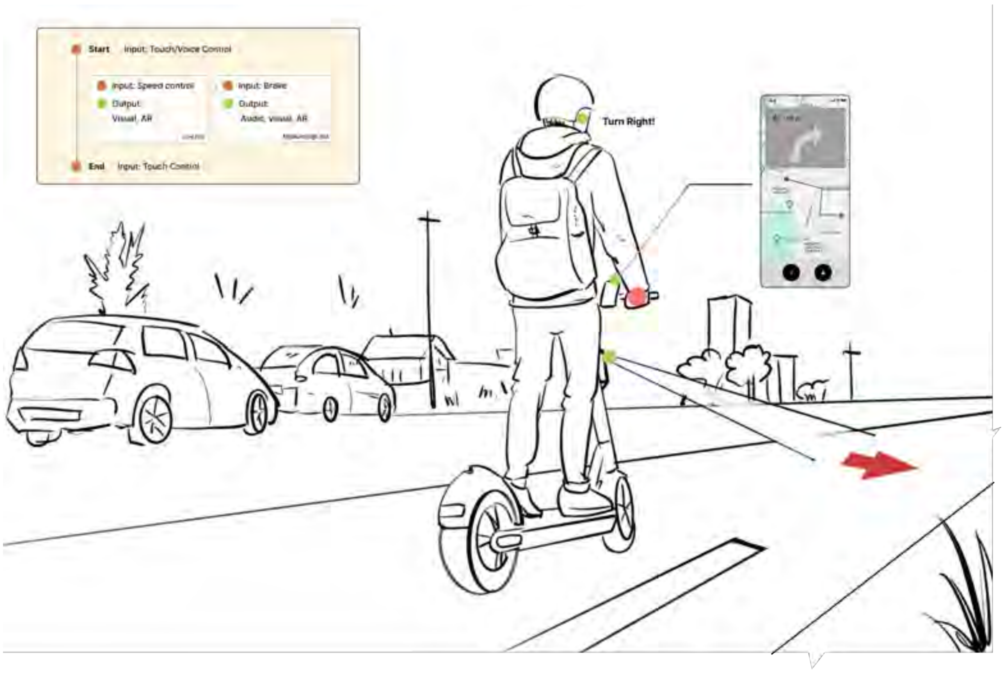
CHALMERS
UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG



Multimodal Approach to Enhance the Navigation System of the Shared E-scooter

Transforming E-scooter Navigation for Safer and Smoother Rides

Master's thesis in Computer science and engineering

Yihong Dai

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

Master's thesis 2025

Multimodal Approach to Enhance the Navigation System of the Shared E-scooter

Transforming E-scooter Navigation for Safer and Smoother Rides

Yihong Dai

UNIVERSITY OF
GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

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

Multimodal Approach to Enhance the Navigation System of the Shared E-scooter
Transforming E-scooter Navigation for Safer and Smoother Rides
Yihong Dai

Supervisor: Georgios Diapoulis, Department of Computer Science and Engineering, Chalmers University of Technology
Examiner: Morten Fjeld, Department of Computer Science and Engineering, Chalmers University of Technology

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

Cover: Schematic diagram of multimodal interaction in riding scenarios.

Typeset in LaTeX
Gothenburg, Sweden 2025

Multimodal Approach to Enhance the Navigation System of the Shared E-scooter
Transforming E-scooter Navigation for Safer and Smoother Rides
YIHONG DAI
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

Shared electric scooters are an important component of sustainable urban transportation, yet current navigation systems rely heavily on smartphone screens, introducing safety risks and usability limitations. This thesis addresses these issues by designing and evaluating a multimodal navigation interface that integrates ground projected augmented reality (AR) with auditory instructions and a visual user interface display (UI).

A user-centered design process guided the research. It started with qualitative methods (interviews, on-site observations) to identify user pain points and contextual needs, followed by quantitative analysis to deepen understanding of common issues and riding behaviors. Iterative design yielded concept sketches and wireframes, then a Unity-built VR prototype. Final usability testing evaluated task performance and cognitive load across interaction modalities, using cognitive efficiency as the key metric.

A controlled evaluation of the design prototype with 30 participants compared six configurations (single-modality, e.g., UI-only, to multimodal combinations). Results showed the AR + audio interaction + UI display configuration achieved the highest cognitive efficiency (E=0.84 ±1.20) and lowest mental workload (NASA-TLX = 33.35 ±16.81), with statistically significant improvements over all other configurations, while the UI-only system had the lowest efficiency (E=1.46 ±1.08) and highest cognitive load (statistical tests confirmed differences in performance, effort, and perceived frustration). Key findings from user testing included the following: (1) Channel redundancy may reduce cognitive efficiency, as users tend to rely primarily on one modality (e.g., augmented reality [AR] significantly outperformed user interface (UI) + AR, efficiency: $p = .001$); (2) Complementary coordination of modalities yields significant improvements (e.g., UI + AR + audio significantly outperformed UI + AR [efficiency: $p = .001$] and UI + audio [efficiency: $p = .002$]).

These findings highlight the potential of multimodal navigation systems: especially those that integrate projected AR - for improving safety, usability, and rider satisfaction in shared micromobility systems. However, this study has limitations: user testing was primarily conducted in a static, stable VR setup, while real-world road environments are far more complex, which would impose greater cognitive load on users.

Keywords: Multimodality in Interaction Design, User Experience, Shared E-scooter, Navigation Systems, Augmented Reality, Cognitive Efficiency

Acknowledgements

As I approach graduation, I want to express my heartfelt gratitude to everyone who has been part of my study-abroad journey. I still vividly remember the excitement and joy I felt two years ago when I received the admission email for the Chalmers scholarship. At that time, I was sitting in my universitys finance office, designing B2B interaction projects. From the end of the pandemic and the lockdown in Shanghai to my sudden resignation, I spent days watching videos of Nordic natural scenery on YouTube, dreaming of studying there someday. Lucky enough, that dream came true. When I first arrived, everything felt fresh, but academic pressures and environmental adjustments overwhelmed me for the first half year. Still, I learned to find peace in this new environment. I was fortunate to meet Professor Jasmina, whose guidance and support over these two years have been invaluable. I also want to thank Thommy, the academic director, for lending us VR equipment during our experiments, which was crucial to completing our research. To Gergeoisthough I often misspelled his namethank you for patiently reviewing our thesis progress every week. His dedication to our work often made me feel like I was pursuing a doctoral project, but it truly honed my academic rigor. Thank you, my supervisor! To my friends who walked alongside me for these two years: thank you for taking the time to participate in my thesis interviews and tests. To the strangers and classmates I met in the library and invited to join my study: your help was indispensable. Without you, I couldnt have recruited 30 participants for testing or gathered the data that shaped my research. I also want to thank my emotionally stable self, the self who persevered through challenges, and the self who has always loved music. Music has been my anchor: there were countless late nights when I wanted to give up, seething with frustration at seeing others carefree lives, unable to write a word. But after singing at home for two hours, Id find the calm to push forward with my thesis. Finally, I hope the future holds bright prospects for me.

In this thesis, I utilized large language models (LLMs) such as GPT-4o, DeepSeek, Qwen, and Kimi to assist with language output and expression. Given that the author is not native English speaker, these AI tools were occasionally employed to explain concepts. The use of AI has contributed to a more standardized and rigorous writing style, ensuring adherence to academic writing conventions while minimizing linguistic errors to the greatest extent possible.

Yihong Dai, Gothenburg, 2025-10-29

Contents

List of Figures

List of Tables

1

Introduction

With the rapid growth of the sharing economy and urban micromobility services [1], shared electric scooters have become an integral part of sustainable transportation globally (Figure 1.1) to solve [1]"the last 10 yards problem". This service not only promotes environmentally friendly travel, but also helps alleviate urban traffic congestion. However, despite its growing popularity, the interaction design for navigation systems used in e-scooters remains underdeveloped and does not address critical safety concerns [1] and User Experience concerns [2].

Figure 1.1: E-scooters are ubiquitous in Stockholm and Gothenburg (the figure shows the e-scooters of the VOI brand.) `https://www.itsinternational.com/its17/its8/news/vianova-boost-micromobility-stockholm`

1.1 Related Work

In light of the recent prevalence of the sharing economy and the surge in the use of e-scooters, the exploration of multimodal interaction in navigation systems has emerged as an area of active research over the past seven years (since 2017), with diverse applications and approaches being investigated.

---

[1] The last 10 yards problem in micromobility refers to the final moments of a user's journey, where the rider must transition from riding to reaching their exact destinationoften on foot or through confusing or constrained environments.

In 2017, during the conference "Defining Multimodal Interactions: One Size Does Not Fit All", Google addressed the application of multimodal interaction in its design. Starting from the perspectives of input and output, as well as users' capacity channels, Google put forward the following three guidelines:

- If one mode goes away, the other should take over
- Optimize for the strongest mode, but allow both
- Leverage the strength of each mode and avoid redundancy across modalities

In 2017, Lock and colleagues proposed a multimodal interaction approach for aiding blind individuals in navigation [3]. This method involved attaching sensors to the arms, hands, and body of the visually impaired and providing navigation assistance through auditory and haptic feedback. This multimodal interface was implemented on Googles Project Tango device developed using Android and Tango SDKs. While this approach offers significant convenience and inclusivity, it may prove cumbersome for everyday cycling scenarios. Additionally, there is the Assistive Sensor Solutions for Independent and Safe Travel (ASSIST) indoor navigation system in 2018 [4].

In 2019, The EyeBeacons system is a framework for multimodal wayfinding communication that makes use of wearable devices. This framework employs three distinct modalities: aural, tactile, and visual to convey navigation instructions [5]. It consists of three main components: a bone-conduction headset, a smartphone, and a smartwatch. The smartwatch is utilized to detect wayfinding messages presented as vibrations. However, the participants who tested the system reported that differentiating between vibrations and audio tunes was challenging.

In 2020, HapAR is a mobile augmented reality application designed to guide users within a university campus. Users can initiate the application by issuing a voice command to Siri [6]. Once the system receives the request, it processes it and attempts to locate the user's point of interest. When the user nears the destination or any area of interest, both aural and haptic feedback mechanisms are activated. However, user feedback indicated that the sound feedback was often drowned out by outdoor environmental noises, such as wind and the chatter of people.

In 2020, The Vibro-Audio map put forward by facilitating environmental learning, the development of cognitive maps, and wayfinding behavior. Vibro-Audio map made use of a low-cost, touchscreen-based multimodal interface on a commercial tablet. It served as an instance of a digital interactive map created using vibro-tactile and auditory information. The tablet device's built-in vibration motor was utilized to generate haptic (vibro-tactile) output. However, the results of this study were restricted solely to indoor building environments [7].

In 2022, Andrii Matviienko et al. conducted a comparison. Using situations with no warnings as a baseline at intersections with approaching vehicles, they compared three unimodal warnings for e-scooter riders [8]. The augmented reality warning presents a text "Warning! Detected Car getting close." (left), the auditory warning gives out beeping signals (middle), and the vibrotactile feedback is activated on the grips of the handlebar (right) (Figure 1.2). Their results indicate that AR and auditory warnings lead to shorter reaction times, have better perception, and create a better feeling of safety than vibro-tactile warnings. Moreover, auditory signals have a higher acceptance by the riders compared to the other two types of warnings.

Figure 1.2: Unimodal Warnings for E-Scooter Riders in Augmented Reality, adapted from 1.2

In 2022, Miao, Ning and colleagues conducted research focusing on the use of haptic interaction to enhance bike traffic safety. Their study aimed to increase cyclists' awareness of traffic information and improve their overall safety. To achieve this, they implemented user interaction methods involving head-mounted displays and wearable haptic displays for bicycle riders [9].

In 2024, Tabatabaie and colleagues proposed a Cross-modality E-scooter Naturalistic Riding Understanding System, namely CENRUS, from a human-centered AI perspective (Figure 1.3).

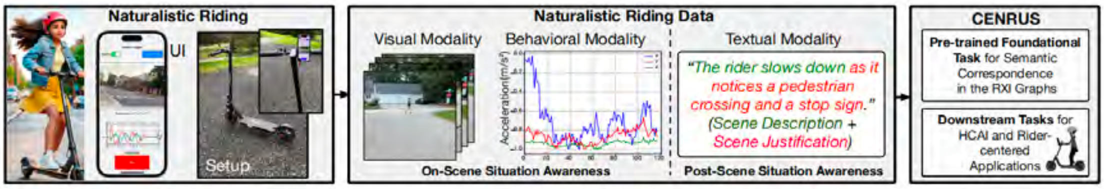


Figure 1.3: Naturalistic riding setup including user interface (UI) and e-scooters, data, and the pre-trained foundation task, adapted from [10]

In 2024, Yucheng and colleagues presented an embodied navigation with multi-modal information with three main components: *Perception and Memorization, Multi-modal Information Understanding, and Action Decision-making* (Figure 1.4).

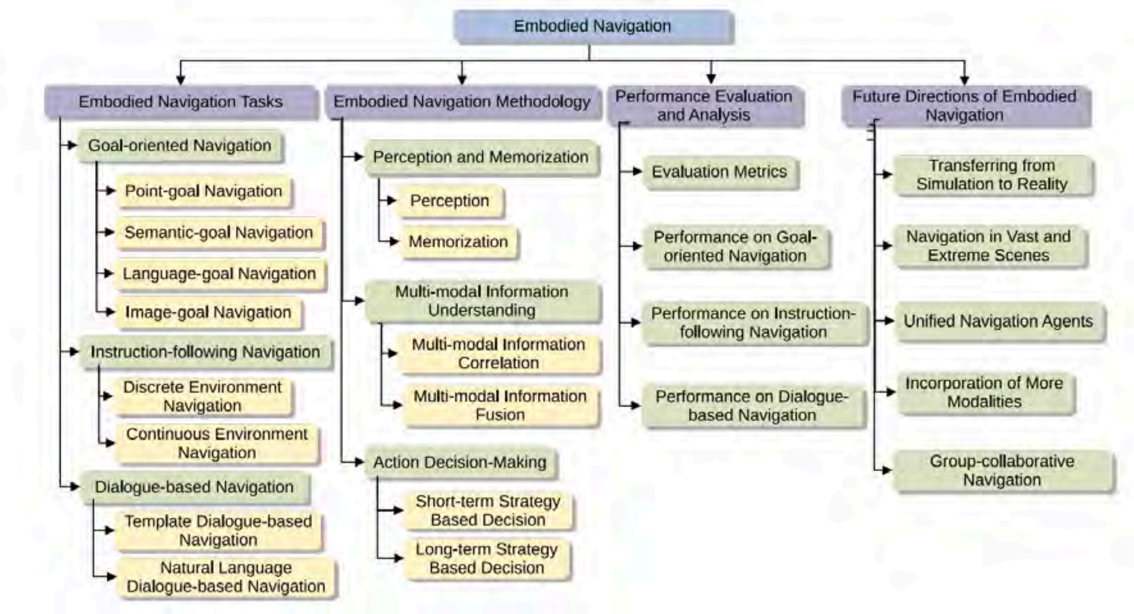


Figure 1.4: Framework of embodied navigation research, adapted from [11]

In 2025, Kazemzadeh, K. conducted an online video experiment and polled 920 e-scooter users in Sweden [12] and investigated the perception of safety which focus more on the riding habits and road situations, ignoring the scooter itself and human interaction. What's more, shared e-scooter brands in Sweden such as Voi, Ryde, and Lime primarily focus on addressing the placement of smartphones. However, during the riding process, users often rely on third-party applications like Google Maps for navigation.

It is evident that the safety issues of shared e-scooters have consistently drawn significant attention from researchers, with numerous strategies proposed to address them. However, when it comes to interaction methods and the overall riding experience, neither companies nor researchers have made substantial breakthroughs or proposed effective solutions. Therefore, this study holds promising theoretical and practical value from the multimodality aspect.

1.2 Purpose

Multimodal interfaces are generally intended to deliver natural and efficient interaction [13]. This thesis aims to use a multimodal interaction approach to design and evaluate the navigation system tailored to the unique challenges of shared E-scooters. By integrating different modality such as visual, auditory, and haptic feedback, we aim to develop a more intuitive, safer, and efficient interaction design. The results of this research can contribute to the fields of human-computer interaction and urban transportation by offering a novel approach to improving micromobility navigation systems.

- Designing a multimodal interface that combines visual, auditory, and haptic feedback for seamless navigation.

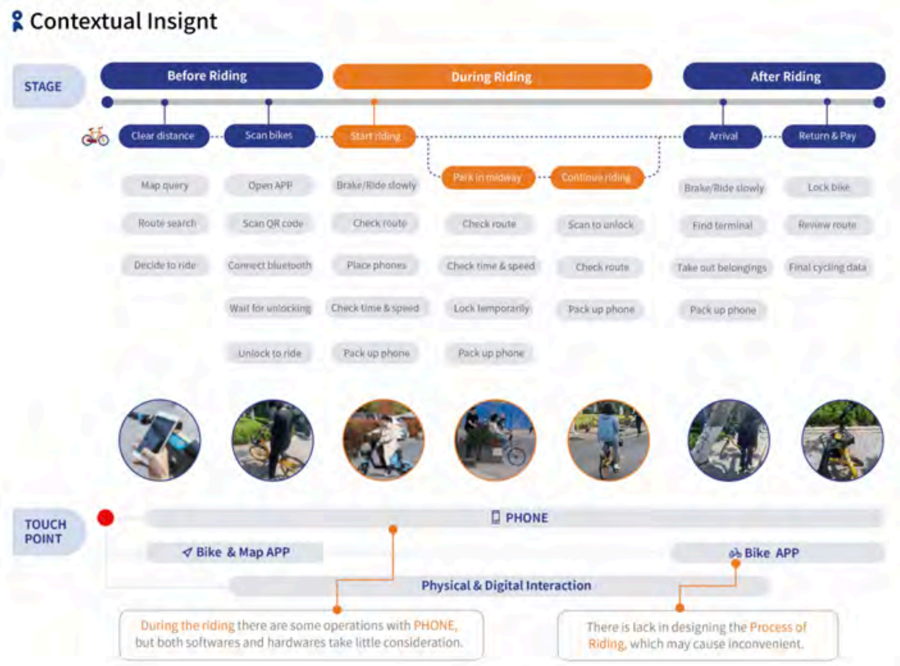
- Ensuring the system is compatible with existing e-scooter hardware, providing reliable data for the e-scooter providers.
- Conducting user-centered evaluations to assess system performance and usability.

1.3 Problem Statement

Currently, riders often rely on their smartphones as primary navigation tools, mounting them on handlebars, or holding them while riding, which creates significant distractions and unsafety on the road (Figure1.5). As for usability concerns, most users use navigation provided by Google Maps or other third-party software. Usually, voice navigation or interface-based navigation is used as the interaction method. For voice navigation, when used on noisy or complex roads, users may not be able to hear the navigation clearly due to background noise. As for interface-based navigation, when used in open outdoor environments, users are prone to being distracted, and the interface may be over-exposed, making it difficult to identify the information. In summary, the interaction methods provided by a single channel have usability deficiencies.

Figure 1.5: Rider holds the smartphone to navigate when riding which may trigger accidences

Through literature research and on-site observation(Figure1.6), the existing interaction problems during the riding can be classified into two major categories: safety issues and experience issues.



During the riding there are some operations with PHONE, but both softwares and hardwares take little consideration.

There is lack in designing the Process of Riding, which may cause inconvenient.

Figure 1.6: Riding User Journey Analysis

1.3.1 Safety Concerns

- **Distraction:** According to Badeau et al. [14], the majority type of injuries (44 %) occurred on sidewalks, lacking of concentration on their way. One major reason from our observation is that riders must visually interact with their smartphones for navigation, diverting their focus from the road.

- **Single Modality Limitations:** Although audio navigation is an option, it can isolate the user from ambient traffic sounds, compromising safety.

1.3.2 User Experience Concerns

- **Lack of Interaction from User Input:** Existing navigation systems primarily rely on one-way communication, delivering information to the user without considering their real-time input or context. For example, changes in the riders speed, environment, or road conditions are not accounted for, making the system static and less adaptive to dynamic situations.

- **Overload of Information:** Multimodal integration also referred to as the fusion engine, which is the key technical challenge for multimodal interaction systems [13]. Visual and auditory instructions often overwhelm users by providing too much information at once, leading to cognitive overload. Riders may struggle to process excessive or irrelevant details during their trips, detracting from the overall usability of the navigation system.

1.4 Research Objectives and Questions

Through preliminary research, we define several research questions to discuss.

RQ1: In what ways does multimodal interaction enhance both user safety and user experience in navigation systems for shared mobility?
Underlying position: Multimodal interaction in shared mobility navigation systems enhances user safety and user experience through multiple pathways. However, the realization of these enhancements is constrained by the inherent limitations of multimodal systems in recognizing the number and type of input modes [15]. Despite users generally showing a preference for multimodal over unimodal interaction, the specific design of a multimodal interface may not enable them to issue every command multimodally.

RQ2: Which modality is dominant when riding and in what contexts? What modality synergies can we construct from observation studies, and what prototypes can we develop to facilitate multimodal interactions?
Underlying position: In unfamiliar and unpredictable scenarios, users may rely on visual and interface interaction to determine their location. When users are interacting with roads that are predictable, they may tend to use audio interaction, as it is often more convenient and natural.

RQ3: Would the number of navigation output modalities during riding affect users' safety and user experience?
Underlying position: There may be a certain learning curve for more than one input modality at the beginning, and it may also impose a certain cognitive load on users in complex scenarios. However, once users have learned it, the interaction will be more convenient and faster. After riding a certain number of times, users may gradually adapt and start to feel the increased efficiency.

1.5 Ethical Considerations

The proposed multimodal navigation system aims to be fair, inclusive, sustainable, and ethical by prioritizing the safety and usability of shared e-scooter navigation. By integrating visual, auditory, and haptic feedback, the design ensures accessibility for a diverse range of users, including those with varying abilities or sensory preferences. The project adopts a user-centered approach, ensuring that feedback from real-world users informs the system's development, which promotes inclusivity and relevance.

From a sustainability perspective, the system utilizes existing phone holders, minimizing resource waste and supporting eco-friendly urban mobility trends. However, as societal values around safety and convenience evolve, the design must continuously adapt to new mobility practices. For instance, balancing privacy with personalized navigation could pose challenges. Overall, the system strengthens urban micromobility by enhancing safety and user experience, fostering trust and engagement in sustainable transportation options.

2

Background

2.1 Micromobility and E-scooter

In recent years, micromobility has emerged as a significant trend in urban transportation worldwide (Figure 2.1), revolutionizing the way people move within cities [16]. This shift is driven by a combination of factors, with e-scooters and shared bikes being a prominent and rapidly evolving component of the micromobility landscape. According to the International Transport Forum (2020) [17], micromobility can be defined as "Personal transportation that makes use of devices and vehicles with a weight of up to 350 kg. In the case of powered ones, their power supply is gradually decreased and completely cut off once a certain speed limit, which does not exceed 45 km/h, is reached."

