020). Voice+ tactile: Augmenting in-vehicle voice user interface with tactile touchpad interaction. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1–12).
Kaiser, H. F. (1974). An index of factorial simplicity. *psychometrika*, *39*(1), 31–36.
Karana, E., Hekkert, P., & Kandachar, P. (2010). A tool for meaning driven materials selection. *Materials & Design*, *31*(6), 2932–2941.

Kershaw, J. F. (2023). *Sae international's dictionary for automotive engineers*. SAE International.

Kim, Y., Reza, M., McGrenere, J., & Yoon, D. (2021). Designers characterize naturalness in voice user interfaces: their goals, practices, and challenges. In *Proceedings of the 2021 chi conference on human factors in computing systems* (pp. 1–13).

Klein, A. M., Hinderks, A., Schrepp, M., & Thomaschewski, J. (2020). Construction of ueq+ scales for voice quality: measuring user experience quality of voice interaction. In *Proceedings of mensch und computer 2020* (pp. 1–5).

Klein, A. M., Kölln, K., Deutschländer, J., & Rauschenberger, M. (2023). Design and evaluation of voice user interfaces: What should one consider? In *International conference on human-computer interaction* (pp. 167–190).

Krüger, S., Bosch, E., Ihme, K., & Oehl, M. (2021). In-vehicle frustration mitigation via voice-user interfaces–a simulator study. In *Hci international 2021-posters: 23rd hci international conference, hcii 2021, virtual event, july 24–29, 2021, proceedings, part iii 23* (pp. 241–248).

Lahoual, D., & Frejus, M. (2019). When users assist the voice assistants: From supervision to failure resolution. In *Extended abstracts of the 2019 chi conference on human factors in computing systems* (pp. 1–8).

Lazarus, R. S. (1991). *Emotion and adaptation* (Vol. 557). Oxford University Press.

Lee, O.-K. D., Ayyagari, R., Nasirian, F., & Ahmadian, M. (2021). Role of interaction quality and trust in use of ai-based voice-assistant systems. *Journal of Systems and Information Technology*, *23*(2), 154–170.

Liu, J., Wan, F., Zou, J., & Zhang, J. (2023). Exploring factors affecting people's willingness to use a voice-based in-car assistant in electric cars: An empirical study. *World Electric Vehicle Journal*, *14*(3), 73.

Mahmood, A., Fung, J. W., Won, I., & Huang, C.-M. (2022). Owning mistakes sincerely: Strategies for mitigating ai errors. In *Proceedings of the 2022 chi conference on human factors in computing systems* (pp. 1–11).

McLean, G., & Osei-Frimpong, K. (2019). Hey alexa… examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28–37.

Meurisch, C., Mihale-Wilson, C. A., Hawlitschek, A., Giger, F., Müller, F., Hinz, O., & Mühlhäuser, M. (2020). Exploring user expectations of proactive ai systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*(4), 1–22.

Mishra, A., Shukla, A., & Sharma, S. K. (2022). Psychological determinants of users' adoption and word-of-mouth recommendations of smart voice assistants. *International Journal of Information Management*, *67*, 102413.

Moore, R. J., An, S., & Ren, G.-J. (2023). The ibm natural conversation framework: a new paradigm for conversational ux design. *Human–Computer Interaction*, *38*(3-4), 168–193.

Murad, C., Candello, H., & Munteanu, C. (2023). What's the talk on vui guidelines? a meta-analysis of guidelines for voice user interface design. In *Proceedings of the 5th international conference on conversational user interfaces* (pp. 1–16).

Murali, P. K., Kaboli, M., & Dahiya, R. (2022). Intelligent in-vehicle interaction technologies. *Advanced Intelligent Systems*, *4*(2), 2100122.

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic books.

Norman, D. A. (1988). *The psychology of everyday things*. Basic books.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research*, *17*(4), 460–469.

Ortony, A., Clore, G. L., & Collins, A. (2022). *The cognitive structure of emotions*. Cambridge university press.

Park, C. W., & Lessig, V. P. (1981). Familiarity and its impact on consumer decision biases and heuristics. *Journal of consumer research*, *8*(2), 223–230.

Pfleging, B., & Schmidt, A. (2015). driving-related activities in the car: Defining driver activities for manual and automated driving. In *Workshop on experiencing autonomous vehicles: Crossing the boundaries between a drive and a ride at chi* (Vol. 15).