Figure 2.1: Overview of the free-floating e-scooter offer in the world during 2019 (adapted from 6T-Research office, 2019)

Traditional modes of transportation, such as cars and buses, often contribute to traffic jams, air pollution, and longer commute times. Micromobility solutions offer solutions for first and last-mile connectivity, as well as assisting in the reduction of traffic [18]. For example, in cities like Shanghai and Beijing, where traffic congestion has become a major concern, e-scooters and shared bikes provide a convenient way for commuters to cover short-to-medium-distance trips, avoiding the gridlock on major roads.

With increasing awareness of environmental issues, there is a global push towards sustainable transportation. E-scooters and bikes are an eco-friendly alternative to gasoline-powered vehicles. They produce zero tailpipe emissions during operation, helping to improve air quality in urban centers. This aligns with the larger goals of many cities to reduce their carbon footprint and combat climate change.

Today's consumers, especially the younger generation, value convenience, flexibility, and shared experiences. Micromobility services, often provided through mobile apps, offer seamless access to e-scooters. Users can locate, unlock, and ride an e-scooter with just a few taps on their smartphones. This on-demand nature of the service fits well with the modern lifestyle. Moreover, the sharing economy model behind e-scooter services appeals to consumers who prefer not to own a vehicle but still want the freedom to move around easily.

In response to the growth of micromobility, many cities and regions are adjusting their transportation policies and regulations. In most European countries (such as the Czech Republic, all Scandinavian countries, Poland and Italy), e-scooters are legal and are categorized as bikes. According to the Swedish Transportation Agency [19], in Sweden, an e-scooter can be regarded as a bike provided that it does not surpass a speed of 25 km/h and has a maximum engine power of 250 Watts. However, in Germany, Austria, Spain, Belgium and France, e-scooters are classified in a separate category. For example, in Germany, electric scooters are a new category of vehicles that is similar to light scooters and bicycles [20]. Some areas have introduced dedicated bike and scooter lanes to ensure the safety of micromobility users. Others are implementing rules on helmet use, parking, and speed limits for e-scooters. These regulatory changes are intended to balance the benefits of e-scooter use with the need to maintain traffic safety and order.

Despite rapid growth and potential benefits, e-scooters also face challenges. During the period April to August 2019, 14% of the people examined in Stockholm hospitals had been injured in traffic accidents involving electronic scooters [21]. Among these accidents, Andreas et al. found that most e-scooters are not associated with an increase in pedestrian or bicycle accidents, but are predominantly involved in accidents with cars [16]. Issues such as sidewalk clutter due to haphazard parking, safety concerns related to rider behavior and vehicle-pedestrian collisions, and the need for proper maintenance of the scooters are areas that require further attention. However, as the micromobility industry continues to evolve, these challenges are likely to be addressed through technological innovation, better user education, and more refined regulations.

2.2 Multimodal Interaction

2.2.1 Human Behavior and User Experience

Human interaction with the world is inherently multimodal [22]. For instance, when we are in a conversation, we do not just rely on the spoken words (auditory modality). We also pay attention to the facial expressions (visual modality) and body language(a visual-kinesthetic modality) of the speaker, which can convey emotions such as happiness, sadness, or surprise.

The advent of multimodal interfaces based on recognition of human speech, gazes, gestures, and other natural behaviors represents only the beginning of a progression toward computational interfaces capable of relatively human-like sensory perception (Figure 2.2) [15].

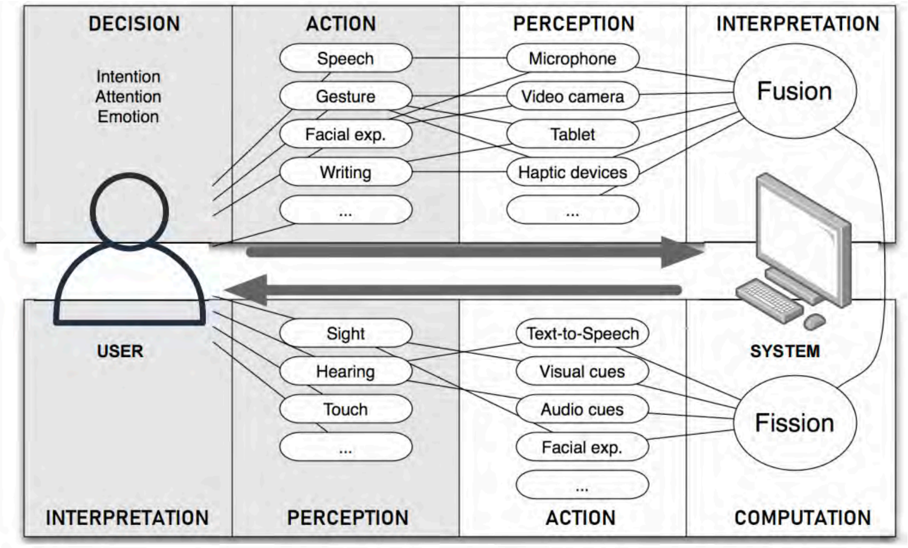


Figure 2.2: Human decision and relation with modality interaction (adapted from [23])

In the past, human-computer interaction mainly focused on unimodal interactions, like pure voice-dialogue systems, graphical user interface (GUI) interactions with mouse or keyboard operations, and touch-based interactions on touch-screen devices. The advantage is that with technological maturation, norms were set for these interactions, and users developed consistent mental models. For example, people now instinctively know that double-clicking an icon opens a program or swiping on a touch-screen performs certain actions, which eases the learning and adoption of new technologies.

Although unimodal interaction is simple and specialized, with the increasing complexity of life scenarios and the widespread popularity of ubiquitous mobile computing, a single interaction modality can no longer meet people's needs. Multimodal interfaces, which coordinate the processing of combined natural input modes such as speech, touch, hand gestures, eye gaze, and head and body movements, represent a paradigm shift at research level away from conventional windows-icons-menus-pointers (WIMP) interfaces to provide users with greater expressive power, naturalness, flexibility, and portability [24].

In 2000, Oviatt et al. [25] pointed out eleven advantages of multimodal interaction:

- They permit the flexible use of input modes, including alternation and integrated use.
- They support improved efficiency, especially when manipulating graphical information.
- They can support shorter and simpler speech utterances than a speech-only interface, which results in fewer disfluencies and more robust speech recognition.
- They can support greater precision of spatial information than a speech-only interface since pen input can be quite precise.
- They give users alternatives in their interaction techniques.
- They lead to enhanced error avoidance and ease of error resolution.
- They accommodate a wider range of users, tasks, and environmental situations.
- They are adaptable during continuously changing environmental conditions.
- They accommodate individual differences, such as permanent or temporary handicaps.
- They can help prevent the overuse of any individual mode during extended computer usage.

Multimodal interaction ought to be personalized, enhancing the user experience by making it more unique. For instance, Siri (Figure 2.3) has the ability to recognize the distinct voice of the phone's owner and then engage in interaction with the user.

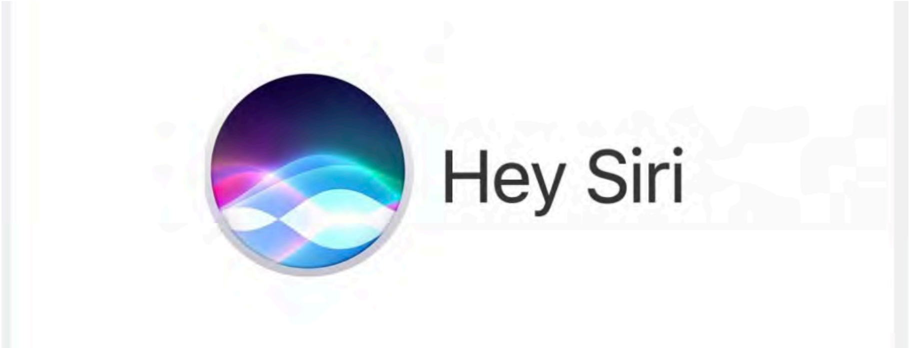


Figure 2.3: Voice mode of Apple's Agent: Siri experience

2.2.2 Factors in Multimodality

In 2017, during the conference "Defining Multimodal Interactions: One Size Does Not Fit All", Google addressed four main factors that influence multimodality:

- **Motion:** Specifically, when compared to bicycle riding, e-scooter riding was associated with more severe vibration events, irrespective of the types of pavement [26].

- **Environment:** In noisy settings, voice commands and hand gestures work together to clarify user intent as sound can be hard to hear. In hands-free or immersive AR, VR and MR environments, gaze-based input combines with touch or speech. Different environments determine the best input modality combinations, showing that the environment significantly influences how users interact with systems and how systems are designed to fit various scenarios.

- **Proximity:** In multimodal interactions, the physical distance between the user and the interface plays a crucial role. When using an e-scooter, the users proximity to the display, handlebars, and other interaction points determines which input methods are most effective. For instance, while touchscreens might be viable when stationary, voice commands or haptic feedback become more practical during motion. Similarly, proximity influences the responsiveness of location-based features, such as navigation prompts that adjust based on the user's real-time position and movement.

- **Human Capacity:** Factors such as cognitive load, attention span, and motor skills affect how users process and respond to multiple interaction modes simultaneously. For e-scooter riders, balancing navigation assistance with situational awareness is essential: overloading them with complex auditory or visual cues can lead to distractions and potential safety risks. An effective multimodal design optimizes interactions to align with human capabilities, ensuring that users can seamlessly interpret and act upon the provided information while maintaining focus on the riding experience.

What's more, modality fusion is also a crucial topic in designing a better and more seamless interaction. Precisely coordinating the integration of diverse input modalities is essential to guarantee that they enhance one another instead of creating conflicts. This process encompasses intricate input synchronization, modality fusion, and context awareness. All these aspects necessitate sophisticated algorithms and machine-learning techniques [27].

Fusion can take place at various levels, as depicted in Figure 2.4, contingent upon the system's requirements. Early-stage or sensor-level fusion entails merging the raw data originating from different modalities during the initial processing phase [28]. One key element of modality fusion is dealing with conflicts that arise between different inputs. In situations where modalities offer conflicting information, for example, when a gesture indicates one object but speech refers to a different one, the system has to figure out which input should be given precedence [29].

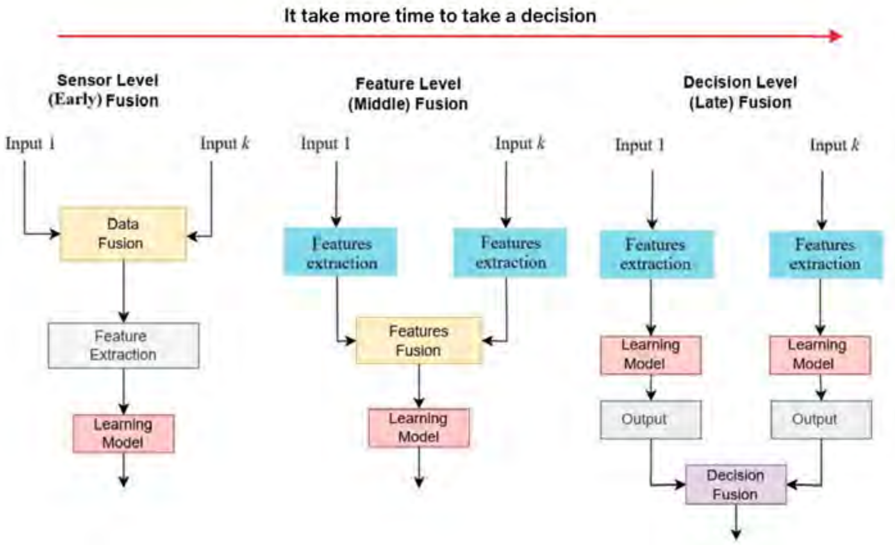


Figure 2.4: An overview of different levels of modalities fusion (adapted from [28])

2.2.3 Technology Development

Advancements in sensor technology and computing technologies have been instrumental in the growth of multimodality interaction in various applications, including virtual assistants, intelligent environments, healthcare, and accessibility technologies [27]. High-resolution cameras, microphones, and touch-sensitive surfaces are now commonly integrated into devices, enabling accurate detection of user input across different modalities. For example, modern smartphones are equipped with front-facing cameras that can recognize facial expressions and gestures, adding a new dimension to user interaction.

As vision-based technologies mature, one important direction is the development of blended multimodal interfaces that combine both passive and active modes [15]. Moreover, AI makes contributions to multimodal systems through its capacity to adapt dynamically to both user behavior and environmental contexts [30].

The Internet of Things (IoT) and augmented reality (AR) are widely popular technologies. They enable interaction systems to integrate the real-world context of the user (agent) with immersive AR content [31]. AR functions by visualizing graphical objects within the view of the real world. Usually, the view is presented through a real-time camera feed on a device or a head-mounted display (HMD) [32].

As multimodal interfaces gradually evolve toward supporting more advanced recognition of users' natural activities in context, they will expand beyond rudimentary bimodal systems to ones that incorporate three or more input modes, qualitatively different modes, and more sophisticated models of multimodal interaction [15]. This trend already has been initiated within biometrics research, which has combined recognition of multiple behavioral input modes [33].

Furthermore, more scientists are exploring emerging sensory technologies like brain-computer interfaces (BCIs) [34].

In summary, multimodality interaction has now become a fundamental aspect of modern technology design, with the potential to create more inclusive, intuitive, and user-friendly experiences for everyone. As technology continues to evolve, the field of multimodality interaction is likely to expand further, incorporating new modalities and applications.

2.3 Navigation Systems

The cognitive process of "navigation" is composed of two main elements: locomotion and wayfinding. Locomotion refers to the physical movement of a person within their immediate environment. Wayfinding, on the other hand, is the process of devising a plan to reach a specific destination [35].

2.3.1 Unimodal Navigation Systems

Typical unimodal navigation systems can be classified into four distinct types by interaction [11]:

- **Point-goal navigation:** This type of navigation task was initially proposed by Anderson et al. [36]. In point-goal navigation, the user simply drags and clicks on a point within a map to navigate to a specific location.
- **Semantic-goal navigation:** Here, the navigation goal is described in semantic terms. For example, a user might be given the task 'find a restaurant'. This requires the system to understand the semantic concept of a restaurant and guide the user to a relevant location.
- **Language-goal navigation:** In language-goal navigation, the destination is defined by a short sentence. A common example is "GO TO X", where "X" usually refers to an object or a room. The system must interpret the natural language instruction to determine the correct navigation path.
- **Image-goal navigation:** In 2017, Zhu et al. designed the first image-goal navigation task and proposed a framework for it. In this type of navigation, an environment with high-quality 3D scenes and a physics engine is presented for embodied AI agents. The task is to find a target object that is represented by a previously shown image. The agent must analyze the visual features of the image and search for the corresponding object in the 3D environment [37].

2.3.2 Multimodal Navigation Systems

Vainio stated that multimodal navigation systems offer users the flexibility to provide inputs or receive outputs in their preferred modality [38]. For instance, in a noisy setting, when receiving directions, vibratory feedback can prove more effective than aural feedback. On the other hand, if the user desires to obtain more details about the navigation environment, like landmarks and traffic signs, audio might be a more appropriate option.

Visual maps offer several benefits as a navigation tool. They can provide an overall view of an environment and have a high information density. In recent decades, many promising technologies have come to the fore, aiming to replace visual maps. These include point and sweep gestures, spatial sound, tactile information, and other multimodal alternatives.

Tactile maps enable users to access geographical representations. While these maps are valuable for acquiring spatial knowledge, they also have certain drawbacks. For example, they often require the user to be able to read braille, which limits their accessibility to those who are proficient in this reading system [39].

Baus et al. [40] presented a hybrid navigation system that relies on visual and haptic modalities to determine the users location and that adapts the presentation of route directions to the limited technical resources of the output device and the limited cognitive resources of the user (Figure 2.5).

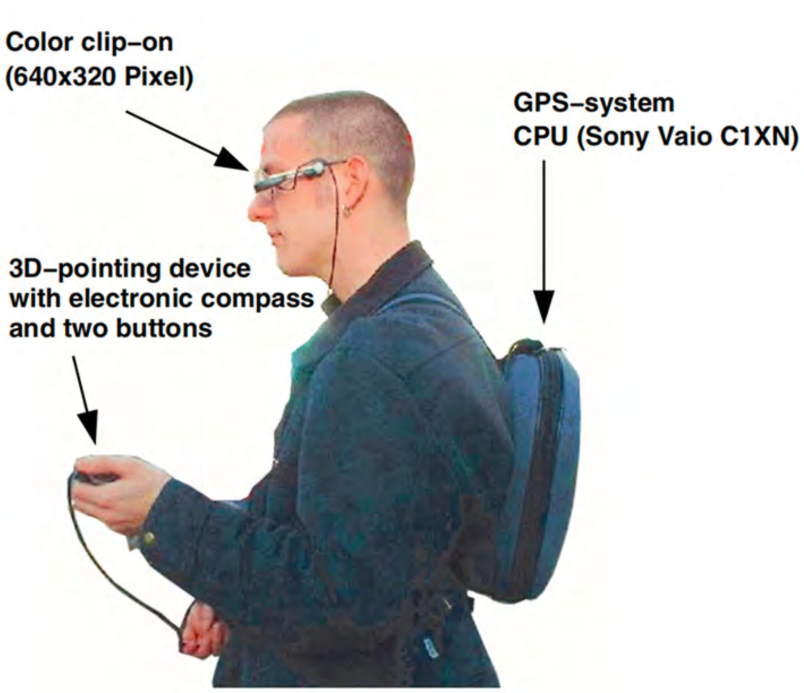


Figure 2.5: The system architecture of the hybrid navigation aid, adapted from [40]

2.4 Stakeholders

2.4.1 E-scooter Riders

They are the primary end-users of the e-scooter navigation system. Their safety, convenience, and overall riding experience are directly affected by the navigation system. They expect a navigation system that is easy to use, provides accurate directions, and does not cause distractions while riding. For example, a commuting rider in a busy urban area needs a navigation system that can quickly adapt to traffic changes and guide them to their destination safely.

2.4.2 Pedestrians

According to Matús Sucha et al., approximately 20-30% of e-scooter riders ride on sidewalks, regardless of whether it is permitted. Those who do not ride e-scooters are inclined to claim that riding an e-scooter is quite dangerous. With the exception of Australians, pedestrians view e-scooter riders and the operation of e-scooters as irritating [41].

2.4.3 E-scooter Providers

Companies like Lime, Bird, and Voi are major stakeholders. They are interested in providing a high-quality service to their customers. A well-designed navigation system can enhance customer satisfaction, leading to increased ridership and loyalty. Moreover, they need a system that is compatible with their existing e-scooter hardware and can provide data on user behavior, such as popular routes and riding patterns. This data can help them optimize their fleet deployment and service offerings.

2.4.4 Urban Planners

They are concerned with the integration of e-scooters into the existing urban transportation infrastructure. A good navigation system can influence the way e-scooters are used in the city, potentially reducing traffic congestion and improving the overall flow of urban mobility. For instance, if the navigation system can direct e-scooter riders to less congested routes, it can help in better traffic management. Urban planners also need to ensure that e-scooter usage aligns with the development of bike lanes, sidewalks, and other transportation-related urban features.

2.4.5 Other stakeholders

Policymakers, who set regulations governing e-scooter operations; safety advocacy groups, which push for safety-enhancing features in the navigation system; and technology developers, who are responsible for innovating and improving the multimodal navigation technology, all play crucial roles as other significant stakeholders in the e-scooter navigation ecosystem.

2.5 Existing Solution

2.5.1 Google Maps

- **Route Planning:** Google Maps offers comprehensive route planning for e-scooter riders. It takes into account factors such as distance, estimated travel time, and traffic conditions. For example, it can calculate the fastest route to a destination, considering both major roads and smaller, less-congested streets that are suitable for e-scooter travel. It also provides options for different types of routes, such as the shortest route or the route with the least amount of elevation change, which can be beneficial for e-scooter riders conserving battery power.

- **Visual Interface:** Google Maps has a highly visual interface. Riders can view their current location, the route to their destination, and nearby points of interest on a detailed map. The map is designed to be easy to read, with clear markers for turns, intersections, and destinations. However, in the context of e-scooter riding, the visual interface can be a distraction when riders need to look at their smartphones while on the move.
- **Voice Navigation:** It offers voice-based navigation, where turn-by-turn instructions are provided through synthesized speech. This allows riders to keep their eyes on the road while getting directions. However, as mentioned earlier, in noisy urban environments, voice instructions can be difficult to hear, especially when riding at higher speeds or in areas with a lot of background noise.

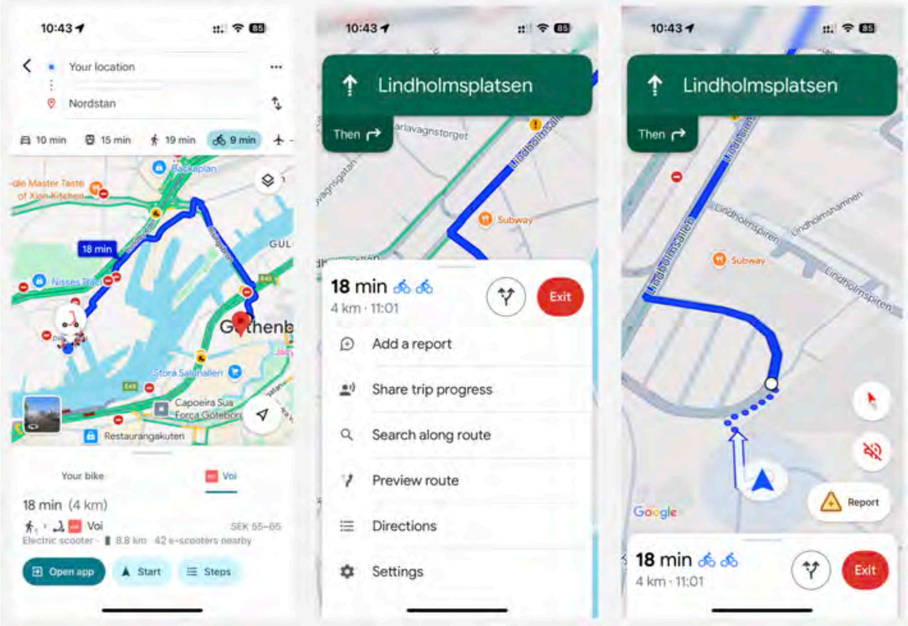


Figure 2.6: Google Maps: IOS version 18.0, captured on 2025-04-01

2.5.2 Ryde

- **Navigation Experience:** The navigation feature in the Ryde app is essential for helping users find their way while riding, but it presents some usability challenges. Currently, users rely on a map-based interface that requires them to glance at their phones for directions, which can be inconvenient and distracting while riding. Unlike standalone navigation apps like Google Maps, Ryde does not provide clear, continuous voice guidance, making it harder for users to follow directions without stopping. Additionally, the interface does not seem optimized for quick decision-making, as turn indications can sometimes be unclear, especially in complex urban areas with multiple intersecting roads. A more intuitive design with larger, more visible directions or a simplified navigation view could improve usability, making it easier for riders to follow their route safely without frequent screen interactions.