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169–190.

Poushneh, A. (2021). Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors. *Journal of Retailing and Consumer Services*, *58*, 102283.

Pu, J., Li, L., Huang, Q., Jiang, W., Luo, X., & Wang, J. (2023). Evaluation study of intelligent vehicle virtual personal assistant. In *International conference on human-computer interaction* (pp. 186–199).

Pyae, A., & Joelsson, T. N. (2018). Investigating the usability and user experiences of voice user interface: a case of google home smart speaker. In *Proceedings of the 20th international conference on human-computer interaction with mobile devices and services adjunct* (pp. 127–131).

Regan, M. A., Lee, J. D., & Young, K. (2008). *Driver distraction: Theory, effects, and mitigation*. CRC press.

Ringfort-Felner, R., Laschke, M., Neuhaus, R., Theofanou-Fülbier, D., & Hassenzahl, M. (2022). It can be more than just a subservient assistant. distinct roles for the design of intelligent personal assistants. In *Nordic human-computer interaction conference* (pp. 1–17).

Ringfort-Felner, R., Laschke, M., Sadeghian, S., & Hassenzahl, M. (2022). Kiro: A design fiction to explore social conversation with voice assistants. *Proceedings of the ACM on Human-Computer Interaction*, *6*(GROUP), 1–21.

Roseman, I. J. (2001, 05). A model of appraisal in the emotion system integrating theory, research, and applications. In *Appraisal processes in emotion: Theory, methods, research.* Oxford University Press. Retrieved from `https://doi.org/10.1093/oso/9780195130072.003.0004` doi: 10.1093/oso/9780195130072.003.0004

Saunders, M. N., Townsend, K., et al. (2018). Choosing participants. *Sage handbook of qualitative business and management research methods*, 480–494.

Scherer, K. (2001). *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.

Schifferstein, H. N., & Hekkert, P. (2011). *Product experience*. Elsevier.

Schmidt, M., Minker, W., & Werner, S. (2020). How users react to proactive voice assistant behavior while driving. In *Proceedings of the twelfth language resources and evaluation conference* (pp. 485–490).

Schmidt, M., Stier, D., Werner, S., & Minker, W. (2019). Exploration and assessment of proactive use cases for an in-car voice assistant. In *Konferenz elektronische sprachsignalverarbeitung* (pp. 148–155).

Shlens, J. (2014). A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*.

Shneiderman, B. (2000). Universal usability. *Communications of the ACM*, *43*(5), 84–91.

Sjöberg, L. (2003). Distal factors in risk perception. *Journal of risk research*, *6*(3), 187–211.

Smith, C. A., & Lazarus, R. S. (1990). Emotion and adaptation.

Stier, D., Munro, K., Heid, U., & Minker, W. (2020). Towards situation-adaptive in-vehicle voice output. In *Proceedings of the 2nd conference on conversational user interfaces* (pp. 1–7).

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2013). *Using multivariate statistics* (Vol. 6). pearson Boston, MA.

Tonetto, L. M., & Desmet, P. M. (2016). Why we love or hate our cars: A qualitative approach to the development of a quantitative user experience survey. *Applied ergonomics*, *56*, 68–74.

Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, *106*, 234–240.

Völkel, S. T., Schödel, R., Buschek, D., Stachl, C., Winterhalter, V., Bühner, M., & Hussmann, H. (2020). Developing a personality model for speech-based conversational agents using the psycholexical approach. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1–14).

Wang, T., Luo, L., & Zhang, X. (2021). Evaluation model of voice user interface. In *Advances in usability, user experience, wearable and assistive technology: Proceedings of the ahfe 2021 virtual conferences on usability and user experience, human factors and wearable technologies, human factors in virtual environments and game design, and human factors and assistive technology, july 25-29, 2021, usa* (pp. 894–899).

Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, *50*(3), 449–455.

Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). *Engineering psychology and human performance*. Routledge.

Wigdor, D., & Wixon, D. (2011). *Brave nui world: designing natural user interfaces for touch and gesture*. Elsevier.

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological review*, *114*(4), 843.

Young, K., & Regan, M. (2007). Driver distraction: A review of the literature. In I. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter, & J. Irwin (Eds.), *Distracted driving* (pp. 379–405). Sydney, NSW: Australasian College of Road Safety.