- **Platform Differences: IOS vs. Android:** The user experience of the Ryde app differs depending on whether it is used on iOS or Android devices, affecting how easily users can find and use the navigation feature. On iOS, the app follows Apples interface guidelines, making navigation tools more visually consistent and accessible. Users may find it easier to locate the map view and start a ride, as the layout appears more refined with smooth animations and predictable interactions. Additionally, Apple Pay integration simplifies the unlocking process, allowing for a quicker start to the ride. On Android, however, there are noticeable variations in performance and design consistency. Some users report that map rendering takes longer to load, and the navigation feature may be slightly harder to find due to different menu placements. The app may also behave inconsistently across different Android devices due to varying screen sizes and system optimizations. These differences can create an uneven experience, making it less convenient for some users to quickly access navigation and start their ride without delays. Improving consistency across both platforms, particularly in how navigation is presented and accessed, would enhance overall usability and ensure a more seamless experience for all users.

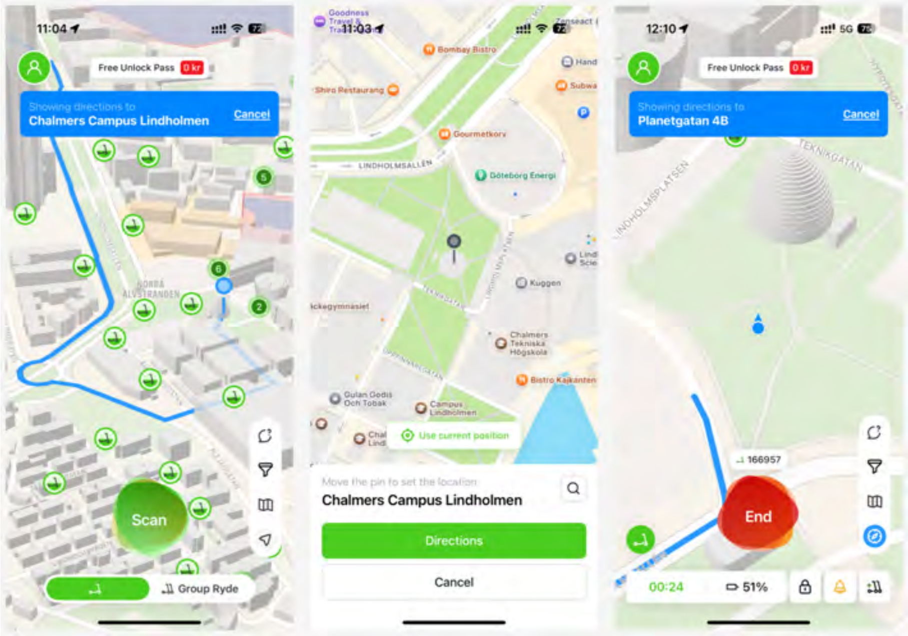


Figure 2.7: Ryde Application: IOS version 18.0, captured on 2025-04-01

2.5.3 VOI

Major developments have occurred in the hardware and software needed to support key component technologies incorporated within multimodal systems, as well as in techniques for integrating parallel input streams [15].

The latest Voi ride has added a new interface, but the current interaction mode is still a single-channel solution at the visual level (Figure 2.8).

Meet the Voi e-scooter

1. **Throttle to control your speed**
The throttle is for accelerating. Kick off with one foot and then accelerate to get going.

2. **Brakes to stop safely**
Voi e-scooters have levers on the handlebars for safe stopping. When you need to slow down or stop, squeeze both brake levers gently and at the same time.

3. **Indicators for turning**
Our latest models have indicators to sig other road users if you plan to turn or

Figure 2.8: Hardware of VOI 7, captured on 2025-04-01

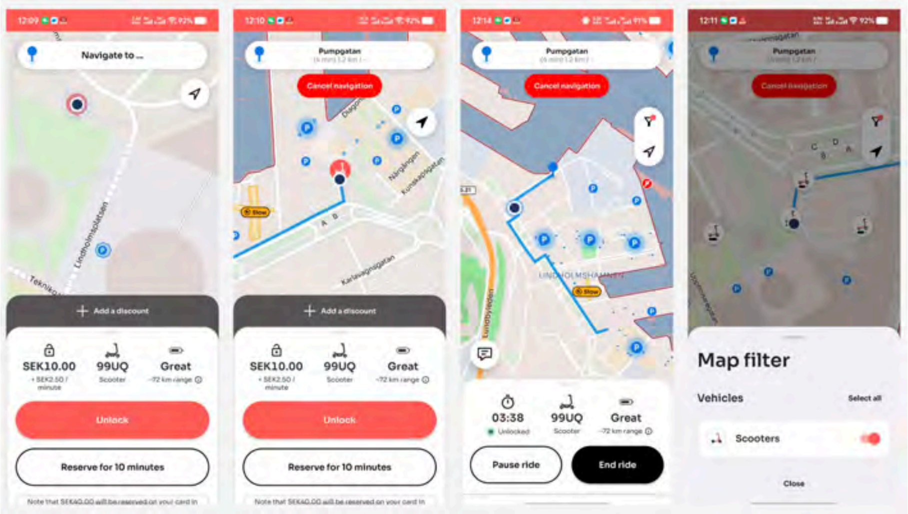


Figure 2.9: Hardware of VOI 7, captured on 2025-04-01

3

Theory

3.1 User-Centered Design Theory

User-centered design (UCD) is a design approach that places users at the core of the development process, ensuring that products meet their needs, preferences, and limitations [42]. The UCD framework emphasizes iterative design, usability testing, and feedback loops to refine and optimize interactive systems. In the context of shared e-scooter navigation, UCD ensures that multimodal interfaces are designed based on user requirements, considering factors such as cognitive load, safety, and ease of use [43].

3.2 Cognitive Load Theory

The cognitive science-related theories on intersensory perception and intermodal coordination provide a foundation of information for user modeling, as well as information on what systems must recognize and how multimodal architectures should be organized [15].

Cognitive Load Theory (CLT) centers on creating instructional strategies. These strategies aim to effectively utilize the restricted cognitive processing capabilities of people, allowing them to transfer the knowledge and skills they've learned to new and unfamiliar circumstances [44] [45].

CLT is founded on a cognitive architecture. This architecture comprises a working memory that has a limited capacity and partially independent processing units for visual and auditory information. Moreover, this working memory interacts with a long-term memory that has an essentially unlimited capacity [46].

3.3 Design Guidelines of Multimodal Interaction

Creating multimodal systems is challenging, as the typical design choices and intuitions from standard computing environments do not necessarily translate well to multimodal environments.

Oviatts "Ten Myths of Multimodal Interaction" [24] offers useful insights for those researching and building multimodal systems, with a few especially apropos:

> *Myth1: If you build a multimodal system, users will interact multimodally.*

Rather, users tend to intermix unimodal and multimodal interactions. Fortunately, multimodal interactions are often predictable based on the type of action being performed.

> *Myth2: Multimodal input involves simultaneous signals.*

Multimodal signals often do not co-occur temporally, and much of multimodal interaction involves the sequential (rather than simultaneous) use of modalities.

> *Myth3: Multimodal integration involves redundancy of content between modes.*

Complementarity of content may be more significant in multimodal systems than redundancy.

> *Myth4: Enhanced efficiency is the main advantage of multimodal systems.*

Multimodal systems may increase efficiency, but not always. Their main advantages may be found in other aspects, such as decreased errors, increased flexibility, or increased user satisfaction.

> *Myth5: Individual error-prone recognition technologies combine multimodally to produce even greater unreliability.*

In an appropriately flexible multimodal interface, people determine how to use the available input modes most effectively; mutual disambiguation of signals may contribute to a higher level of robustness.

Reeves et al. [47] defined the following guidelines for multimodal user interface design:

> *Multimodal systems should be designed for the broadest range of users and contexts of use.*

Designers should support the best modality or combination of modalities anticipated in changing environments (for example, private office vs. driving a car).

> *Designers should take care to address privacy and security issues in multimodal systems.*

For example, non-speech alternatives should be available in a public context to prevent others from overhearing information or conversations.

> *Maximize human cognitive and physical abilities, based on an understanding of users' human information processing abilities and limitations.*

For example, match the output to an acceptable user input style, such as constrained grammar or unconstrained natural language.

> *Multimodal interfaces should adapt to the needs and abilities of different users, as well as different contexts of use.*

Individual differences (for example, age, preferences, skill, sensory or motor impairment) can be captured in a user profile and used to determine interface settings.

> *Be consistentin system output, presentation and prompts, enabling shortcuts, state switching, etc.*

3.4 Related Concepts

3.4.1 Input and Output modalities

Humans primarily interact with the world through their five major senses sight, hearing, touch, smell, and taste. In perception, a mode or modality refers to receiving stimuli from a particular sense [13].

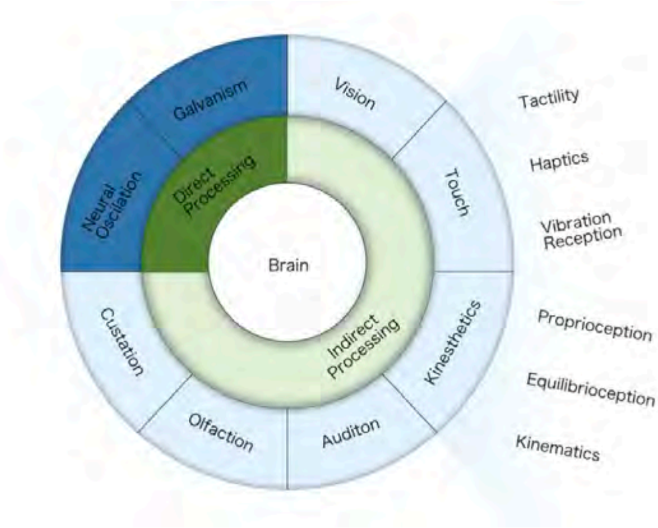


Figure 3.1: Taxonomy of interaction modalities proposed by Augstein and Neumayr [48]

3.4.2 Cognitive Efficiency

According to Pass et al., there are three main aspects to measure the cognitive load [49]:

- *Mental load* is the aspect of cognitive load that originates from the interaction between task and subject characteristics.
- *Mental effort* is an element of cognitive load. It pertains to the cognitive capacity that is actually assigned to meet the demands set by a task. As a result, it can be regarded as a reflection of the real-time cognitive load.
- *Performance*, which is another aspect of cognitive load, can be defined by a learner's accomplishments. These include metrics like the number of correct test items, the number of errors made, and the time spent on the task. Per-

formance can be evaluated either while individuals are engaged in a task or after they have completed it.

$$E = \frac{z_{\text{Performance}} - z_{\text{Mental Effort}}}{\sqrt{2}} \tag{3.1}$$

An instructional condition efficiency score (E) is calculated. This score is determined as the perpendicular distance between a dot-represented by the z-score for mental effort and the z-score for performance-and the diagonal line where E =0 (Figure 3.2). The formula used for this calculation is based on the general formula for finding the distance from a point p(x, y) to a line ax+by+c=0. In this case, the square root of 2 in the formula is derived from that general distance-calculation formula.

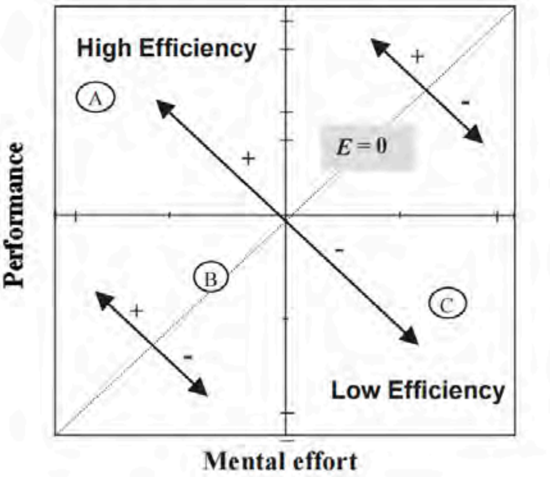
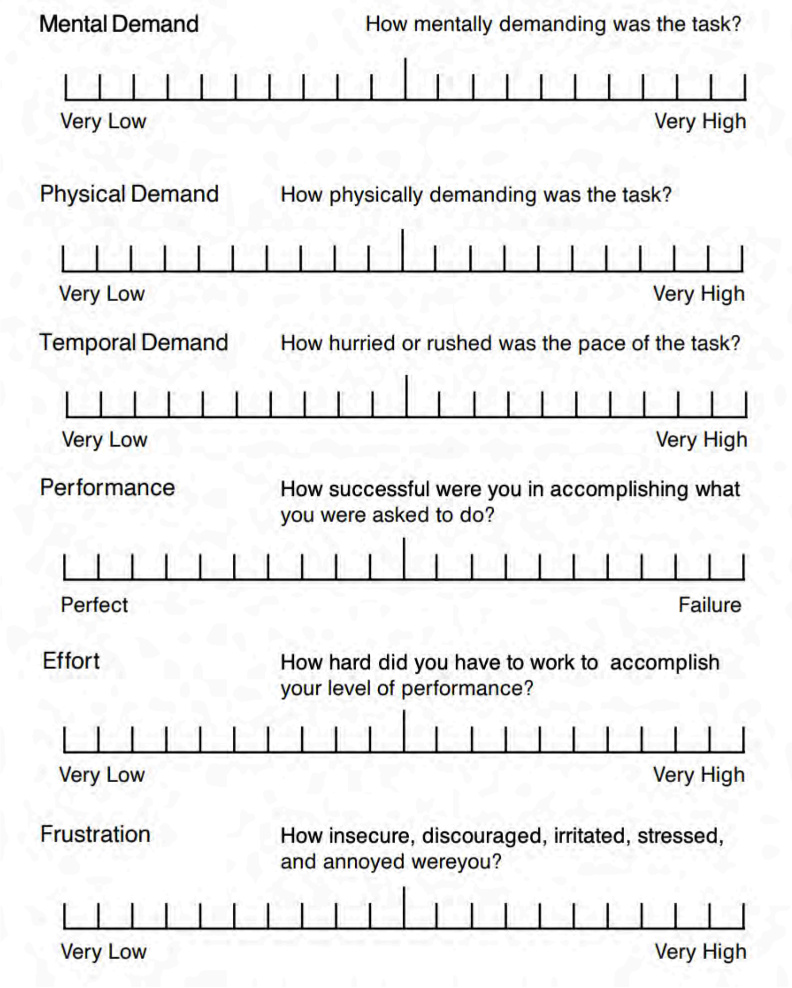


Figure 3.2: Graphic presentation used to visualize instructional efficiency, adapted from [50]

3.4.3 NASA-task load index

NASA-task load index(NASA-TLX) [51] consists of six subscales that represent somewhat independent clusters of variables: Mental, Physical, and Temporal Demands, Frustration, Effort, and Performance (Figure 3.3).



Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

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

Very Low — Very High

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

Perfect — Failure

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

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure 3.3: NASA-task load index, adapted from [48]

4

Methodology

To investigate how a multimodal approach can enhance the interactive navigation system of shared e-scooters, this study employs a mixed-methods research design combining qualitative and quantitative techniques. The methodology is structured as follows:

4.1 Research Approach

A Research through Design (RtD) approach will be employed. As Gaver (2012) discussed in "What Should We Expect From Research Through Design?", research through design acquires insights into specific issues via design practice. The theories it generates tend to be provisional, context-specific, and inspiring [52]. This approach enables us to continuously explore and discover new possibilities during the iterative design of the multimodal navigation system, rather than merely relying on pre-established theoretical frameworks. Moreover, Gaver pointed out that research through design is generative and is more concerned with creating what might be. This aligns with our goal of enhancing the shared e-scooter navigation system through design innovation, helping us break free from traditional thinking and create a better navigation experience for users.

4.2 Design Approach

This study adopts a user-centered design (UCD) methodology, ensuring that the navigation system improvements align with user needs and behaviors. By applying UCD principles, this study aims to design a multimodal navigation system that enhances user experience while minimizing distractions. Through iterative prototyping and user evaluations, the system will be refined to align with real-world navigation challenges faced by e-scooter users.

4.3 Design Process

The design process follows the Double Diamond framework (Figure 4.1), a structured approach to problem-solving that consists of four phases: Discover, Define, Develop, and Deliver. This methodology ensures a balance between user research, ideation, prototyping, and refinement, leading to a user-centered navigation system for e-scooter riders [53].

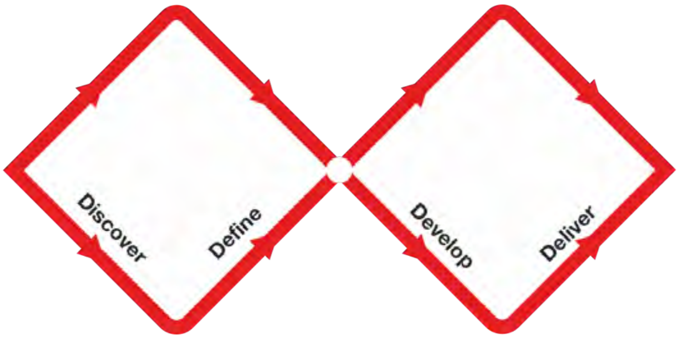


Figure 4.1: Double Diamond Model from Design Council, adapted from [54]

4.3.1 User Research

To develop a user-centered navigation system that effectively addresses the needs of e-scooter riders, various research methods will be employed. These methods will provide both quantitative and qualitative insights into user preferences, navigation challenges, and real-world behaviors. The combination of surveys, interviews, and observational studies will ensure a well-rounded understanding of user needs, guiding the design and refinement of the system.

4.3.1.1 Questionnaires

Questionnaires will be distributed to e-scooter users to gather quantitative data on their current navigation experiences, preferences, and expectations. These structured forms will include multiple-choice questions, Likert scale ratings, and open-ended responses to assess key aspects such as the frequency of navigation use, common pain points, and desired improvements. By reaching a broad audience, surveys will help identify patterns in user behavior and highlight recurring challenges, such as difficulties in following turn-by-turn directions or distractions caused by smartphone-based navigation. The results will inform design decisions by prioritizing the most common user concerns.

4.3.1.2 Interviews

To gain deeper qualitative insights, semi-structured interviews and focus groups will be conducted with e-scooter users. These discussions will explore context-specific challenges that may not be fully captured in surveys. Interviews will provide in-depth perspectives on user frustrations, safety concerns, and feature preferences, while focus groups will facilitate interactive discussions, allowing participants to compare their experiences and suggest potential improvements. These conversations will help ensure that the system is designed to align with actual user needs and expectations.

4.3.1.3 Observational Studies

Field observations will be conducted to analyze real-world navigation behaviors and identify pain points in actual usage conditions. Researchers will observe e-scooter users navigating different environments, such as busy city streets, bike lanes, and pedestrian areas, to assess how they interact with existing navigation tools. Key factors such as how often users check their screens, whether they rely on audio cues, and how they respond to environmental distractions will be documented. These observations will provide unfiltered behavioral data that may not be fully captured through self-reported surveys or interviews.

4.3.1.4 First-person perspective

First-person perspective studies will involve equipping participants with wearable cameras or head-mounted devices to record their real-time interactions with navigation tools during rides. This method captures unfiltered, context-rich footage of how users visually scan their environment, engage with UI elements, and divide attention between navigation cues and traffic conditions. By analyzing these recordings, researchers can identify subtle behavioral patterns (e.g., glance frequency, hand movements) that reflect usability issues or cognitive load in naturalistic settings.

4.3.2 Problem Definition and Design Goal

Defining the problem accurately is crucial to ensuring the design addresses real user challenges effectively. The research findings from interviews, focus groups, and observational studies will be synthesized using an Affinity Diagram [55] to categorize insights into key themes. This method helps in identifying patterns in user behaviors, pain points, and expectations, allowing for a structured approach to problem-solving.

4.3.2.1 Thematic Analysis

Thematic analysis is a method for identifying, analyzing, and interpreting patterns of meaning (themes) within qualitative data [56]. This method will be applied to qualitative data (e.g., interview transcripts, focus group discussions) to identify, organize, and interpret recurring themes related to navigation experiences. Researchers will systematically code data to uncover patterns in user needs, frustrations, and preferences (e.g., "difficulty interpreting complex visual cues," "reliance on audio instructions in high-risk scenarios"). This iterative process ensures themes are grounded in participant insights, providing a foundation for design requirements that address core user challenges.

4.3.2.2 Text Quantitative Analysis

Text quantitative analysis allows the identification of themes and the text relations with semantic analysis [57]. This method will involve using computational tools to quantify and statistically analyze linguistic patterns in open-ended survey responses, first-person perspective and interview data (e.g., word frequency, sentiment scores).

For example, researchers may measure the prevalence of keywords related to "distraction", "clarity" or "safety" to identify dominant user concerns. This approach complements qualitative thematic analysis by offering numerical insights into the salience of specific issues, helping prioritize design improvements based on data-driven trends.

4.3.2.3 MoSCoW

The MoSCoW method will be used to prioritize design requirements by categorizing them into four tiers: Must-have (critical for basic functionality, e.g., clear turn directions in high-risk zones), Should-have (important but non-essential, e.g., personalized audio tone options), Could-have (nice-to-have enhancements, e.g., AR-based landmark highlighting), and Wont-have (out-of-scope for this iteration, e.g., social sharing features). This framework ensures resources are allocated to high-impact solutions first, aligning the design process with user priorities and technical feasibility.

4.3.3 Design & Ideation

The design and ideation phase is essential in transforming user research insights into actionable design concepts. This phase ensures that the navigation system is developed with user needs at the forefront, allowing for iterative improvements based on real feedback.

Concept development begins with generating initial design ideas informed by user research findings. These ideas are translated into low-fidelity sketches and wireframes, focusing on key navigation features, interface layout, and interaction flows. The goal is to explore multiple possible solutions before refining the most promising ones. By visualizing different approaches early, potential usability issues can be identified and addressed before investing time in high-fidelity prototypes.

4.3.4 Prototyping

Prototyping is a critical step in the design process as it allows designers to test, evaluate, and refine ideas before committing to a final solution. According to Lim, Stolterman, and Tenenberg (2008), prototypes serve as both *filters* and *manifestations* of design ideas, helping designers explore specific interactions while narrowing down effective solutions [58].

Prototypes can vary across several dimensions as a manifestation, including [58]:

- **Material:** The substance or medium used to create the prototype, such as paper, digital interfaces, or physical models.
- **Resolution:** The level of detail and refinement, ranges from rough sketches (low resolution) to polished, detailed representations (high resolution).
- **Scope:** The extent of the design aspects covered, whether focusing on a specific component or encompassing the entire system.

Prototypes can also vary across different dimensions as filters, influencing what aspects of a design are tested and refined. These filtering dimensions include:

- **Appearance:** Visual elements and aesthetics of the design.
- **Data:** Information handling and processing within the design.
- **Functionality:** Operational features and behaviors of the design.
- **Interactivity:** The nature and quality of user interactions are facilitated by the design.
- **Spatial Structure:** The arrangement and organization of elements within the prototype, influence how users perceive and interact with the system.

In the context of a multimodal navigation system for e-scooters, prototyping enables iterative improvements, ensuring that design decisions are driven by real user needs and usability testing rather than assumptions. By developing prototypes early, potential usability issues can be identified and resolved before the final implementation, reducing development risks and enhancing the overall user experience.

4.3.4.1 Visual Instruction

The goal is to create clear and easily readable on-screen and even more out-of-screen navigation instructions(such as AR or head-up display) that riders can quickly glance at without distraction. This includes optimizing font size, contrast, iconography, and layout to ensure high visibility, even in outdoor conditions where lighting may vary.

4.3.4.2 Auditory Instruction

Since riders cannot always look at a screen, the aim is to design adaptive voice prompts that provide navigation instructions while ensuring ambient traffic sounds remain audible for safety. Factors such as volume control, speech clarity, and frequency adjustments will be tested to create a non-intrusive yet effective audio experience.

4.3.4.3 Haptic Feedback

The advantage of haptic feedback is that users can use it anywhere, anytime, without interrupting others [59]. To further minimize distractions, we are designing vibration-based cues embedded in the handlebars to signal turns, directional changes, or nearby hazards. This ensures that riders can receive navigation guidance without needing to rely on vision or hearing, making the system more accessible and intuitive.

4.3.5 Evaluation & Testing

To ensure the effectiveness, safety, and usability of the proposed navigation system, a series of structured evaluation methods will be employed. These assessments will provide both qualitative and quantitative insights into user experience, cognitive demands, and overall performance compared to existing solutions.

4.3.5.1 A/B Testing

A/B testing is a method used to compare two versions of a design to determine which performs better in meeting user needs [60]. In this thesis, A/B testing was employed to compare the original Google Maps navigation interface with the improved navigation interface design. The goal was to evaluate whether the redesigned interface could reduce user cognitive load and improve usability. The study used the NASA-TLX scale to measure mental workload across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. Wilcoxon tests were then conducted to assess significant differences between the two interfaces. The results provided empirical evidence for the effectiveness of the redesigned interface in enhancing navigation experience, highlighting whether modifications led to statistically significant improvements in user interaction and cognitive efficiency.

4.3.5.2 Cognitive Load Analysis

To measure the mental effort required during navigation, cognitive load analysis will be performed using the NASA-TLX (Task Load Index) tool [51]and the Efficiency equation [49]. Participants will rate different aspects of their cognitive burden, such as mental demand, frustration, and perceived effort, after completing navigation tasks. This analysis will help determine if multimodal guidance reduces cognitive strain, making navigation more efficient and less overwhelming for users.

In this study, the NASA-TLX used default/standard weight assignment, meaning that in the cognitive scale for the user driving task, each parameter was equally assigned a weight of 1/6. The details are as follows:

$$\text{Weighted Score} = \frac{\text{Mental} + \text{Physical} + \text{Temporal} + \text{Performance} + \text{Effort} + \text{Frustration}}{6}$$

4.3.5.3 Wilcoxon test

The Wilcoxon test (a non-parametric statistical method) [61] will be employed to analyze ordinal or non-normally distributed data (e.g., NASA-TLX ratings, user preference rankings) collected from paired samples (e.g., comparing mental effort scores between the old and new UI designs). This test helps determine whether there are statistically significant differences in user performance or subjective perceptions across conditions (e.g., low vs. high-risk cycling scenarios) when parametric assumptions (e.g., normal distribution) are not met. By ranking data points and assessing shifts in distribution, the Wilcoxon test provides robust insights into the effectiveness of design modifications while accommodating the studys mixed-methods data structure. When conducting repeated validations across multiple groups to prevent data bias, Bonferonni correction [62] is performed.

$$\alpha_{\text{corrected}} = \frac{\alpha}{\text{number of comparisons}} \tag{4.1}$$

4.3.6 Iterative Design

The usability optimization phase is a critical step in refining the multimodal navigation system to ensure a seamless and intuitive user experience. Building on insights gathered from usability testing, we will address key usability issues by identifying pain points, such as confusing feedback cues or inefficient interaction flows, and implementing targeted improvements. A primary focus will be on enhancing feedback mechanisms- visual, auditory, and haptic- by ensuring they are clear, timely, and contextually appropriate to guide users effectively without causing distractions. Additionally, we will streamline interaction flows to reduce cognitive load, making it easier for riders to process navigation instructions quickly and intuitively, even in fast-paced or complex environments. The system interface will be simplified by eliminating unnecessary steps and aligning interaction patterns with users' natural behaviors and expectations. Through iterative design rides, we will continuously test and refine these improvements, ensuring that the final prototype is not only efficient but also user-centered, promoting safety and ease of use in real-world riding conditions.

5

Planning

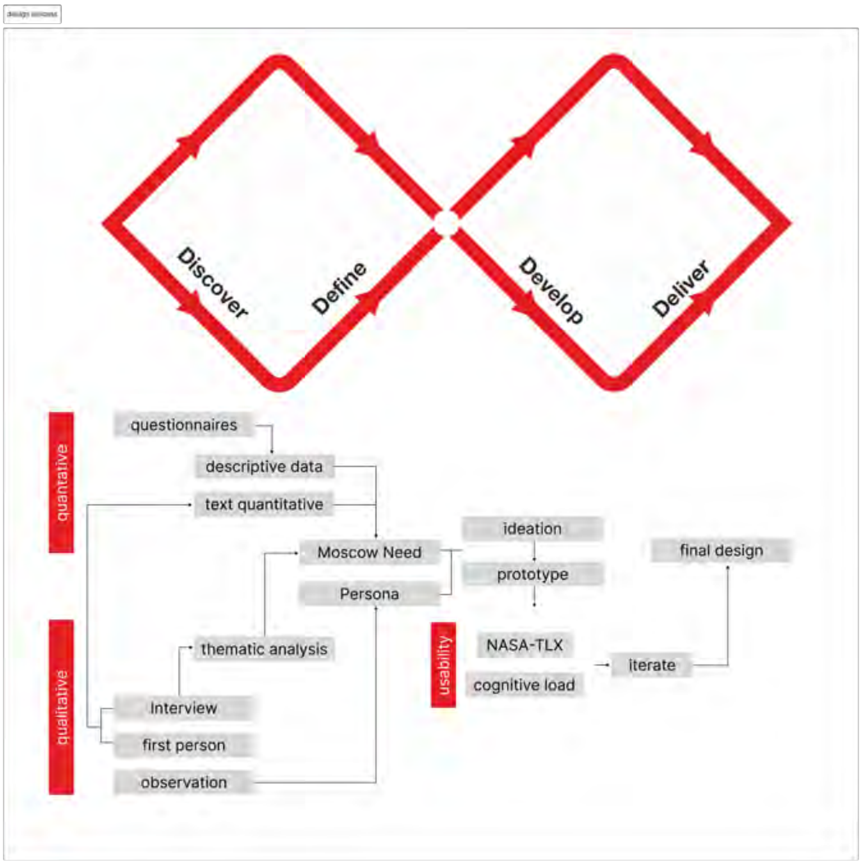


Figure 5.1: Design process and user studies of the thesis

The thesis project follows a structured user-centered design (UCD) methodology, integrating iterative feedback loops to ensure continuous improvement within a compressed three-month timeline (Figure 5.1). By continuously involving users in the design process, it can be identified usability challenges early, refine solutions based on real feedback, and create a system that aligns with actual rider behavior and preferences. This iterative approach ensures that our final product is not only functional but also enhances user experience and safety in real-world scenarios.

The planning is divided into 8 key phases, some of which overlap to maximize efficiency. The literature review phase lasts for 1 to 2 weeks, focusing on reviewing relevant research on navigation, micromobility, and user-centered design theories. It runs in parallel with the user research phase, which spans 1.5 weeks and involves surveys, interviews, and observational studies to gather insights into user needs and behaviors.

Following this, the design and ideation phase takes place over 1 to 2 weeks, translating research findings into initial design concepts through brainstorming, storyboarding, and wireframing. This phase is crucial because it transforms user insights into tangible design solutions. It is important to explore multiple ideas, test different interaction models, and validate design assumptions before committing to a final approach. A well-executed ideation phase ensures that the navigation system is intuitive, accessible, and aligned with user expectations. Additionally, this phase runs parallel to prototyping, which lasts for two weeks and involves the development of a high-fidelity interactive prototype incorporating visual, auditory, and haptic feedback elements.

Evaluation and testing begin midway through prototyping and last for 1 week. It includes usability testing, cognitive load assessments, and performance benchmarking. The insights gained from testing feed into the data analysis phase, which lasts for half a week and involves qualitative and quantitative analysis of user feedback and performance metrics.

The iteration and refinement phase, lasting 1 week, focuses on making necessary improvements based on evaluation results, ensuring optimal usability. The final phase, documentation and final submission, takes one week, involves the compilation of findings, preparation of the final report, and submission of the thesis.

Figure 5.2: Gantt Chart of Planning

However, during the final implementation phaseparticularly the prototyping and user testing stagesadjustments were made to the initially planned design. In the planning phase, tangible interaction designs had been envisioned, including hardware components such as haptic vibration feedback and a real-time interactive AR projection lamp. Nevertheless, due to constraints in time and economic costs during the actual execution, user usability testing was ultimately conducted in a virtual reality (VR) environment.

During this VR-based testing phase, a critical limitation of the vibration feedback module emerged: its effective operation relied on highly precise hardware calibration, which was not feasible to achieve within the projects resource and timeline boundaries. Consequently, the vibration feedback modality was excluded from the final deliverables.

6

Execution

6.1 User Research

From the "what-how-why" user research framework [63], at this stage, we mainly study the behaviors of users at these three levels to help us define the goals of the navigation systems in e-scooters.

As Christian Rohrer mentioned [64], the research on user experience is mainly divided into a 3-dimensional framework with the following axes:

- Attitudinal vs. Behavioral: This distinction can be summed up by contrasting "what people say" versus "what people do".
- Qualitative vs. Quantitative
- Context of Use

In order to explore the characteristics of user behaviors, this research mainly adopted qualitative analysis methods such as interviews and field investigations to think about the user behaviors at the "why" and "how" levels. Quantitative analysis is mainly used to explore "how many" and "how much", helping us to consider the duality of problems: for example, whether the pain point exists and whether users care about this behavior (Figure 6.1).

Figure 6.1: The user research methods which the thesis conducted

6.1.1 Quantitative Research

6.1.1.1 Questionnaires

In this user research, questionnaires were mainly used in the initial stage. The main research objectives are as follows:

- To understand users' usage habits, which helps us define users' main applicable needs for navigation.
- To obtain users' riding behavior patterns and scenarios.
- To explore the riding capabilities and habits of the main user group.
- To understand users' expectations regarding riding safety and experience-side requirements.

In the questionnaire, we mainly divided it into five parts, namely personal information, riding experience, safety perception, navigation and interaction, and future improvements (Appendix B). Among all the questionnaire questions, there are 13 columns of categorical variables and 10 columns of numerical variables. For numerical variables, we used the 5-points Likert scale [65] to conduct a satisfaction analysis of the factors that influence the riding experience(Table 6.1).

In this study, questionnaires were distributed online in European cities where scooters are available, such as Stockholm, Gothenburg, Milan, Copenhagen, etc. The distribution channels included WeChat groups and social media platforms like LinkedIn. The collection period lasted for one week.

6.1.1.2 Text Quantitative Analysis

The text sources of this study mainly come from the interviews in the qualitative analysis stage, some open questions from questionnaires and the log texts written from a first-person perspective. First, all text data(around 4,900 words) was cleaned and encoded. For instance, other languages were translated into English.

Then, this study conducted data sorting and key information extraction: Carefully read the text data and extract key information relevant to the 5 research questions, such as usage frequency, usage scenarios, opinions on safety, requirements for navigation functions, and evaluations of vehicle design.

Finally, through semantic and thematic analysis, it is possible to gain an in-depth understanding of users' opinions, needs, and concerns regarding electric scooters. Regarding semantic analysis, this study utilizes the Rawgraph(`https://app.rawgraphs.io/`), which is mainly used to display the proportions of positive, neutral, and negative elements in the text.

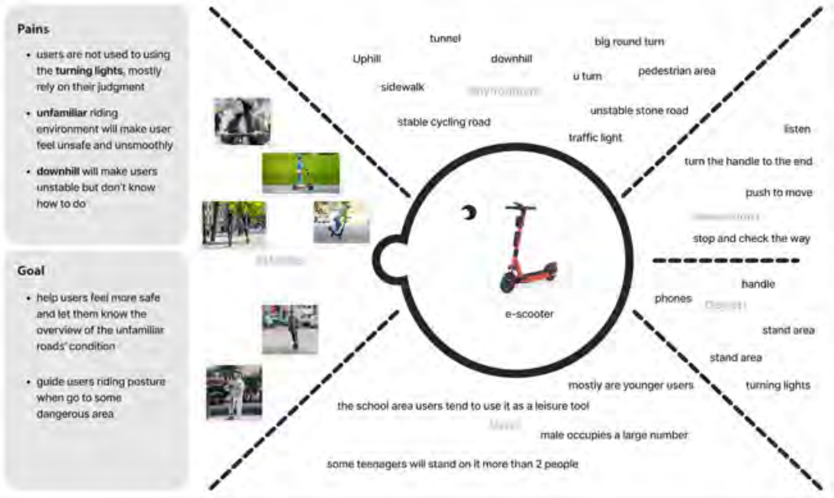
Table 6.1: Factors that influence the riding experience

No.	Category	Question
F1	Safety Perception (overall)	How safe do you feel while riding an e-scooter on a scale
F2	Safety Perception (attachments)	How important do you think it is for e-scooters to have safety features like lights, brakes, and helmets
F3	Safety Perception (handlebars)	How comfortable do you find the handlebars of the e-scooter on a scale?
F4	Safety Perception (foot-rest)	How comfortable do you find the foot-rest area of the e-scooter on a scale?
F5	Navigation and Interaction (phone holder)	To what extent do you want to put your phones inside the phone holder on a scale?
F6	Navigation and Interaction (audio)	Auditory (beeping, voice prompts)
F7	Navigation and Interaction (visual)	Visual (lights, text messages)
F8	Navigation and Interaction (vibratactile)	Vibrotactile (vibrations on the handlebars)
F9	Future Improvement (software)	If an e-scooter had a more advanced interaction system, like touch-screen controls or voice commands, would you use it?
F10	Future Improvement (hardware)	If in the future there is a helmet/handlebar that combines vibration and voice interaction, would you be willing to use it?

6.1.2 Qualitative Research

6.1.2.1 On-site observation

In this study, through a week of continuous tracking and observation, the behaviors of people using scooters were observed on specific road sections, such as areas around schools, near bus stops, and in bustling downtown regions. Finally, the empathy map tool was used to record these observations (Figure 6.2).


Pains

- users are not used to using the **turning lights**, mostly rely on their judgment
- **unfamiliar** riding environment will make user feel unsafe and unsmoothly
- **downhill** will make users unstable but don't know how to do

Goal

- help users feel more safe and let them know the overview of the unfamiliar roads' condition
- guide users riding posture when go to some dangerous area

Figure 6.2: Empathy map from the observation study

6.1.2.2 First-person perspective

This study invited three senior e-scooter users to write down their riding record and some details thoughts (Figure 6.3). The records were mainly in text form. Eventually, we obtained the initial data. Then, through data cleaning and encoding, we used these data as the text source for quantitative analysis.

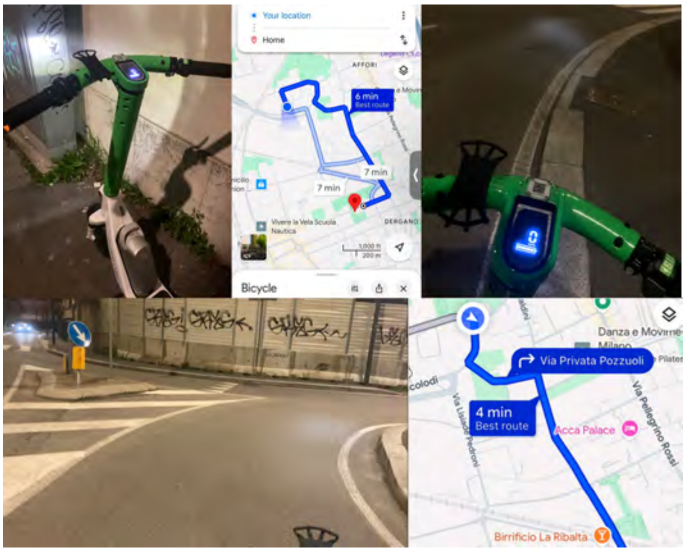


Figure 6.3: First-person perspective riding log in Milan, photographed on 2025-04-25

6.1.2.3 Semi-structured interview

This research mainly adopts the semi-structured interview method for interviews (Appendix A). The main implementation means include face-to-face interviews, online interviews, and telephone conferences. The duration of a single interview is controlled at around 30 minutes. The main purposes of this interview are as follows:

- To understand users' attitudes towards the current riding situation and their expectations.
- Gain an in-depth understanding of users' riding habits and navigation interaction habits.
- To explore users' acceptance level of multimodal interaction methods.

To this end, this research first adopted the method of stratified sampling interviews [66], inviting 12 interviewees(Table 6.2). First, based on the frequency and experience of users' riding, this research divided users into three categories: occasional users (who have a small amount of riding experience and occasionally choose to travel by scooter), extreme users (who have no riding experience and will not consider riding as a means of transportation), and typical users (who often ride). From different categories of users, this research can obtain the expectations and needs of different users regarding riding behavior, thus achieving more accurate and diverse results.

Table 6.2: 12 participants in the interview phase

No.	Gender	Age	User Category	Riding Experience
P1	Female	27	Occasional User	2 times
P2	Female	28	Occasional User	2 times
P3	Female	27	Extreme User	Never
P4	Male	22	Typical User	3-4 times a week
P5	Female	24	Occasional User	More than 10 times
P6	Male	23	Typical User	2 times a week
P7	Male	24	Typical User	4-5 times a week
P8	Male	28	Typical User	3-4 times a week
P9	Male	26	Occasional User	More than 10 times
P10	Male	41	Extreme User	Never
P11	Male	31	Occasional User	More than 10 times
P12	Male	23	Typical User	3-4 times a week

In this study, all interviews were recorded and then directly transcribed into text. For non-native speakers in the interviews, the text was translated into English using the GPT-4o model. Subsequently, information was extracted from the text for thematic analysis.

6.1.3 Context of Use

One strand of research has zeroed in on objective driving risk analysis. This area can be partitioned into two key aspects: macroscopic traffic data and microscopic operating parameters of the driving process. These two approaches are employed to identify and analyze risk factors [67].

In the scenario of riding a scooter, macroscopic traffic can be regarded as the ordinate indicator-predictable, and microscopic operating parameters can be regarded as the abscissa indicator-familiar (Figure 6.4).

Figure 6.4: Risk analysis based on riding context

In terms of classifying users' riding scenarios, according to users' familiarity with the riding scenarios and the predictability of the riding conditions, it can be divided into three types of risky riding: low-risk, moderate-risk, and high-risk. Zheng et al. proposed in 2021 that there are two main factors influencing users' decisions during driving, namely safety and efficiency [68]. For different riding risk modes, users' main decision-making factors also vary.

- **Low-risk:**
  When the riding time in a single direction exceeds 1.5 seconds, users will feel extremely safe [68]. This situation mostly occurs in familiar and predictable environments. At this time, the main goal of users' riding is dominated by efficiency (Figure 6.5).
- **Moderate-risk:**
  For roads that are familiar but unpredictable, users are mainly dominated by the pursuit of efficiency.
  For roads that are unfamiliar but predictable, users are mainly guided by safety. At this time, the riding characteristics of users are generally as follows: The navigation only occupies a single channel for interaction with the users, and the users' main attention remains on the road environment rather than the route.

- **High-risk:**
  When safety is the dominant factor, users need to make dynamic decisions in a dynamic environment.

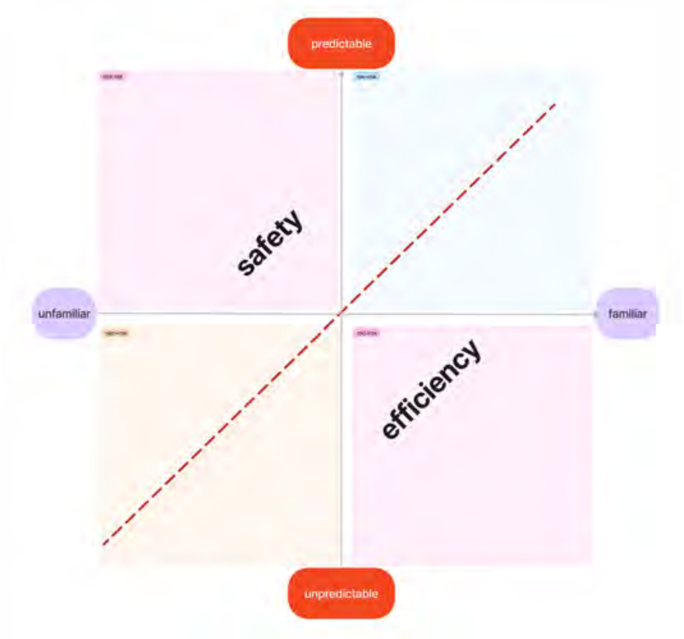


Figure 6.5: User goal when meeting different riding contexts

6.2 Problem Definition

6.2.1 Persona

To design an effective multimodal navigation system for shared e-scooters, a deep understanding of user behavior, pain points, and contextual needs is critical. Such insight ensures that the system aligns with real-world usage patterns and addresses the specific challenges riders face. Building on findings from user research, this persona analysis identifies key rider profiles, each representing distinct expectations and navigation challenges that inform user-centered design decisions (Table 6.3).

Table 6.3: User Persona Analysis for Shared E-Scooter Design

Aspect	**Persona 1: Anna**	**Persona 2: Ling**
Character	22-year-old student at the University of Gothenburg; lives near Lindholmen campus; commutes 1.53 km daily for school, groceries, and social visits using e-scooters.	25-year-old tourist visiting Stockholm for the first time; first-time e-scooter rider; unfamiliar with operations and concerned about safety.