Zwakman, D. S., Pal, D., & Arpnikanondt, C. (2021). Usability evaluation of artificial intelligence-based voice assistants: The case of amazon alexa. *SN Computer Science*, *2*(1), 28.

Appendix

Appendices

A Survey on Voice Assistant Usage in Cars

Survey on Voice Assistant Usage in Cars 2025-01-03, 17:53

Survey on Voice Assistant Usage in Cars

Introduction:

Welcome, and thank you so much for participating in our survey! 😊
In the evolving landscape of automotive technology, the traditional physical buttons used to control various car features and settings are increasingly being complemented by in-vehicle voice assistants. This survey aims to gather your insights and opinions on the use of voice assistants (VUI) for operating features in your car. We are particularly interested in understanding your perspective on reducing or eliminating traditional physical controls in favour of voice-only interaction within the car interior.

Your valuable feedback will aid us in enhancing future voice assistant designs and user experiences. This survey is estimated to take approximately 5-10 minutes to complete.

Definition:
Vehicle features: all the special things a car has to make it safe, fun, comfortable, and easy to use

General Information

1. What is your age? *

https://forms.office.com/Pages/DesignPageV2.aspx?prevorigin...zBAmyg0x1y6v0tUOE41TEJQQVMzTFFIT1BWRFpYREdNV1RYVyQlQCN0PWcu Page 1 of 18

Survey on Voice Assistant Usage in Cars 2025-01-03, 17:53

2. What is your gender? *

- Woman
- Man
- Non-binary
- Prefer not to say

3. Where are you from? *

4. How familiar are you with using voice assistant (such as Siri, Google assistant,etc.)?

	5	4	3
Voice assistant	○	○	○

5. How often do you use voice assistants (such as Siri, Google Assistant, Alexa, etc.)?

	5	4	3
Voice assistant	○	○	○

https://forms.office.com/Pages/DesignPageV2.aspx?prevorigin...zBAmyg0x1y6v0tUOE41TEJQQVMzTFFIT1BWRFpYREdNV1RYVyQlQCN0PWcu Page 2 of 18

Survey on Voice Assistant Usage in Cars 2025-01-03, 17:53

6. How many years have you been driving? *
If you do not possess a driving license or if you are merely a passenger, please enter '0'.

7. How frequently do you use a car? *

- Daily
- Several times a week
- Once a week
- Rarely
- Never

https://forms.office.com/Pages/DesignPageV2.aspx?prevorigin...zBAmyg0x1y6v0tUOE41TEJQQVMzTFFIT1BWRFpYREdNV1RYVyQlQCN0PWcu Page 3 of 18

Survey on Voice Assistant Usage in Cars 2025-01-03, 17:53

Opinion on Vehicle's Features with physical buttons

8. Do you know the car features listed below that entail physical buttons (Directly related to driving activities)?

	5	4	3
Adaptive Cruise Control (ACC)	○	○	○
Speed Control	○	○	○
Parking Assist	○	○	○
Gear Shift	○	○	○
Driving mode switch	○	○	○
EPB (Electronic parking brake)	○	○	○
Side Mirror adjustment	○	○	○
Steering Wheel Adjustment	○	○	○
Windshield wipers and automatic cleaning	○	○	○
Exterior Light control	○	○	○
Circulation	○	○	○

https://forms.office.com/Pages/DesignPageV2.aspx?prevorigin...zBAmyg0x1y6v0tUOE41TEJQQVMzTFFIT1BWRFpYREdNV1RYVyQlQCN0PWcu Page 4 of 18

Control	○	○	○
Defogging and Defrostting control	○	○	○
Media control	○	○	○
Trunk/Front Trunk Switch	○	○	○
Climate system control	○	○	○
Glove Box Control	○	○	○
Sunroof Control	○	○	○
Power window switches	○	○	○
Exterior Door handles	○	○	○
Interior Door handles	○	○	○

9. How often do you use each of the following car interior functions with physical buttons that are directly related to driving activities?

	5	4	3
Adaptive Cruise Control (ACC)	○	○	○
Speed Control	○	○	○
Parking Assist	○	○	○

Parking Assist	○	○	○
Gear Shift	○	○	○
Driving mode switch	○	○	○
EPB (Electronic parking brake)	○	○	○
Side Mirror adjustment	○	○	○
Steering Wheel Adjustment	○	○	○
Windshield wipers and automatic cleaning	○	○	○
Exterior Light control	○	○	○
Circulation Control	○	○	○
Defogging and Defrostting control	○	○	○
Media control	○	○	○
Trunk/Front Trunk Switch	○	○	○
Climate system control	○	○	○
Glove Box Control	○	○	○
Sunroof Control	○	○	○

Power window switches	○	○	○
Exterior Door handles	○	○	○
Interior Door handles	○	○	○

Usage and Opinion on Voice Assistants in Cars

Including voice assistants that you have used in your car, such as mobile devices, car play.

10. **Did you use in-car voice assistant?** *

The in-car voice assistant here includes voice assistants that you have used in your car, such as mobile devices.

○ Yes

○ No

○ Not sure

11. **How often do you use the in-car Voice user interface** *

If you do not possess a driving license or if you are merely a passenger, please enter 'never'.

	5	4	3
In-car Voice user interface	○	○	○

12. **For which purposes do you use the voice assistant?** (Please select all that apply) *

- [ ] Navigation
- [ ] Making phone calls
- [ ] Sending/reading text messages
- [ ] Playing music or media
- [ ] Controlling car settings (e.g., temperature, seat adjustment)
- [ ]

13. (If you have used,) For In-car Voice User assistant

In this questionnaire, the numbers from low to high indicate an increasing level of agreement or understanding. For example, 1 indicates a complete lack of familiarity with voice interaction products or assistants, while 5 indicates a complete understanding of voice interaction assistants.

	1	2	3
Does this help me in attaining your goals and motives?	○	○	○
Is this pleasant or enjoyable?	○	○	○
Does this confirm your expectations?	○	○	○
Does this match with your social norms and standards?	○	○	○
How certain are you about this voice user interface	○	○	○

Perceptions and Preferences

In this questionnaire, the numbers from low to high indicate an increasing level of agreement or understanding. For example, 1 indicates a complete lack of familiarity with voice interaction products or assistants, while 5 indicates a complete understanding of voice interaction assistants.

14. Do you know that the following car features can be controlled using a voice assistant instead of physical buttons?

	1	2	3
Adaptive Cruise Control (ACC)	○	○	○
Speed Control	○	○	○
Parking Assist	○	○	○
Gear Shift	○	○	○
Driving mode switch	○	○	○
EPB (Electronic parking brake)	○	○	○
Side Mirror adjustment	○	○	○
Steering Wheel Adjustment	○	○	○
Windshield wipers and automatic cleaning	○	○	○

Exterior Light control	○	○	○
Circulation Control	○	○	○
Defogging and Defrostting control	○	○	○
Media control	○	○	○
Trunk/Front Trunk Switch	○	○	○
Climate system control	○	○	○
Glove Box Control	○	○	○
Sunroof Control	○	○	○
Power window switches	○	○	○
Exterior Door handles	○	○	○
Interior Door handles	○	○	○

15. Do you think voice assistants could replace the following car features that currently use physical buttons if the car maintain these physical buttons or interaction?

	1	2	3
Adaptive Cruise			

Adaptive Cruise Control (ACC)	◯	◯	◯
Speed Control	◯	◯	◯
Parking Assist	◯	◯	◯
Gear Shift	◯	◯	◯
Driving mode switch	◯	◯	◯
EPB (Electronic parking brake)	◯	◯	◯
Side Mirror adjustment	◯	◯	◯
Steering Wheel Adjustment	◯	◯	◯
Windshield wipers and automatic cleaning	◯	◯	◯
Exterior Light control	◯	◯	◯
Circulation Control	◯	◯	◯
Defogging and Defrostting control	◯	◯	◯
Media control	◯	◯	◯
Trunk/Front Trunk Switch	◯	◯	◯
Climate system control	◯	◯	◯

Glove Box Control	◯	◯	◯
Sunroof Control	◯	◯	◯
Power window switches	◯	◯	◯
Exterior Door handles	◯	◯	◯
Interior Door handles	◯	◯	◯

16. Do you think voice assistants could replace the following car features that currently use physical buttons if the car no longer has these physical buttons?