Aspect	Persona 1: Anna	Persona 2: Ling
Goals	• Quickly complete short-distance commutes. • Reduce walking time and discomfort in cold weather.	• Easily visit tourist attractions. • Explore the city with convenient short-distance transport.
Pain Points and Needs	• Fixed phone mount angle requires looking down, reducing visibility and control. • Frequent urban road maintenance and changing layouts. • Navigation fails to reflect updates, causing misdirection (e.g., near Chalmers).	• Unfamiliar with braking; sudden acceleration causes stress. • Silent scooter operation makes it risky around unaware pedestrians. • Navigation voice is in English, while street names are in Swedish; prefers reading maps.
Behavior Patterns	• Uses scooters mainly in daylight and good weather. • Relies on memory for familiar routes; uses navigation only when necessary.	• Avoids riding at night or in bad weather. • Uses social media to share travel moments (e.g., photo stops).

6.2.2 MoSCoW Analysis

This study adopted the MOSCOW method to define design objectives, starting from the user group to understand, summarize, and finalize the product scope (Table 6.4).

Table 6.4: MoSCoW Analysis for Shared E-Scooter System Design

Category	Features
Must Have	• Optimization of brake system • Improvement of mobile phone holder reliability • Design of anti-slip tires and vehicle body stability • Integration of basic navigation functions
Should Have	• Real-time traffic update and route optimization • Adjustment of economic model • Windproof and rainproof design

Category	**Features**
Could Have	• Vibration feedback and multilingual voice prompts • Nighttime lighting and reflective design • User feedback and emergency support system
Won't Have	• Shared helmet rental system • Long-range battery

6.3 Design Concept

During the preliminary questionnaire analysis and interviews, many users expressed anticipation for vibration-based interaction. Therefore, in the earliest concept development phase, this study primarily explored three modalitiesvisual, auditory, and vibration-as navigation input modalities to assist user interaction (Figure 6.6).

Figure 6.6: Design concept of the multi-modal navigation systems

For the visual component of the conceptual design: 1. Optimize the interface interaction design to better align with users' mental models during riding. 2. Introduce an AR projector interaction to assist users with directional recognition, reducing frequent head-down movements and stops during driving.

For auditory interaction: 1. Shorten the duration of voice interaction. 2. Refine the content prompts of voice interaction (e.g., reduce street names that may cause confusion due to language differences and focus on directional cues).

For vibration interaction: Activate handlebar vibration prompts before users make turns.

What's more, combined with the context of use, three riding concept hypotheses under different driving risk modes are proposed as follows (Figure 6.7):

- **Low-risk scenario:** This scenario is where users can perform single-operation within 1.5 seconds [68]): The main interaction methods are single-modality interactions, such as auditory or AR.

- **Medium-risk scenario:** Dynamic interaction design based on users' driving behaviors (mainly reflected in users' control of handlebar speed):
  1) When users press the brake, the UI dynamically adjusts information, and audio and AR interactions are displayed simultaneously.
  2) When users only coast, audio and AR interactions serve as the main input modalities, with the UI assisting in displaying directional information.

- **High-risk scenario:** Maintain the effectiveness of the UI at all times, with AR and audio as auxiliary reminders.

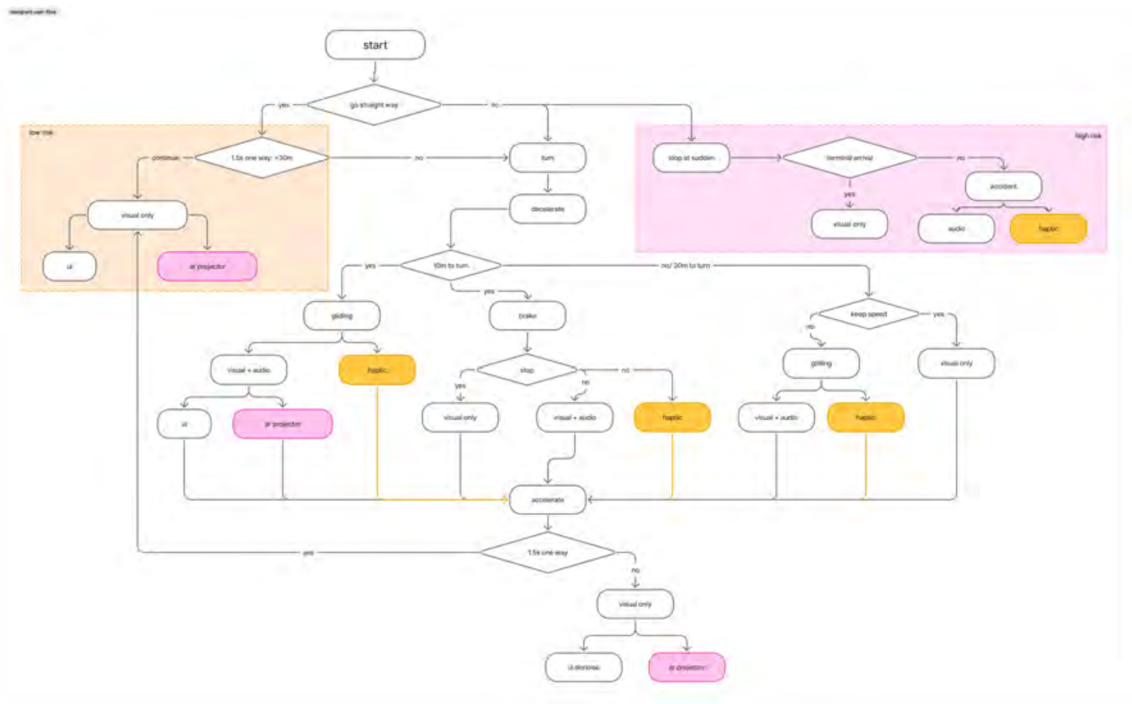


Figure 6.7: User flow and interaction design based on different scenarios

6.4 Prototype

At this stage, low-fidelity prototypes were primarily used to conduct early pivot studies on the design. Through continuous feedback from users and iterative design, a more refined design solution was developed, which also helped reduce the pressure during the usability testing phase. This section of the thesis only presents the low-fidelity prototypes, while the high-fidelity prototypes are shown in the results section.

6.4.1 Software design

To address the issues of excessive information display on the original UI interface, overly granular navigation interaction (e.g., navigation instructions changing every 10 meters with frequent directional reminders), and difficult touch interaction due to the phone being placed on a stand, the interface design introduces three key innovations:

- Remove street information display and enlarge directional indicators to reduce cognitive load (Figure 6.8).
- Dynamically adjust interface information based on the users riding speed and distance to the next turn. Set 30 meters as the first turn reminder (the headway distance [69] calculated using the formula: 1.5s (critical point for low/medium-risk scenarios) * 25km/h + braking distance Œ friction coefficient 30m). At 10 meters before the turn, highlight the prompt and enlarge junction details (Figure 6.9).
- Minimize accidental navigation exit by placing the exit button in the left-hand operation area and the thumbnail toggle function in the right-hand area for easy access.

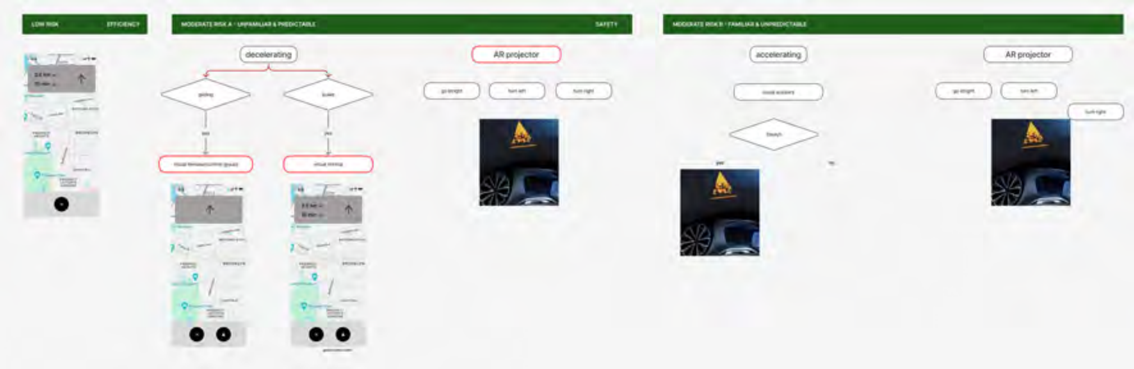


Figure 6.8: Low fidelity of the app: low risk and medium scenario

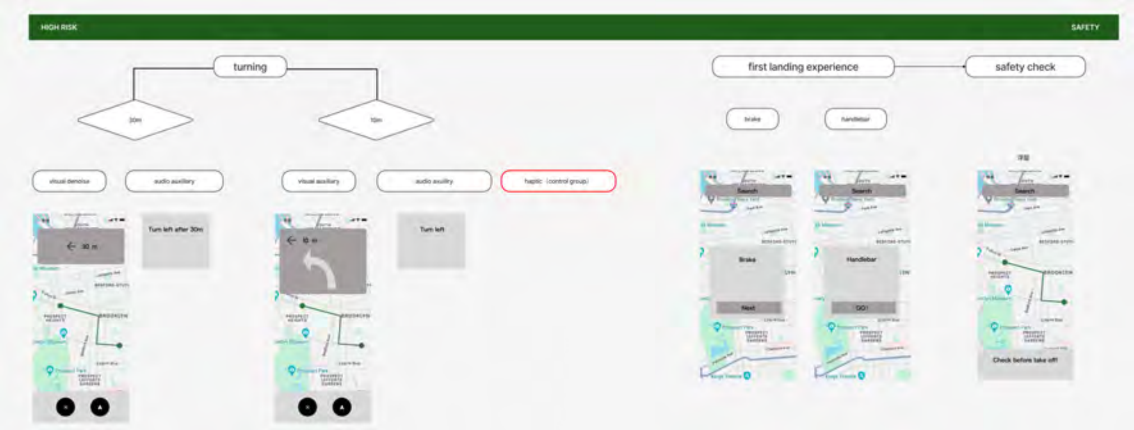


Figure 6.9: Low fidelity of the app: High risk scenario

6.4.2 AR Projector design

In terms of hardware support, this design draws inspiration from the Head-Up Display (HUD) in vehicles. It explores the possibility of developing a physical carrier to present specific directional information, thereby enabling users to identify directions more efficiently. Against this backdrop, the concept of an AR projector was conceived.

The core rationale behind this inspiration lies in the inherent advantage of HUD to deliver critical information (e.g. navigation cues, vehicle data) directly within the users line of sight, eliminating the need for users to shift their focus between the environment and a separate display, an attribute considered highly transferable to scenarios of direction guidance. By extending this logic, the AR projector is intended to serve as a dedicated physical interface: unlike abstract digital prompts on a mobile or desktop screen, it projects directional indicators (e.g., arrows, landmarks) onto real-world surfaces (e.g., roads), creating a more intuitive, context-integrated way for users to perceive and act on directional information. This design choice addresses the common challenge of disorientation in digital navigation, where users often struggle to map on-screen directions to their physical surroundings.

The preliminary design details for the projection light hardware are as follows:

- **Position and Projection Distance:** The projection light is placed on the front of the e-scooter. The specific height and angle require testing across different brands and are not discussed further here (Figure 6.10).

- **Projection Frequency:** Existing research provides limited theoretical guidance on blinking patterns. Given the challenges of defining optimal frequencies and durations at the low-fidelity prototype stage, the initial interaction involves projecting indicators for 2 seconds exclusively at turns.

- **Color Coding:** Green for straight paths and red for turns.

- **Signal Indicators:** The initial interaction includes only straight, left, and right directions. More complex scenarios requiring higher data precision are not addressed at this stage.

- **Interaction and Connectivity:** The projection light synchronizes with map data via Bluetooth from the users smartphone.

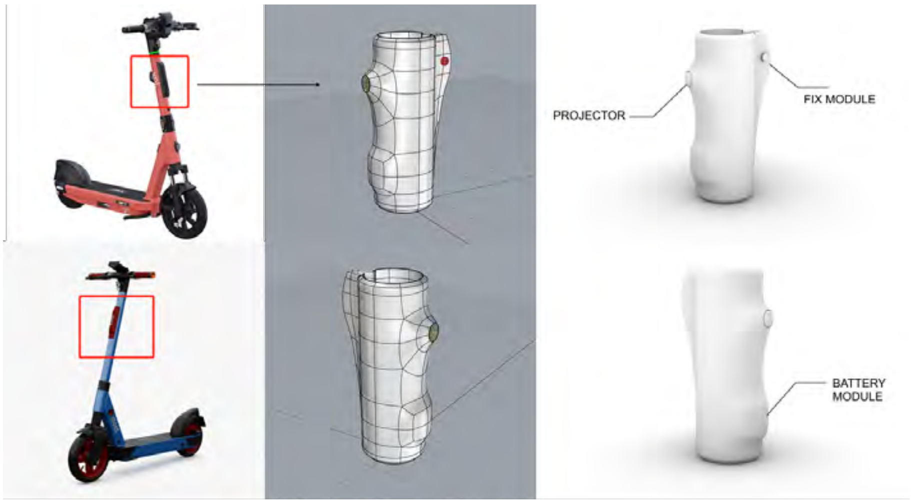


Figure 6.10: Low fidelity of Model in the Rhinoceros

6.5 Iterative Design

During the preliminary design phase, five users who had previously participated in interviews were invited to conduct a pilot study based on low-fidelity prototypes. The study aimed to observe whether the design met their initial expectations and effectively addressed their pain points. Following this, a second round of optimization was carried out before proceeding to high-fidelity prototype development and the setup of a usability testing environment.

To enhance participants' task experience and obtain more objective psychological cognition ratings among different modalities, this study constructed a simulated riding environment within a VR setting. Users could control the riding movement by standing or sitting and operating a directional keyboard, and the UI display was toggled through head-up and head-down movements, mirroring the real-world behavior of looking at a mobile navigation device. The VR environment meticulously replicated the complexity of road conditions, integrating various risk scenarios (Figure 6.11). In low-risk situations, users were required to ride straight for 1.5 seconds, while high-risk scenarios featured moving ridists and roadblocks that demanded users to make evasive maneuvers.

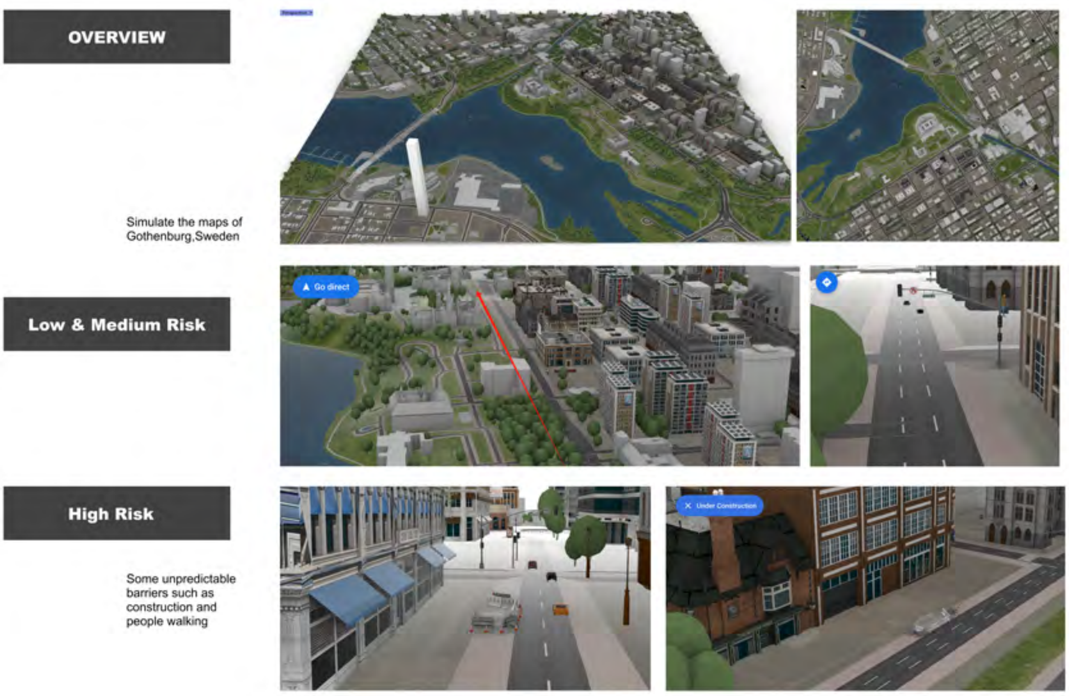


Figure 6.11: Environment build-up for testing the user's driving attitudes on the design concept in Unity, the video of usability test `https://drive.google.com/file/d/1SeaHAHI4Y9_0RG5nwu7Xwym0sCfB-zxM/view?usp=sharing`

During testing, while participants manipulated controls, researchers observed their movements via a screen and provided limited assistance, recording driving behaviors and errors (Figure 6.12). After each ride, participants were asked simple questions, such as: "Which modality (or combination) did you prefer most, and why?" and "How do you think this would differ or what additional concerns might arise in a real riding environment?" This mixed approach of observation and post-task interviews aimed to capture both quantitative usability data and qualitative insights into user preferences and real-world applicability.

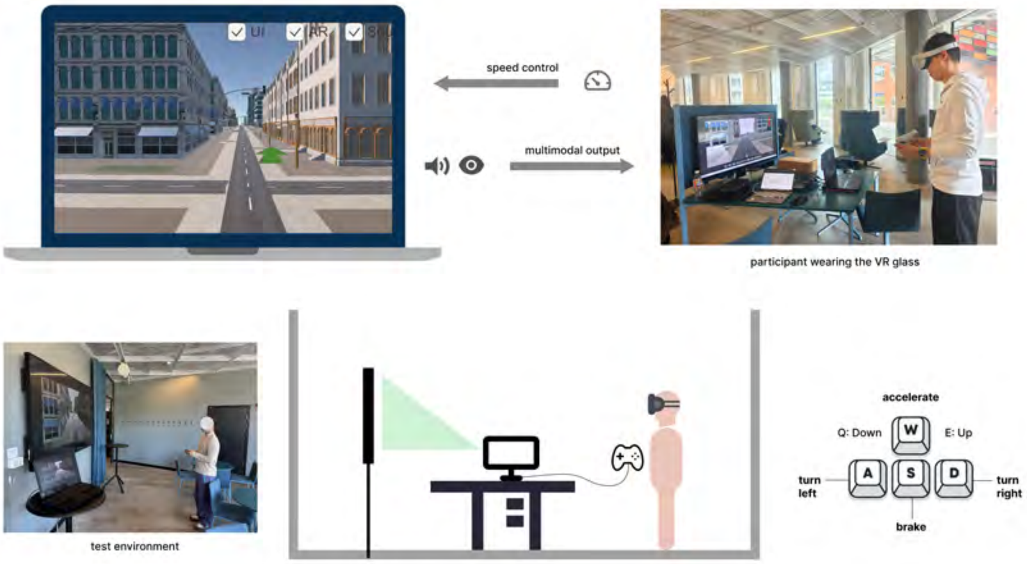


Figure 6.12: Experiment for participants to do usability test

When establishing the testing environment, the research used Unity to simulate scenarios in a VR environment for user testing. However, during this process, it was found that virtual simulation of vibration modality was highly challenging. Forcing vibration feedback (e.g., sudden 1-second vibrations from VR controllers during simulated riding) risked introducing errors into the research results, rendering them unreliable. As a result, discussions of this modality were omitted from the final outcomes.

After analyzing the results, the design solutions were optimized by integrating Google Guidelines and multimodal design principles, with the final outcomes presented in the *Final Result* section of the thesis.

7

Results

This chapter primarily introduces the results of the research phase and the final design outcomes. The research achievements include conclusions from the user research phase and results from the usability testing phase, spanning from the outputs of the preliminary research stage to the verification and iteration of the high-fidelity prototype.

7.1 Research Results

7.1.1 Results of Qualititive Research

7.1.1.1 Interview

In this study, the information from user interviews was divided into "user need" (the pain points encountered by users, such as the too-high price) and "user want" (the functions expected by users, such as a navigation system that can adapt to the weather). According to the previous text quantitative analysis, the four main themes(External environmental conditions, Braking and warning system, The conditions of the scooter design, Commercial Strategy) are defined and more detailed root-theme are explored here (Figure 7.1):

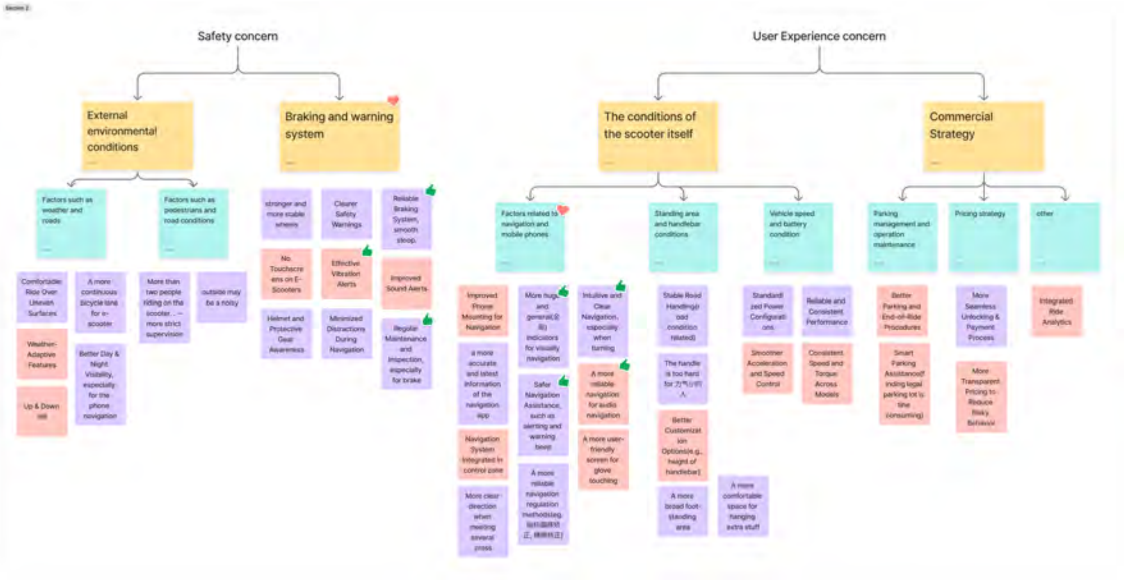


Figure 7.1: Affinity diagram from the semi-structured interview

In terms of user experience, the needs on the commercial side have also been mentioned by a considerable number of people, mainly involving the charging standards and the QR code scanning process. This study will not delve too deeply into these aspects. Regarding the riding experience of the scooter design, apart from the standing area and battery life, what users are most concerned about is the navigation issue. There are problems at the software level, and more specifically in the design, currently the interaction is not very intuitive, unreliable, and the interaction methods are rather monotonous and dangerous. For example:

> "Yes, I use Google Maps on my phone if I'm going somewhere new. Since I cant hold my phone while riding, I usually start the navigation in audio mode and listen through my earphones. Sometimes, I also glance at my phone while stopped, but thats not ideal." [P8]
>
> "I think it's better to combine different types of warnings. Sometimes, the visual navigation position on apps like Google Maps can be inaccurate." [P6]

There are also problems at the hardware level, and what is most frequently mentioned is the height and angle issues of the phone holder. For example:

> "So, I think they do it on purpose. They don't make the phone holder properly, so that you have to use these fragmented moments of time to check the navigation, and in this way, they can make more money." [P2]
>
> "The current placement for securing the phone isn't very convenient, especially for larger phones." [P7]

7.1.1.2 Themetic Analysis

Through semantic analysis, it is found that users' positive evaluations of the scooter riding experience are generally few (9.6%). The negative evaluations mainly focus on aspects such as safety, navigation, and cost (60.5%), while the neutral evaluations mainly concentrate on objective comparisons, such as the weather, roads, etc. (29.8%). The specific emotion-theme correlations are as follows:

Table 7.1: Semantic and thematic Analysis

Theme	Negative	Neutral	Positive	Typical words
Safety and design	45%	10%	-	brake failure, unstable, accident, vibration
Navigation and Interaction	30%	5%	-	inaccuracy, difficulty using, distraction
Cost	25%	15%	-	unfair pricing, public transport comparison, extra fees
Convenience	-	20%	60%	time-saving, instant access, no waiting, short distance

According to the semantic analysis, this study has found that within the aspect of safety concern, the braking and warning system is of great importance in enhancing the perception of safety. Many users have once encountered safety crises due to braking issues. For example:

> "Its mainly the brakes-especially when accelerating and braking, the scooter is very abrupt. When you brake, the scooter wobbles from side to side, making it unstable. It doesnt feel very secure." [P1]

> "Sometimes, when going downhill and needing to brake urgently, the brakes of the Ryde e-scooter slow down gradually instead of immediately. So, sometimes the braking might be too late. You need to predict the stopping distance in advance, and it can be difficult to stop in time, which makes people feel a bit dangerous. " [P12]

Figure 7.2: Visualization of semantic and thematic analysis

By quantifying the results from qualitative texts, the following insights were finally discovered:

- **Users have a high dependence on external navigation and have a willing to improve the existing navigation:**
  Users mostly just check the parking area and scan the code when using the navigation within the Voi or Ryde apps. In fact, they still rely on Google Maps(14 in frequency) as an auxiliary navigation tool. At the same time, users are not highly satisfied with the current navigation and interaction status overall (30% of the evaluations are negative), mainly focusing on aspects such as distraction, inaccuracy, and lack of intuitiveness.
- **Speed affects the perception of safety and the sense of experience:**
  Whether it is in terms of braking, stability, or the convenience and comfort of the experience, the perception of speed needs to be further improved.

7.1.2 Results of Quantitative Research

7.1.2.1 Questionnaire

A total of 100 questionnaires were collected this time, resulting in a data matrix of 100 X 23. The obtained data were cleaned and normalized, and then the mean and mode of several key indicators were calculated(Table 7.2).

Table 7.2: Result of the 10 numerical variables

No.	Category	Average	Mode	Data Features
F1	Safety Perception (overall)	3.4	4	The data is left-skewed, with some extremely low scores. For example, scores of 1 point account for 5%.
F2	Safety Perception (attachments)	3.8	4	High scores are concentrated, and users generally attach great importance to safety functions.
F3	Safety Perception (handlebars)	3.5	3	The distribution is relatively uniform.
F4	Safety Perception (footrest)	3.3	3	There are quite a lot of low scores, which reflect problems regarding space or materials.
F5	Navigation and Interaction (phone holder)	3.1	3	The distribution is relatively uniform.
F6	Navigation and Interaction (audio)	3.4	3	There is a large difference in the preference scores. Some users consider it to be a source of interference.
F7	Navigation and Interaction (visual)	3.2	3	There are quite a lot of low scores, indicating concerns about distraction.
F8	Navigation and Interaction (vibratile)	3.0	3	Users have feedback that the sensitivity is insufficient.
F9	Future Improvement (software)	4.1	5	85% of the users are willing to try out new features.
F10	Future Improvement (hardware)	3.9	5	The acceptance level is high, but some users are concerned about hygiene issues.

The data normalization process was carried out for each of the above 10 indicators respectively. Then, a similarity matrix analysis was conducted on the data, and the results of the heat map are as follows:

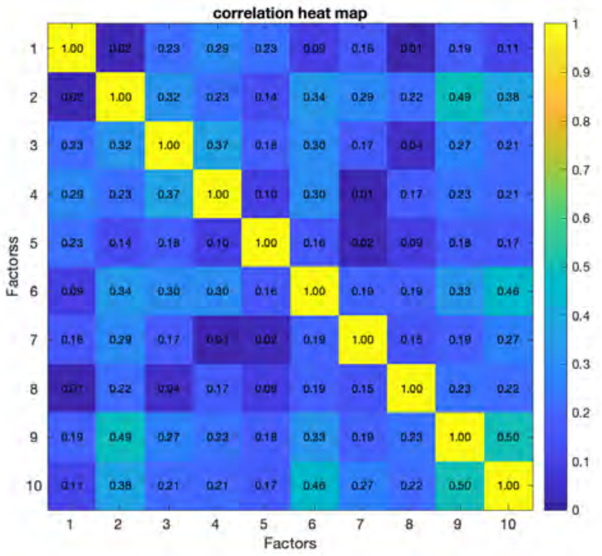


Figure 7.3: Similarity matrix of 10 factors that influence riding experience

Through the analysis and exploration of the data, the following Key insights and speculations were finally obtained:

- **People who have a higher acceptance of the audio interaction mode are more likely to accept a more comprehensive navigation system**
  There is a relatively strong positive correlation (r=0.49) between F6 and F10, and there is a relatively obvious positive correlation (r=0.33) between F6 and F9. This indicates that if users are more willing to accept audio as an interaction channel, they will also have a higher acceptance of the richness of software and hardware.

- **People who attach greater importance to safety have a higher degree of acceptance of advanced interaction systems**
  The correlation coefficient between F2 and variable F9 is r=0.49, and the correlation coefficient between F9 and F10 is r=0.50, indicating that there is a strong positive correlation among these variables. That is, when the value of one variable increases, the value of the other variable also tends to increase. This implies that people who attach more importance to this safety function are more willing to use the advanced interaction system. At the same time, currently the overall users' recognition of safety is relatively low (with an average score of 3.4, and 5% of the users gave the lowest score for safety), which indicates that users have a relatively strong willingness towards advanced interaction.

7.1.2.2 Text Quantitative Analysis

This study first conducted a word frequency analysis on approximately 4,900 pieces of text, and finally presented the word frequencies. Words without main meanings, such as subjects and prepositions, were removed. At the same time, the two topic keywords "scooter" and "e-scooter" were also removed. The 10 words with the highest occurrence frequencies were selected to find the correlations among them, as follows(Table 7.3):

Table 7.3: Result of word frequency analysis

No.	Words	Counts	Related Theme
W1	ride	34	Riding experience, distance, time
W2	safety	28	Safety, braking problems, accident risks
W3	navigation	22	Navigation system, map usage, direction guidance
W4	brake(s)	18	Defects in the braking system, emergency braking issues
W5	speed	15	Speed control, speed limit issues
W6	Google Maps	14	Dependence on navigation tools and functional defects
W7	cost	12	Rental fees, economic efficiency
W8	route	12	Route planning, navigation accuracy
W9	handlebar(s)	10	Handle design, grip comfort level
W10	phone holder	10	Practicality and design flaws of mobile phone holders

7.1.3 Usability Testing

The usability testing of this study is divided into two parts: a cognitive scale test for the new and old UI to analyze users' cognitive preferences for the new UI interface design; and an analysis of the interaction efficiency of multimodalities in different riding scenarios.

7.1.3.1 User Interface test

The design of the interface optimization scheme primarily involves A/B testing between the original Google Maps UI and the redesigned UI (Figure 7.4). In this study, 30 users were tested through online interviews, where they completed NASA-TLX scoring based on task contexts. The scale includes numerical ratings across six dimensions: such as mental demand, physical demand, temporal demand, etc. Each scale row has 20 intervals with a step of 5, and final scores are recorded on a 0 to 100 scale.

Figure 7.4: A/B test of designed interface and Google map UI

To analyze the NASA-TLX scores, a Wilcoxon test was conducted to validate effectiveness, while mean values were used to assess whether results trended positively. After each user completed the scale, brief interviews were conducted to gather feedback, yielding the following key insights:

- **Overall new UI shows positive, though not significantly (Table 7.4):** Users generally accepted some optimized elements, but certain design flaws persisted. For example, some users noted: "Distance and time information are redundant-distance should be deprioritized as non-critical and further optimized."

- **Usability issues in niche interactions (4/30 users):** For example, the left-hand placement of the exit button contradicted original ergonomic intentions, with users reporting discomfort. And for turning indications, the dynamic zoom feature during turns required a learning curve, confusing some riders during navigation.

The majority of users (25/30) provided positive feedback, stating that the new UI effectively reduced visual strain. However, significant individual differences were observed(can be found in Figure 7.5): users already familiar with navigation interactions tended to rate it as high as 10 or as low as 5, while those unfamiliar with such interactions generally scored it around 50.

Users largely recognized improvements in mental demand, performance, and effort(the difference of mean is more than 10 points).

Table 7.4: New UI vs Google Map UI: Statistics of 6 different aspects in NASA-TLX

Variables	New-UI		Old-UI		Significance
	Mean	Std	Mean	Std	
Mental Demand	38.50	16.67	43.83	18.55	$p = .212$, Non-Significant
Physical Demand	39.67	14.26	41.17	18.65	$p = .560$, Non-Significant
Temporal Demand	39.83	19.76	44.33	19.20	$p = .298$, Non-Significant
Performance	30.33	19.43	36.67	21.63	$p = .235$, Non-Significant
Effort	38.83	20.46	43.33	19.80	$p = .410$, Non-Significant
Frustration	37.83	19.99	41.17	22.00	$p = .593$, Non-Significant
Score	37.50	14.28	41.75	15.59	$p = .174$, Non-Significant

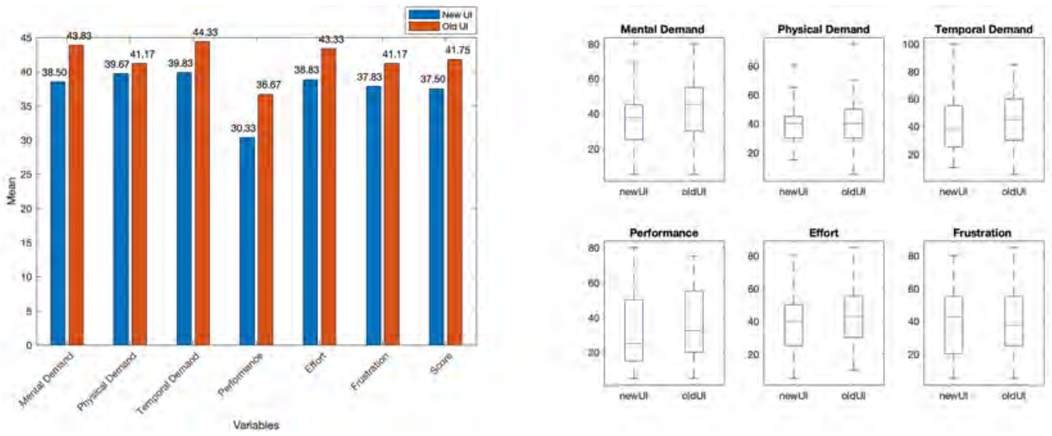


Figure 7.5: Difference of the NASA-TLX between the new UI and old Google UI in detail

7.1.3.2 Cognitive Load Test of interaction design

This study employed a cognitive load efficiency analysis formula to conduct usability analyses of task performance across different modalities. Users' cognitive efficiency in riding tasks consisted primarily of two components: *user performance* and *mental effort.*

In the end, this study invited a total of 30 participants to take part in the test, recording users' performance and mental effort:

- **Performance:** was measured by the time users took to complete the same distance.
- **Mental effort:** was quantified using the NASA-TLX subjective cognitive scale for riding tasks, with standard weighting applied consistently.

According to the calculations, users in the UI-only group spent significantly more time and had higher NASA TLX scores compared to other groups(Table 7.5). Additionally, the mean cognitive efficiency of this modality was the only one shown as negative across all groups, indicating that users generally perceived this interaction as inefficient. During interviews about their experiences with the UI-only modality, users commonly reported feedback such as, *"It's too troublesome to keep looking down and up,"* and *"I sometimes have to stop other tasks and focus on the UI, which is very inefficient."*

Table 7.5: Descriptive Statistics of Various Indicators

Variables	Mental(Nasa-tlx)		Performance(seconds)		Efficiency	
	Mean	Std	Mean	Std	Mean	Std
UI	59.98	10.32	219.83	143.90	-1.46	1.08
AR	36.00	17.16	110.00	103.06	0.72	1.11
Audio	36.00	17.16	130.13	110.73	0.22	1.02
UI+Audio	39.3	15.97	139.50	129.52	0.28	1.16
UI+AR	40.93	14.22	128.83	97.20	0.16	0.85
UI+AR+Audio	33.35	16.81	120.70	106.24	0.84	1.20

Note: Mental measures users' cognitive effort (assessed via the NASA-TLX scale); Performance refers to the total time (in seconds) users took to complete a fixed distance in a lab (VR) environment; Efficiency denotes the cognitive efficiency calculated from Mental and Performance, where positive/negative values indicate high/low efficiency, and the magnitude represents the degree of efficiency.

In addition, a higher E-value (efficiency) indicates more effective modality interaction: AR (E=0.72 ±1.11) and UI+AR+audio (E=0.84 ±1.20) significantly outperformed other groups, suggesting AR intervention can improve riding efficiency to some extent.

Regarding why the UI+AR groups efficiency value lagged far behind UI+AR+audio: testing and post-interviews revealed that most users only utilized AR+audio in the full-modality group, treating the UI as a psychological safety net for occasional route checks. When asked about preferred interaction modalities:

- 18/30 users favored the AR+audio combination, citing it as the most efficient. They noted that AR and audio complemented each other without interference. While most users(10/18) still saw value in UI during real-world riding, particularly for quick route overviews at the start or during stops.
- 8/30 users preferred UI+AR: some users(3/8) express that they like listening to music while riding and wanted audio as an optional feature. Others emphasized that ambient noise in real environments could hinder road condition awareness.
- 4/30 users preferred the full UI+AR+audio combination.

7.1.3.2.1 Performance First, the user's task performance was analyzed. According to the QQ plot (Figure 7.6), the data did not exhibit a significant normal distribution, so the Kruskal-Wallis H test (nonparametric test) was used. The results showed significant differences between the parameters. Subsequently, Wilcoxon rank-sum tests were conducted for pairwise comparisons of data between different groups. Finally, Bonferonni correction was applied for multiple comparisons.

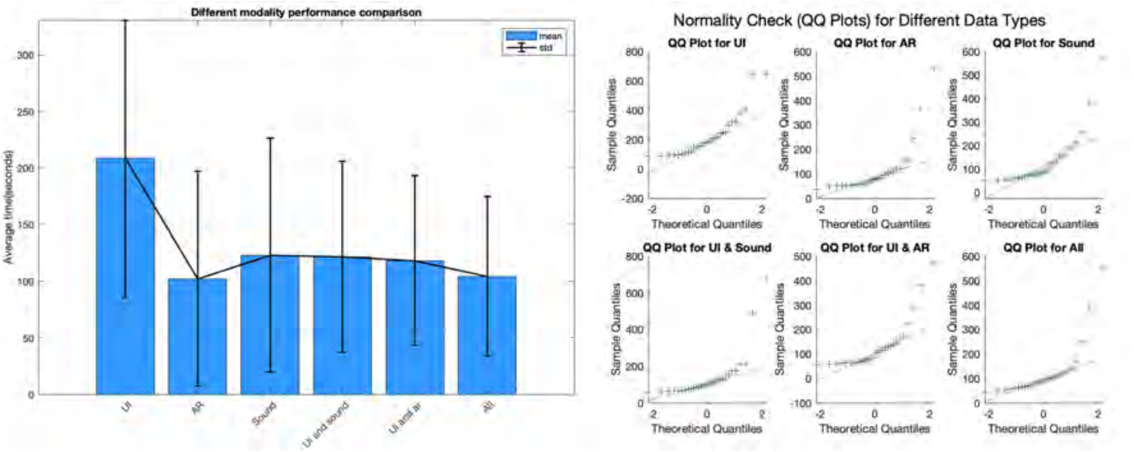


Figure 7.6: User Performance time(seconds) in different modalities and qq plot normality analysis

The final significance results show that the UI modality performed worst in terms of time (average duration: 219.83s), while other modalities averaged around 120s. The AR modality demonstrated the best time performance (average duration: 110s). Wilcoxon tests for each group revealed the following patterns(Table 7.6):

- The UI modality showed significant differences compared to all other modalities ($p < .001 < \alpha = .003$).
- The single AR modality significantly outperformed the audio modality ($Z = 2.921$, $p = .003$, $r = .533$), the UI+AR combination ($Z = 2.927$, $p = .004$, $r = .553$), and the UI+Audio combination ($Z = 2.778$, $p = .006$, $r = .507$). Insights from interviews and usability testing indicate that AR provides nearly instant visual feedback, whereas audio often requires waiting for the voice prompt to complete, introducing a delay in user guidance.
- The UI+Audio+AR combination relatively outperformed the UI+Audio combination ($Z = 2.397$, $p = .017$, $r = .438$), suggesting that AR as an additional modality enhances efficiency in time.

7.1.3.2.2 Mental Effort Regarding the measurement of mental effort, this study used the NASA-TLX scale (0-100, with users reporting values in steps of 5). Visual comparisons of the six dimensions across six groups were made via radar charts (Figure 7.7), yielding the following findings:

- The UI group showed the most mediocre performance in both mean scores and dimensional ratings, with an average user score of 60 (a score above 50 signifies a high mental workload [70]).
- The comprehensive mental effort scores of other groups ranged between 30 and 41, with relatively consistent performance. The all (UI+AR+Audio) group

Table 7.6: Performance in Significance Analysis Between Different Modalities with Wilcoxon's test

P-value	UI	AR	Audio	UI+AR	UI+Audio	ALL
UI	-	-	-	-	-	-
AR	$\boldsymbol{p < .001}$	-	-	-	-	-
Audio	$\boldsymbol{p < .001}$	$p = .004$	-	-	-	-
UI+AR	$\boldsymbol{p < .001}$	$\boldsymbol{p = .003}$	$p = .510$	-	-	-
UI+Audio	$\boldsymbol{p < .001}$	$p = .006$	$p = .153$	$p = .552$	-	-
ALL	$\boldsymbol{p < .001}$	$p = .245$	$p = .334$	$p = .371$	$p = .017$	-

Note: After Bonferroni correction, a total of 15 repeated comparisons were conducted, so the corrected significance level was set $\alpha = .003$.

performed best (weighted score: 33.35), indicating that the input of three modalities did not impose cognitive burdens on users. Instead, the ability to choose modalities enhanced convenience.

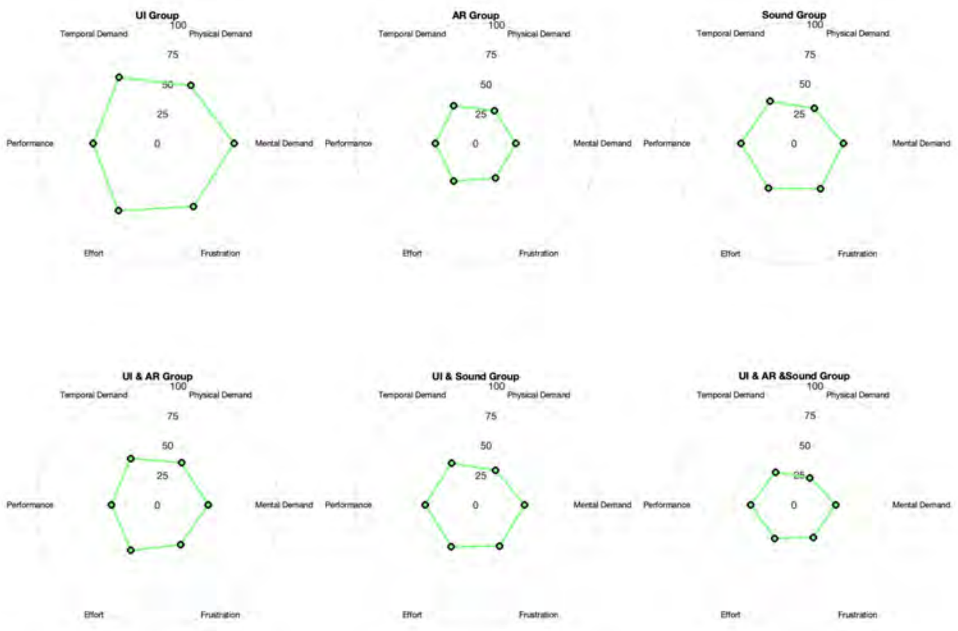


Figure 7.7: Radar Map visualization of Mental effort Analysis in different modalities

For the measurement of mental effort, this study directly employed Wilcoxon tests to conduct pairwise comparisons across different modalities to verify their significance(Table 7.7). The final results are as follows:

- The UI group's psychological cognitive performance remained significantly lower than that of other groups (p<.001). Through usability testing, when asked about the reasons, users universally reported that constantly looking down and up was troublesome, significantly creating "seams" in interaction behavior.

- Notably, the UI+AR group performed worse (weighted score: 40.93) than the AR group (Z = 2.386, p =.017, r= -.451), suggesting that the two visual input modalities (UI and AR) created some degree of conflict.

- The UI+audio+AR group significantly outperformed the audio group (Z = 3.498, p =.001, r= .661), the UI+AR group (Z = 2.964, p =.003, r= .550), and the UI+audio group (Z = 3.386, p =.001, r= .640).

Table 7.7: Significance Analysis Between Different Modalities (NASA TLX) with Wilcoxon's test

P-value	UI	AR	Audio	UI+AR	UI+Audio	ALL
UI	-	-	-	-	-	-
AR	$\boldsymbol{p < .001}$	-	-	-	-	-
Audio	$\boldsymbol{p < .001}$	$p = .050$	-	-	-	-
UI+AR	$\boldsymbol{p < .001}$	$p = .017$	$p = .869$	-	-	-
UI+Audio	$\boldsymbol{p < .001}$	$p = .327$	$p = .069$	$p = .308$	-	-
ALL	$\boldsymbol{p < .001}$	$p = .144$	$\boldsymbol{p = .001}$	$\boldsymbol{p = .003}$	$\boldsymbol{p = .001}$	-

Note: After Bonferroni correction, a total of 15 repeated comparisons were conducted, so the corrected significance level was set $\alpha = .003$.