	1	2	3
Adaptive Cruise Control (ACC)	◯	◯	◯
Speed Control	◯	◯	◯
Parking Assist	◯	◯	◯
Gear Shift	◯	◯	◯
Driving mode switch	◯	◯	◯
EPB (Electronic parking brake)	◯	◯	◯
Side Mirror adjustment	◯	◯	◯

Steering Wheel Adjustment	◯	◯	◯
Windshield wipers and automatic cleaning	◯	◯	◯
Exterior Light control	◯	◯	◯
Circulation Control	◯	◯	◯
Defogging and Defrostting control	◯	◯	◯
Media control	◯	◯	◯
Trunk/Front Trunk Switch	◯	◯	◯
Climate system control	◯	◯	◯
Glove Box Control	◯	◯	◯
Sunroof Control	◯	◯	◯
Power window switches	◯	◯	◯
Exterior Door handles	◯	◯	◯
Interior Door handles	◯	◯	◯

17. **What are the main benefits you see in using voice assistants over physical buttons in cars?** (Please select all that apply) *

- ☐ Increased safety by reducing distractions
- ☐ More convenient
- ☐ Efficiency: Quick operation execution.
- ☐ Visual impairment aid
- ☐ Personalized settings
- ☐ Adaptability: Simplifies complex functions.
- ☐

18. **What are your main concerns or drawbacks regarding the use of voice assistants in cars?** (Please select all that apply) *

- [ ] Privacy issues
- [ ] Accuracy Issues: Voice misunderstands commands.
- [ ] Slow Response: Delays in executing commands.
- [ ] Operational Errors: Unwanted or accidental activations.
- [ ] Noise Issues: Struggles in loud environments.
- [ ] Compatibility Problems: Issues with other devices.
- [ ] Learning Curve: Time needed to learn usage.
- [ ] Limited Functions: Not all needs met.
- [ ] Other

19. Do you prefer future car to rely on voice assistants for controlling functions rather than physical button.

0	1	2	3	4	5	6	7	8	9	10

Not at all likely — Extremely likely

Survey on Voice Assistant Usage in Cars 2025-01-03, 17:53

Thank you!

Your insights will help shape the future of in-car technology, making driving a safer, more enjoyable, and convenient experience for everyone. Thank you in advance for your valuable input!

B Interview Questions for User Study

Introduction: When designing interview questions, the focus is on deeply understanding users' experiences, preferences, encountered issues, and suggestions for improvement regarding the use of voice assistants for in-car interactions. Below is a series of interview questions designed to comprehensively collect feedback on users' use of in-car VUI (including Voice assistant).

Beginning

1. Are you familiar with voice assistants? (The voice assistants referred to here are broadly defined.)
   (a) How often do you use the voice assistant? And why (Any reasons behind the frequency of usage)?
2. When you use a car, do you use a voice interaction assistant (car-mounted) or other voice assistants (such as those on mobile devices)? How often do you use voice assistants?
3. As Driver or Passenger (Not only as a driver but also as a passenger):
   (a) Under what circumstances do you usually use voice assistants? (For example: while driving, when parked, etc.)
      i. The reasons behind using voice assistant under these circumstances.
   (b) What was your first impression when you started using the in-car voice assistant?

User Experience and Preference

If they are familiar with products and know how to use them:

1. In your use of an in-vehicle voice interaction interface (like using voice assistant), which tasks do you frequently perform?
   (a) Why do you use voice user interface (voice assistant) to help you complete these tasks?
2. In your use of an in-vehicle voice interaction interface (like using voice assistant), does VUI perform up to your expectations? Do you think the experience with these features conforms to your idea of standards and norms?
   (a) Which features of the VUI do you think perform well and which do not meet expectations? And why?
   (b) Is this product [voice assistant] pleasant or enjoyable? (User experience)
3. What challenges or difficulties have you encountered while using the in-car voice assistant?

   (a) Describe an experience where you felt the use of a voice assistant was particularly successful (or unsuccessful).

4. If [a certain function] is used infrequently, do you think it's acceptable to remove the physical button corresponding to that function?
   (a) Ask for the reasons in terms of the choices made in the questionnaire.