7.1.3.2.3 Efficiency For the calculation of efficiency, this study first performed z-score normalization on performance and mental effort values to facilitate operations between different types of data. Finally, Wilcoxon tests were conducted on the calculated [1] "E values". for significance analysis.

To more clearly display the cognitive efficiency across different modalities, the data were normalized to an interval of [-3, 3], making it easier to observe data concentration (Figure 7.8). The results are as follows:

- The vast majority of users perceived the UI-only group as inefficient (E < 0), and the data points were far from the diagonal, indicating a high degree of inefficiency.

- The AR, UI+Audio, and all (UI+AR+audio) groups showed the best performance with generally concentrated data, suggesting these three modality combinations offer the highest cognitive efficiency and are suitable for scenarios requiring urgent responses.

- The UI+AR and audio-only groups exhibited more average distributions. Key influencing factors included the poor situational applicability of audio (users

---

[1] The E value refers to cognitive efficiency, a concept derived from Cognitive Load Theory. It is calculated as $E = \frac{Z(\text{performance}) - Z(\text{mental effort})}{\sqrt{2}}$, where Z(performance) is the z-score normalized completion time over a 2 km riding task, and Z(mental effort) is the z-score normalized NASA-TLX score reflecting perceived cognitive load across different interaction modalities.

easily missed information) and visual input conflicts in the UI+AR group, which caused certain visual stress for users.

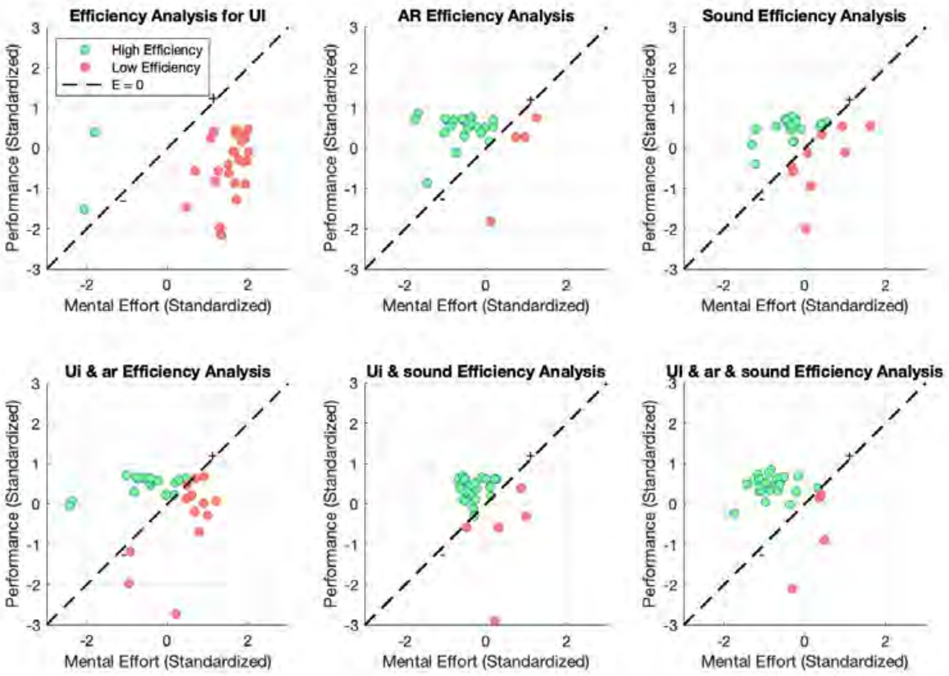


Figure 7.8: Scatter plot of Efficiency Analysis in different modalities

Finally, through comparative verification among different modalities, the following results were found (Table 7.8):

- The cognitive efficiency of the UI group was significantly lower than that of other modalities (p = .001).
- The cognitive efficiency of the AR group was significanly better than the UI + AR group ($Z = 3.445$, p =.001, r= .629). The main reasons were the timeliness of feedback in the visual channel and the redundancy of a single channel.
- The cognitive efficiency of the All group (audio + AR + UI) was significantly better than that of the audio group (Z = 2.952, p =.003, r= .539), the UI + AR group (Z = 3.260, p =.001, r= .595), and the UI + Audio group (Z = 3.137, p =.002, r= .573).

In summary, the study revealed that users of e-scooter navigation systems face significant challenges in safety perception (average score 3.4/5), navigation intuitiveness, and hardware design flaws (e.g., phone holder ergonomics). Qualitative interviews highlighted heavy reliance on external tools like Google Maps and dissatisfaction with current UIs due to distraction, inaccuracy, and monotonous interaction. Quantitative data showed strong correlations between safety prioritization and acceptance of advanced features. While A/B testing of a new UI showed no statistically significant reductions in NASA-TLX workload metrics compared to the old design,

Table 7.8: Cognitive Efficiency Significance Analysis between different modalities with Wilcoxon's test

P-value	UI	AR	Audio	UI+AR	UI+Audio	ALL
UI	-	-	-	-	-	-
AR	$\boldsymbol{p < .001}$	-	-	-	-	-
Audio	$\boldsymbol{p < .001}$	$p = .013$	-	-	-	-
UI+AR	$\boldsymbol{p < .001}$	$\boldsymbol{p = .001}$	$p = .572$	-	-	-
UI+Audio	$\boldsymbol{p < .001}$	$p = .026$	$p = .629$	$p = .478$	-	-
ALL	$\boldsymbol{p < .001}$	$p = .349$	$\boldsymbol{p = .003}$	$\boldsymbol{p = .001}$	$\boldsymbol{p = .002}$	-

Note After Bonferroni correction, a total of 15 repeated comparisons were conducted, so the corrected significance level was set $\alpha = .003$.

25/30 users reported reduced visual strain. Multimodal interaction testing found that AR + audio achieved the highest cognitive efficiency (E=0.84 ±1.20) and lowest mental workload (NASA-TLX=33.35 ±16.81), outperforming UI-only systems (E=-1.46 ±1.08), which caused frequent gaze-shifting frustrations. The full multimodal system (UI+AR+audio) balanced redundancy and efficiency, preferred by 18/30 of users for complementary feedback. Key recommendations include prioritizing AR and audio as core interaction modes, refining UI redundancies (e.g., distance displays), and integrating real-time safety alerts to address user concerns.

7.2 Final Design

The final design presentation is structured around five key dimensions of interaction design: Strategic, Scope, Interaction, Information, and Sensoryderived from the usability test findings. Each of these dimensions is discussed in the following subsections to provide a comprehensive analysis of the design outcomes.

7.2.1 Strategic Design

Through qualitative and quantitative research, the final product goals are positioned to design a user-friendly, usable, and accessible product, focusing on user safety and user experience through multimodal interaction navigation. The specific design strategies are as follows:

- **User Safety:** Optimize the cognitive efficiency of the UI interface, using UI as a safety guarantee for riding. Dynamically display prioritized information based on the users primary objectives in different scenarios.

- **User Experience:** Adopt AR as the primary interaction medium to enhance users cognitive efficiency (Figure 7.9).

Figure 7.9: AR projector position and functionality

Table 7.9: Multimodal Navigation Strategies Under Varying Risk Contexts

Scenario Type	**Environmental Features**	**User Needs**	**Multimodal Application**
Low-Risk	Familiar routes, straight roads, minimal traffic	Maintain speed, occasional direction checks	- AR projection for turns - Optional audio feedback for straight paths
Moderate-Risk	Urban main roads, intersections	Timely response to traffic changes, collision avoidance	- Dynamic AR arrows + audio countdown for turns - UI overview when braking
High-Risk	Rainy/night-time, unfamiliar routes, construction zones	High safety needs (e.g. high pedestrian flow), multiple confirmations	- AR + audio + UI alerts (red arrows + screen highlights)
Special Contexts	Noisy environments (e.g., near subway stations), hands occupied	Navigation without visual/auditory input	- AR projection for turns - Optional audio feedback

7.2.2 Scope Design

The final research targets two main user groups:

- First-time commuters/tourists unfamiliar with urban roads
- Frequent commuters facing route changes due to ongoing urban construction

Additionally, adhering to inclusive design principles, the multimodal options provide navigation feedback for users with single-modality impairments (e.g., hearing loss, myopia). Specific dynamic input modalities include (Figure 7.10):

- **Speed control:** Real-time recognition of cycling speed changes to adapt the riding interface and provide intuitive feedback.
- **Screen interaction:** Adjusted button placement to prevent physical barriers to touch inputs (e.g., inaccessible click areas).
- **Brake detection:** Identification of speed fluctuations to trigger brake signals, enhancing the comprehensiveness and reliability of the riding interface.

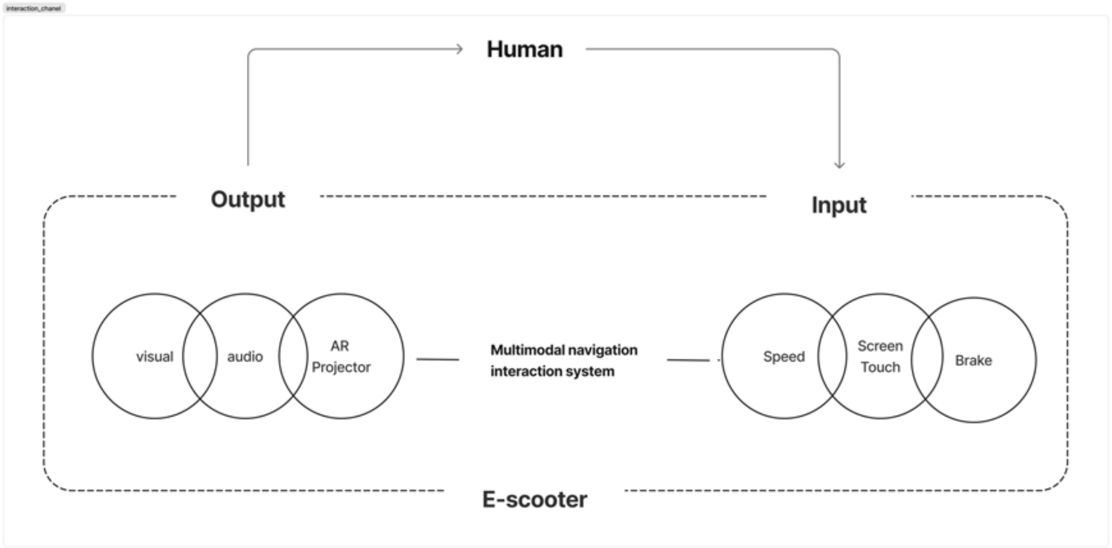


Figure 7.10: Input and output of the interaction design

7.2.3 Interaction Design

The details of Dynamic Modality Switching and corresponding interactions are as follows (also refer to Figure 7.11):

- **Accelerating:** Automatically reduce visual feedback to minimize distraction and rely primarily on AR guidance.
- **Braking:** Simultaneously highlight AR arrows and the UI interface, with repeated audio instructions to prevent missed cues during sudden stops. The UI should display detailed route information, including total remaining distance and distance to the next turn.

- **Decelerating:** Use AR and audio interaction as the primary channels. When approaching a turn, the UI should amplify route details to allow users to reconfirm the path visually.

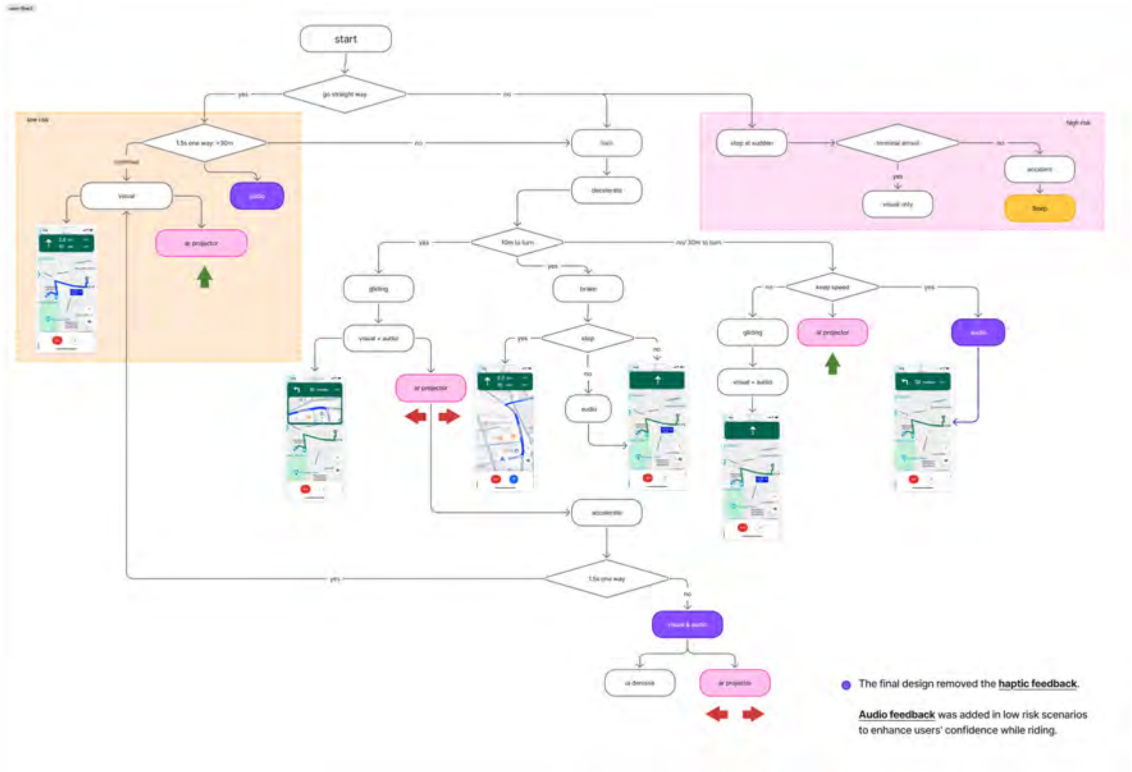


Figure 7.11: User flow and interaction design based on different scenarios

7.2.4 Information Design

For low-risk mode cycling scenarios, the final UI optimization removes street name details (addressing the contradiction of users unfamiliar with road names while needing navigation) and enlarges arrow indicators (Figure 7.12). Meanwhile, both distance (km) and time information are displayed. As an efficiency-focused scenario, the design prioritizes users primary needs by streamlining non-essential data.

Figure 7.12: High fidelity of the app: Low risk scenario

For medium-risk scenarios, the UI incorporates dynamic information prompts and behavioral distinctions (Figure 7.13). This layered design ensures the UI dynamically aligns with users mental workload and physical actions, balancing situational awareness and navigational efficiency.

- Deceleration Intent Analysis
  - Active Braking: Indicates high cognitive load; prompts users to pause and view holistic navigation details.
  - Gliding Deceleration: Signals ongoing mobility; the interface only enlarges directional cues for the forward path.
- Distance-Based Adaptation:
  - 10m Deceleration: Triggers enlarged road details (e.g., junction layouts) to support turn execution and improve task completion (Figure 7.14).
  - 30m Deceleration: Marks the threshold of low-risk driving; pre-alerts users of an upcoming turn to facilitate proactive adjustment.
- Behavior-Information Mapping:
  - Heavy braking correlates with overwhelming information load, necessitating simplified, high-priority UI (e.g., route summaries).

  - Gradual coasting allows for focused but non-intrusive guidance (e.g., enlarged arrows without full-map expansion).

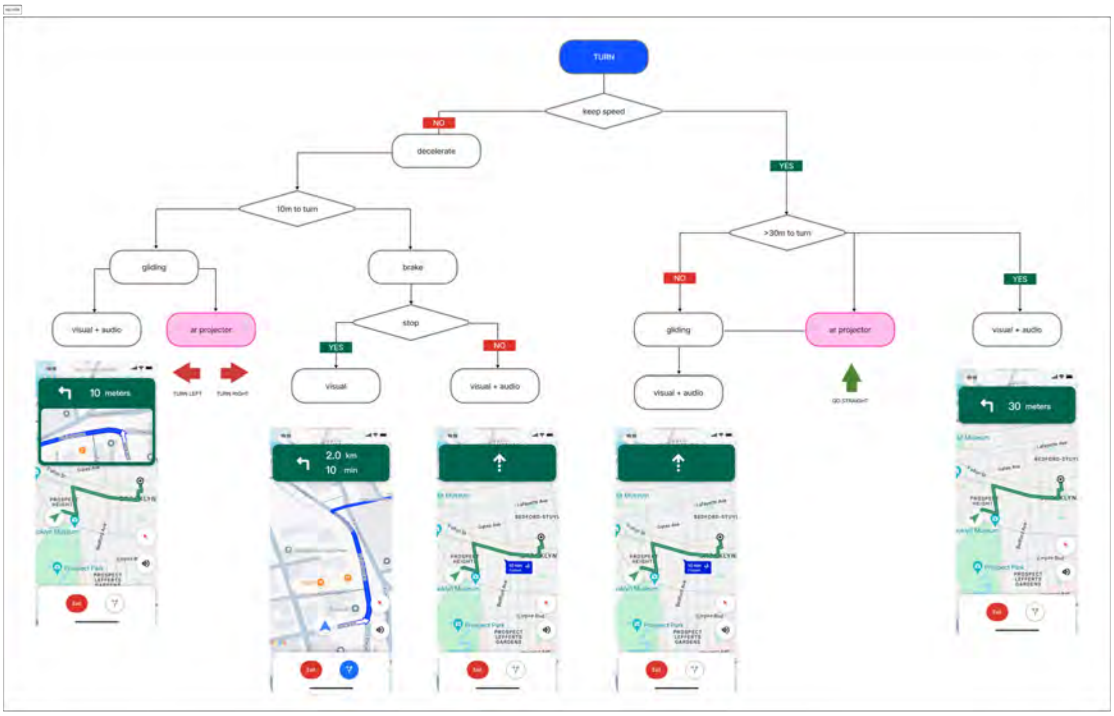


Figure 7.13: High fidelity of the app: Moderate risk scenario

Figure 7.14: High fidelity of the components: amplified route details

7.2.5 Sensory Design

The final iterative design deliverables include *pattern design* and *hardware design*. The hardware portion remains in the conceptual design phase, with key specifications as follows:

- Primary Materials: Foamed TPE/TPU for shock absorption and lightweight construction.
- Structure: Bolt-and-tubular nesting connections for modular assembly and adjustability(Figure 7.15).
- Projection Geometry: The projector is angled to optimize road coverage, with a raised lower section housing the battery and chip module.

Figure 7.15: Different views of the rendering of AR projector Model

Since this AR projector design does not actually involve hardware development, the conceptual design was carried out in a VR environment. The specific AR projection parameters are as follows:

- AR indicators will only be projected 3 meters before a turn, with no interaction displayed on other road sections, minimizing interference with the user's driving behavior.
- Each AR prompt will flash for 2 seconds and then stop.
- A green arrow will indicate a straight path, while a red arrow will indicate a turn (Figure 7.16).

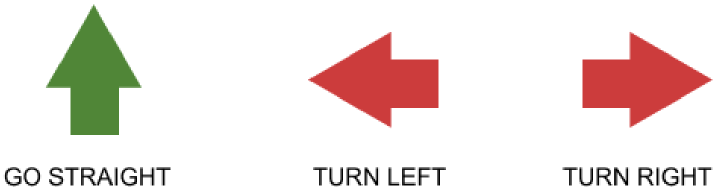


Figure 7.16: The pattern design of the AR projector

8

Discussion

8.1 Result Discussion

8.1.1 Design Result

The final prototype integrated multiple sensory channels into a cohesive navigation interface. Key features included:

- **Augmented Reality (AR) visual cues:** Directional arrows were overlaid onto the riders view (via a smartphone-based AR display), providing clear, eyes-on-road guidance for upcoming turns.
- **Auditory prompts:** General audio instructions signaled turn timing in sync with the AR arrows.
- **Map-based UI:** A smartphone app displayed an overview map and route; this screen-based UI was intended for occasional strategic checks rather than constant use.

Performance testing showed that this multimodal design substantially outperformed single-modality baselines. In controlled riding trials, the full AR+audio+UI prototype achieved the highest throughput, with an efficiency measure of E=0.84. This efficiency was significantly greater than that of any one- or two-modality condition. However, this improved performance came with a trade-off: riders in the full-modality condition reported slightly higher workload. In particular, NASA-TLX scores were marginally higher for the triple-modality system than for the best dual-modality setup, a difference confirmed by a Wilcoxon test ($p < .001$). Overall, the multimodal interface enabled faster navigation and lower perceived effort than the traditional smartphone-only system. Participants navigated faster and felt more at ease under AR+audio+UI conditions compared to single cues. In other words, combining sensory channels reduced mental load while improving speed and confidence.

In sum, riders found that an AR head-up display paired with audio cues provided a smooth, hands-free experience, with the map UI needed only intermittently. These results underscore that the most effective design centers on visual AR guidance augmented by audio prompts, which together enable rapid, eyes-on-road navigation. Such a multimodal combination dramatically improved task efficiency and reduced workload compared to conventional navigation (smartphone map only), thereby enhancing both objective performance and subjective safety.

8.1.2 Research Questions Result

8.1.2.1 Research Question 1

In what ways does multimodal interaction enhance both user safety and user experience in navigation systems for shared mobility?

The findings indicate that integrating multiple modalities (visual, auditory, augmented reality [AR]) significantly enhances navigation safety and user experience compared to a single-screen interface. In performance tests, riders using multimodal cues completed routes more quickly and with fewer errors than those relying on the user interface (UI) alone. For example, conditions with AR and audio produced the highest efficiency (Efficiency of 0.72 ±1.11 for AR alone and 0.84 ±1.20 for AR+audio). Statistical analysis confirmed these gains: pairwise Wilcoxon tests showed that the UI-only condition was significantly inferior across all comparisons ($p < .001$). Notably, two key patterns emerged from the data and user feedback.

First, channel redundancy may reduce cognitive efficiency, as users tend to rely primarily on one modality. This was evidenced by performance metrics: AR alone significantly outperformed the combined UI+AR condition (Efficiency: $Z = 3.445$, $p = .001$, $r = .629$), indicating that adding the UI did not improve outcomes and instead introduced redundant information. User feedback further supported this: one participant noted, When AR is accurate, I basically dont check the UI at alllooking at the UI would even be a waste of time. This suggests that redundant modalities do not complement each other but rather impose unnecessary cognitive effort, as users default to the more intuitive modality (AR) and disregard the secondary one (UI).

Second, complementary coordination of modalities yields significant performance and experience improvements. This was demonstrated by comparisons of multimodal combinations: the UI+AR+Audio condition significantly outperformed both UI+AR (Efficiency: $Z = 3.260$, $p = .001$, $r = .595$) and UI+Audio (Efficiency: $Z = 3.137$, $p = .002$, $r = .573$). Subjectively, users highlighted the synergistic value of non-redundant cues: a participant commented, I think the combination of AR and audio is excellent. The two modalities dont compete for attention, and if one is missed, the other can serve as a backup. This aligns with objective data, as the complementary blend of visual (AR), textual (UI), and auditory cues addressed different user needsAR for real-time environmental alignment, audio for glance-free confirmation, and UI for detailed referencewithout overloading cognitive resources.

In other words, users ride faster and more accurately when they receive guidance through AR displays and audio cues rather than solely looking at a phone screen. Subjectively, multimodal feedback also reduced cognitive workload: NASA-TLX scores were highest under the UI-only condition, whereas any condition involving AR or audio yielded significantly lower effort ratings. Wilcoxon analyses of TLX ratings confirmed that the UI-only condition was significantly more mentally demanding than all other conditions ($p < .001$).

In practice, riders reported that the AR projector pattern on the ground and audio prompts allowed them to keep their eyes on the road, instead of frequently glancing down at the screen. Most participants (18 out of 30) noted that the projector pattern for turns was easy to feel even without looking, and that the cues confirmed directions when I didnt want to take my eyes off traffic. In contrast, the traditional UI required taps and glances that interrupted riding. These observations reinforce the value of complementary modalities: by offloading navigation instructions to non-visual channels (audio) and context-integrated visual cues (AR), users maintained situational awarenessaddressing the inefficiency of redundant UI+AR while leveraging the strengths of complementary cues to reduce distraction.

Overall, multimodal navigation provided a more intuitive and efficient user experience. Riders reported greater confidence and comfort when guided by combined cues, with many remarking that the system felt natural and hands-free. These subjective reports align with objective metrics: riders in AR+audio conditions not only navigated faster but also felt more at ease. In sum, the multimodal systemby avoiding channel redundancy and prioritizing complementary modality coordination-improved performance and reduced mental load, thereby enhancing perceived safety and overall user satisfaction during shared-scooter navigation.

8.1.2.2 Research Question 2

Which modality is dominant when riding and in what contexts? What modality synergies can we construct from observation studies and what prototypes can we develop to facilitate multimodal interactions?

The results and user feedback consistently showed that AR (augmented reality) combined with auditory cues tends to dominate the navigation experience, while the visual UI serves as a backup or secondary reference. In practice, riders relied primarily on head-up AR arrows with audio prompts reinforcing the timing of turns. Among combinations, the AR+audio prototype was thought to be the most efficient; 18 out of 30 participants explicitly favored the AR+audio combination as most efficient, noting that these cues worked together without interfering. These riders explained that the AR arrow provided clear directional guidance, and the audio confirmatory cues arrived precisely when needed, allowing them to focus on riding.

Context influenced which modalities were used. In low-risk, straight-road scenarios, users tended to depend almost entirely on the AR visual cue to maintain course. In moderately complex contexts, they naturally augmented AR with audio signals to ensure timely turns. In high-risk or unfamiliar environments, riders sometimes briefly checked the UI map for an overview before resuming AR guidance. Post-test interviews revealed that many participants treated the UI as a safety net: they used it mainly at the beginning of a route or at major intersections to confirm the overall path, then switched their eyes back up once the AR cues resumed. For example, one user said, "I only looked at the map when I came to a stopotherwise the arrows and beeps handled everything." This pattern suggests that while AR and audio are the primary channels for dynamic guidance, the UI map provides an occasional strategic overview. Prototyping also reflected these findings. The fully integrated UI+AR+audio prototype achieved the highest efficiency in controlled testing (E=0.84 ±1.20), confirming that all three cues together are powerful. However, because participants rarely used the UI in motion, a leaner prototype combining only AR and audio may be optimal.

Indeed, the fact that no test condition isolated the AR+audio duo (due to our test matrix) is a limitation, but user preference data strongly imply its effectiveness. We conclude that an ideal multimodal navigation design would center on an AR display (for on-road directions) with synchronized audio prompts, reserving the screen-based UI for situational checks rather than continuous use. Such a combination was consistently reported to best support a smooth, eyes-on-road navigation experience, especially in busy or complex riding contexts.

8.1.2.3 Research Question 3

Would the number of navigation output modalities during riding affect users' safety and user experience?

The findings reveal that the quantity of modalities itself is not the decisive factorinstead, two interrelated variables drive outcomes: the learning curve and operational cost of each modality, and the relational nature of modalities (whether they complement or conflict with one another).

The impact of adding more modalities is mixed, reflecting a trade-off between information richness and cognitive load, but this trade-off is shaped by the two key variables above. In this study, introducing a second modality (e.g., AR+audio or UI+AR) generally improved performance and reduced perceived workload compared to single-modality conditions. For instance, dual-modality conditions outperformed the UI-only baseline on both speed and accuracy. However, adding a third modality (the full UI+AR+audio combination) provided diminishing returns and introduced a small cognitive penaltyoutcomes rooted in how users engage with each modalitys learning demands and how modalities interact.

First, the learning curve and operational cost of individual modalities directly influence safety and experience, far more than sheer modality count. This is most evident in comparisons involving the UI: the UI-only condition exhibited significantly lower efficiency (encompassing both performance and mental effort) relative to all other conditions ($p < .001$), driven by its inherently higher learning and operational costs. Unlike audio cues (which require no visual attention) or AR indicators (projected into the users line of sight), the UI demands explicit, disruptive actions: users must lower their gaze or even pause riding to locate, interpret, and interact with on-screen mapsbehaviors that break situational awareness and increase safety risks. In contrast, audio and AR modalities have minimal learning curves: their intuitive, context-integrated design (e.g., directional audio prompts, ground-projected AR arrows) allows users to process information without altering their riding posture or focus, reducing cognitive and physical burdens. This explains why even single-modality AR or audio outperformed the UI-only condition: modality quantity matters less than how easily users can adopt and operate each channel during dynamic riding.

Second, the relational dynamic between modalitieswhether they complement or conflictdetermines whether adding a modality enhances or undermines outcomes. Complementary modalities (e.g., AR + audio) synergize to improve performance: AR provides visual, environment-aligned directional cues (e.g., turn arrows on the ground), while audio offers glance-free confirmation (e.g., Turn right in 50 meters), and neither competes for the same cognitive or attentional resources. This non-conflicting design explains why dual-modality AR+audio struck the best balance between efficiency and workloadusers reported it felt seamless and low-effort, as the two channels reinforced each other without overload. In contrast, conflicting modalities (e.g., AR + UI) introduce redundancy and cognitive friction: AR, as the more intuitive, context-integrated channel, became the primary reference for users, while the UI required additional visual and attentional effort to processeffort that felt unnecessary and distracting. This conflict is reflected in performance data: AR alone significantly outperformed the UI+AR dual-modality condition, confirming that adding a conflicting modality does not enhance safety or experience, even if it increases the total number of output channels.

The full three-modality condition (UI+AR+audio) illustrates how these two variables interact. It achieved the highest efficiency (E=0.84 ±1.20) of all groups, significantly higher than any one- or two-modality condition, yet this came with slightly higher NASA-TLX ratings than the best two-modality set: riders in the full condition reported marginally more effort than those in the AR+audio condition (Wilcoxon $p = .001$ for UI+Audio vs. all modalities). The higher workload stemmed from both the UIs persistent operational cost (needing to ignore or occasionally check a redundant screen) and the initial learning curve of managing three concurrent signals. In practical terms, most users learned to manage the extra information, but some noted that initially it felt like a lot to take in. Once adapted, they acknowledged that the combined cues made navigation fasterthe system allowed quick corrections without visual confirmationalbeit at the cost of needing to interpret more signals simultaneously. These results align with an inverted-U effect: performance improves as modalities are added, but only if those modalities have low learning/operational costs and complementary relationships; beyond that point, conflicting channels or high-cost modalities (like the UI) cause cognitive demands to offset gains.

Participants implicitly recognized this balance. Although the full UI+AR+audio prototype had the best measured throughput, only 4 of 30 participants ultimately preferred it as their top choice, whereas many were satisfied with AR+audio alone. Several riders mentioned that the UI screen felt redundant when AR and audio were already activeechoing the conflict between AR and UIand described a learning curve with multiple outputs: At first I was confused by all three signals, but by the second ride it felt much smoother, one participant said.

In summary, more simultaneous modalities can enhance efficiency, but they also demand more cognitive processingespecially during initial learningonly if those modalities carry high operational costs or conflict with one another.

8.2 Process Discussion

8.2.1 Research Process

The study adopted a user-centered design framework with iterative interviews and testing, integrating qualitative and quantitative analysis methods while deeply examining user interaction scenarios. This thesis demonstrates strong extensibility at the design research level, offering both depth and research significance from the perspectives of data analysis and design insights. However, several methodological limitations require discussion.

First, participant sampling was restricted: the interview phase included 12 riders, predominantly university-aged, which may not fully represent the diversity of real-world scooter users (e.g., older adults, non-urban commuters). Consequently, the findings reflect a specific demographic.

Second, due to safety concerns, all testing was conducted on a closed path or simulated environment rather than open roads, lacking dynamic factors such as moving vehicles, pedestrians, and complex signage. These elements could unpredictably interact with the interfacefor instance, loud street noise potentially masking audio prompts.

Finally, regarding participant selection, testing was conducted in a campus environment due to time constraints, meaning most data reflect the mental models of 18 to 30 years old. While this group is a primary user demographic for scooters, broader research implications are limited.

These limitations are acknowledged: future research should incorporate high-fidelity traffic simulations or staged field trials to evaluate multimodal navigation under realistic cognitive loads. Despite these constraints, the research process effectively isolated modality effects in a repeatable manner and aided in identifying core user needs.

8.2.2 Concept Process

Our concept development centered on static AR overlays (e.g. smartphone-augmented arrow) and conventional app/phone UI. A key limitation was that the AR prototype did not fully emulate a wearable head-up display; riders had to balance a device rather than see cues directly in their line of sight. Users noted this issue: one commented that having to point the phone at the road was less natural than a true HUD. The static AR design also lacked depth and integrationfor example, it could not dynamically highlight moving objects or adapt to changing lighting. A dynamic AR approach (such as a windshield HUD or AR glasses) would better align cues with the riders gaze.

What's more, the prototypes themselves were relatively low-fidelity: visual designs were wireframes and static mock-ups, and the audio prompts were generic turn signals with the young female soft tone(such as turn left and turn right). While sufficient for initial testing, these did not capture real-life complexity (e.g. ambient audio interference, GPS jitter).

In future iterations, adding contextual audio (like turn in 30 meters) and more realistic AR visuals (e.g. semi-transparent overlays that adjust with head movement) could enhance immersion and fidelity. Prototype realism is crucial: participants tend to respond differently to high-fidelity systems. Thus, the current concepts should be seen as functional approximations; they demonstrated proof-of-concept, but a fully integrated hardware prototype would be needed to validate the design under authentic conditions.

8.2.3 Usability Test Process

Overall, the usability testing conducted in the VR-simulated environment and the comparative lab data were scientifically rigorous. By using the cognitive efficiency formula and taking performance and mental effort as metrics to measure users' perceptual efficiency, followed by non-parametric comparison of the data, this thesis has strong reference value in terms of methodology, both academically and practically. However, several issues require further consideration:

First, the experimental environment was overly controlled. Users rode on a simplified, low-traffic route without real-world hazards, which reduced cognitive load compared to actual urban riding. Consequently, the measured benefits of multimodal cues may be underestimated; in real traffic with distractions, the advantages (or potential drawbacks) of the design could be more pronounced.

Second, vibration feedback was not incorporated into the testing**, limiting the exploration of more diverse modalities. Future research could conduct in-depth comparative studies on scenarios with vibration cues to enhance the comprehensiveness of multimodal design evaluations.

Additionally, consideration for users with disabilities was insufficient: Some participants experienced motion sickness from the VR equipment, leading to obvious outliers in the data. Color-blind users faced difficulties distinguishing red/green indicators for AR turns, highlighting the need for color-agnostic visual coding (e.g., shape or pattern differentiation).

Finally, the A/B test of the old and new UI was based on static interface comparisons. If future research uses interactive prototypes directly, the data would more accurately reflect real-world usability, as dynamic interactions may reveal subtle issues (e.g., latency, gesture recognition errors) not captured in static evaluations.

These limitations suggest avenues for refinement, but the studys systematic approach to isolating modality effects and integrating quantitative/qualitative data remains a valuable foundation for advancing multimodal navigation design.

8.3 Generalizability

The generalizability of this studys findings is explored through three complementary perspectives: qualitative analysis, quantitative analysis, and Research Through Design (RTD). These approaches collectively strengthen the applicability of the results to broader contexts and user populations while acknowledging methodological limitations. Each dimension contributes distinctively to evaluating how insights from this study can inform broader contexts in micromobility navigation systems.

8.3.1 Quantative Research

Quantitative results were derived from controlled usability tests (N=30) and an online survey (N=100), focusing on NASA-TLX workload scores, riding efficiency, and modality preferences. While the experimental sample was not population-representative in a strict statistical sense, the use of randomized modality conditions and nonparametric tests (Wilcoxon, Kruskal-Wallis) strengthens the internal validity of observed effects.

Significant differences were consistently found between UI-only and AR+audio conditions across all cognitive metrics (e.g., E-values, NASA-TLX scores). These results offer evidence of functional generalizability: the cognitive benefits of combining AR and auditory cues for navigation likely extend beyond the studys exact interface. However, because the tests were conducted in a simulated VR environment, extrapolation to real-world physical settings should be made cautiously. Real-world factors like noise, lighting, weather, and traffic behavior could moderate the observed advantages of multimodal interaction.

8.3.2 Qualititive Research

The qualitative phase of the research - comprising interviews, first-person perspective, and field observations - offered a rich, contextualized understanding of rider behavior, pain points, and expectations. While the sample size was intentionally small(12 interviewees) and stratified to represent diverse user types (e.g., tourists, commuters, non-users), it supported analytical generalization rather than statistical inference. Key themes such as cognitive overload caused by screen dependence, the need for hands-free navigation, and uncertainty in unfamiliar routes appeared across different user types and urban environments, suggesting that these findings are transferable to similar shared mobility systems in comparable cities. Nevertheless, qualitative responses may carry bias due to users prior experience with technology or personal preference.

This aligns with Lincoln and Gubas concept of transferability in qualitative research [71], where the depth and clarity of contextual description allow readers to judge applicability to other settings. Although direct replication is not assumed, the behavioral patterns and interaction needs documented here can inform design for other shared micromobility platforms in urban European contexts.

8.3.3 Research through Design

From an RtD perspective, the studys prototypes and interaction strategies represent design knowledge that can be reused or reinterpreted by future researchers and practitioners. This includes the dynamic modality-switching framework (e.g., braking triggers full UI display) and scenario-specific mappings of sensory feedback (e.g., AR turn indicators for medium-risk zones).

Such outputs are not meant to be universally prescriptive but rather generative, providing templates and heuristics that inspire adaptations in other contextssuch as cycling interfaces, AR pedestrian guidance, or ride-hailing scooter systems. However, further parameter adjustment and modality optimization are required based on speed. The modularity of the final design enables selective application: systems may adopt the AR projection logic without the audio layer, or apply distance-triggered UI scaling independently.

The conceptual contributions (e.g., modality synergy frameworks, cognitive efficiency evaluation) extend the design space for human-mobility interaction. Thus, generalizability in this context is best understood as design transposability, where results enrich the broader discourse of multimodal navigation rather than claim universal coverage.

8.4 Future Work

While this study offers valuable insights into multimodal interaction for shared e-scooter navigation, several limitations should be acknowledged that may affect the generalizability and depth of the findings.

First, the study relied on the NASA Task Load Index (NASA-TLX) to evaluate cognitive load across different modalities. However, the standard weightings used in NASA-TLX may not fully capture the unique characteristics of dynamic navigation tasks like e-scooter riding. Since riding involves continuous motion and real-time decision-making, the relative importance of mental and physical load might deviate from typical use cases in static environments such as computer-based tasks.

Second, although augmented reality (AR) was incorporated into the system design and evaluation, the implementation emphasized relatively static AR interactions (e.g., route indicators and heads-up displays). In reality, riding an e-scooter is a dynamic and context-sensitive activity, often involving rapid environmental changes, obstacles, and unpredictable user behavior. Future research should explore more adaptive and real-time AR feedback mechanisms that adjust according to movement, environment, and rider intent.

Third, one major limitation was the omission of testing the AR+audio modality combination in the usability study. Despite being mentioned by 18 out of 30 users during the qualitative interviews as their most preferred interaction style, this condition was not included in the final usability test protocol. This omission may have restricted the findings regarding modality synergy and user preference. A more comprehensive testing matrix that includes high-potential combinations like AR+audio could lead to more holistic conclusions about optimal modality integration.

Lastly, while the controlled test environment allowed for consistency and reliability in cognitive load and performance measurement, it may not fully replicate the complexities of real-world navigation. Variables such as weather conditions, pedestrian density, and noise levels can significantly influence interaction effectiveness and should be considered in future field studies.

Future work should focus on:

- Incorporating adaptive and context-aware AR systems,
- Developing weighted evaluation frameworks tailored to mobile multimodal interactions,
- Including all major modality combinations in usability testing (e.g., AR+audio),
- And conducting real-world longitudinal studies to examine user learning curves, behavioral adaptation, and long-term usability.

Addressing these limitations will not only strengthen the design and evaluation of multimodal navigation systems but also enhance their real-world applicability in micromobility solutions.

9

Conclusion

This thesis addresses the critical need for safer and more intuitive navigation systems in shared electric scooters by exploring multimodal interaction design. Through a systematic blend of user-centered research, iterative prototyping, and empirical testing, the study contributes novel insights into how combining visual, auditory, and augmented reality (AR) modalities can enhance both safety and user experience in micromobility navigation.

A set of core findings emerged from the empirical evaluation, clarifying the key drivers of effective multimodal navigation-not the sheer quantity of modalities, but two interrelated factors. First, channel redundancy may reduce cognitive efficiency, as users tend to rely primarily on one modality when faced with overlapping information. This was confirmed by performance data: AR alone significantly outperformed the dual-modality UI+AR configuration (Efficiency: $Z = 3.445$, $p = .001$, $r = .629$), as the UI introduced redundant visual demands that users ultimately disregarded. Second, complementary coordination between modalities yields significant improvements in performance and experience. Statistical comparisons showed the triple-modality UI+AR+Audio condition outperformed both UI+AR (Efficiency: $Z = 3.260$, $p = .001$, $r = .595$) and UI+Audio (Efficiency: $Z = 3.137$, $p = .002$, $r = .573$), as AR (ground-projected directional cues) and audio (glance-free prompts) reinforced each other without competing for attention. Together, these findings underscore that modality quantity is not decisivesuccess depends on avoiding redundancy and leveraging complementary relationships between low-cost, intuitive channels.

The research systematically evaluated single, dual, and triple-modality configurations, revealing distinct advantages of multimodal approaches over unimodal interfaces. Quantitative results from usability tests demonstrated that the UI-only condition exhibited the lowest cognitive efficiency (E = -1.46 ±1.08) and highest mental workload (NASA-TLX = 59.98 ±10.32), primarily due to frequent gaze shifts between the road and screen. In contrast, the AR+audio combination (and the full UI+AR+audio condition) achieved significantly higher efficiency (E = 0.84 ±1.20 for the full condition) and lower mental effort (NASA-TLX = 33.35 ±16.81), as users could maintain situational awareness through head-up AR projections and auditory prompts. Qualitative feedback reinforced these findings, with 18 out of 30 participants prioritizing AR+audio for its seamless integration and reduced distraction, highlighting that multimodal cues allow riders to stay focused on the road while receiving clear directional guidance.

Future work should explore higher-fidelity implementations (e.g., smartglasses with dynamic overlays) and conduct prolonged field studies to assess long-term use and safety. Nonetheless, this research lays a foundation for next-generation micromobility navigation tools by showing that thoughtfully combined modalitiesgrounded in minimizing redundancy and maximizing complementaritycan significantly enhance rider experience and safety.

Bibliography

[1] J. Cubells, C. Miralles-Guasch, and O. Marquet, “E-scooter and bike-share route choice and detours: Modelling the influence of built environment and sociodemographic factors,” *Journal of transport geography*, vol. 111, p. 103 664, 2023.

[2] A. Nikiforiadis, E. Paschalidis, N. Stamatiadis, A. Raptopoulou, A. Kostareli, and S. Basbas, “Analysis of attitudes and engagement of shared e-scooter users,” *Transportation research part D: transport and environment*, vol. 94, p. 102 790, 2021.

[3] J. C. Lock, G. Cielniak, and N. Bellotto, “A portable navigation system with an adaptive multimodal interface for the blind,” in *2017 AAAI Spring Symposium Series*, 2017.

[4] V. Nair, M. Budhai, G. Olmschenk, W. H. Seiple, and Z. Zhu, “Assist: Personalized indoor navigation via multimodal sensors and high-level semantic information,” in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 0–0.

[5] J. van der Bie, S. Ben Allouch, and C. Jaschinski, “Communicating multimodal wayfinding messages for visually impaired people via wearables,” in *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2019, pp. 1–7.

[6] A. H. Basori, “Hapar: Handy intelligent multimodal haptic and audio-based mobile ar navigation for the visually impaired,” *Technological Trends in Improved Mobility of the Visually Impaired*, pp. 319–334, 2020.

[7] N. A. Giudice, B. A. Guenther, N. A. Jensen, and K. N. Haase, “Cognitive mapping without vision: Comparing wayfinding performance after learning from digital touchscreen-based multimodal maps vs. embossed tactile overlays,” *Frontiers in Human Neuroscience*, vol. 14, p. 87, 2020.

[8] A. Matviienko, F. Müller, D. Schön, *et al.*, “E-scootar: Exploring unimodal warnings for e-scooter riders in augmented reality,” in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 2022, pp. 1–7.

[9] N. Miao, M. Hou, X. Hong, G. Wang, X. Peng, and G. Ren, “Enhancing the cyclist traffic safety by multimodel interaction design with wearable haptic devices and optical see-through head-mounted displays,” in *2022 IEEE 4th Eurasia Conference on IOT, Communication and Engineering (ECICE)*, IEEE, 2022, pp. 286–289.

[10] M. Tabatabaie, S. He, H. Wang, and K. G. Shin, “Beyond" taming electric scooters": Disentangling understandings of micromobility naturalistic riding,”

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 8, no. 3, pp. 1–24, 2024.

[11] Y. Wu, P. Zhang, M. Gu, J. Zheng, and X. Bai, “Embodied navigation with multi-modal information: A survey from tasks to methodology,” *Information Fusion*, p. 102 532, 2024.

[12] K. Kazemzadeh, “Assessing e-scooter rider safety perceptions in shared spaces: Evidence from a video experiment in