If people properly know about the voice assistant but seldom use them:

1. Why did they seldom or never use these voice assistants inside the car?
2. Ask them the requirements of voice assistant:
   (a) When would you like to use voice user interface (for instance, voice assistant)? (For example: while driving, when parked, etc.)
   (b) If you are going to use an in-vehicle voice interaction interface (like using voice assistant), which tasks would you perform?
      i. Why do you choose voice user interface (voice assistant) to help you complete these tasks?
   (c) How do you think this voice interaction assistant should perform to fulfill your expectations?
      i. Which features of the VUI do you think could perform well and which could not meet expectations? And why?
   (d) What challenges or difficulties have you encountered while using the in-car voice assistant?
      i. Describe an experience where you felt the use of a voice assistant was particularly successful (or unsuccessful).

About Switching Physical Features into VUI Control

1. Compared to physical buttons or touchscreens, what do you see as the advantages and disadvantages of using Voice User Interface?
2. In which situations do you tend to prefer voice control over physical buttons? And why?
   (a) Why do you use voice user interface (voice assistant) to help you complete these tasks?
3. If (a certain function) is used infrequently, do you think it's acceptable to remove the physical button corresponding to that function?
   (a) Could you give me the reasons behind your choice?
4. Why did you choose to use voice assistant instead of [a certain feature] with physical buttons?
   (a) Ask for the reasons in terms of the choices made in the questionnaire.
5. In what ways do you think voice assistants could offer features or experiences that would make you prefer voice control over physical buttons?

About the Features of Voice Assistant or Virtual Assistant

1. **Response Time:** Have you experienced delays in the voice assistant's responses? How did that affect your experience?
2. **Accuracy in Task Execution:** How accurately does the voice assistant perform the tasks you request?
3. **Environmental Factors:** How well does the voice assistant function in different driving conditions (e.g., noise levels, driving speed)?
4. **Integration and Compatibility:** Have you faced any issues with the voice assistant integrating with other in-car systems or external devices (e.g., smartphones)?
5. **User Interface:** What improvements would you suggest for the voice assistant's user interface to make it more intuitive?
6. **Customization Options:** Are there any additional customization options you would like to have for the voice assistant?
7. **Feedback Mechanism:** Would you prefer having a feedback mechanism to report issues or provide suggestions directly through the voice assistant? How would you envision it working?

About Future Expectations

1. **Improvements or New Features:** What improvements or new features do you hope to see in future in-car VUI or voice assistant?
   (a) If voice assistants could offer a more personalized experience, such as adjusting their responses and suggestions based on your preferences and habits, how appealing would you find this feature? *(This question must be answered.)*
2. (Optional) **Role and Importance:** How do you see the role and importance of voice assistant technology evolving in future automobiles?

Interview Notes

During the interview, ensure not to lead the users but feel free to provide examples (e.g., question ”a”) to clarify the context. Note that question ”a” is compulsory for users to answer.

C Evaluation Scale for comparative study

Instructions:
Please respond to the following questions based on your interaction experience with the voice assistant interface. Use the following 5-point rating scale for your answers:

- 1 = Strongly Disagree
- 2 = Disagree
- 3 = Neutral
- 4 = Agree
- 5 = Strongly Agree

Dialogue Evaluation

Question	Rating Scale (1-5)
How do you rate the quality of this feedback?	

Integrated Usability and User Experience Evaluation Survey (Part 1)

Questions	Rating Scale (1-5)
I think that I would like to use this interface frequently.	
I found the interface unnecessarily complex.	
I thought the interface was easy to use.	
I think that I would need the support of a technical person to be able to use this interface.	
I found the various functions in the interface were well integrated.	
I thought there was too much inconsistency in this interface.	
I imagine that most people would learn to use this interface very quickly.	
I found the interface very cumbersome/awkward to use.	
I felt very confident using the interface.	
I needed to learn a lot of things before I could get going with this interface.	
I felt afraid when it was unable to work properly.	
I think I would concern and evaluate its safety when using.	

Integrated Usability and User Experience Evaluation Survey (Part 2)

Questions	**Rating Scale (1-5)**
Does this VUI help you in attaining your goals and motives?	
Is this whole process(event) pleasant or enjoyable?	
Does this interface confirm your expectations?	
Does this event match with your standards?	
Can I handle or change the event?	
How certain are you about the event?	
After using it, Do you prefer to rely on voice for controlling functions rather than physical buttons?	

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY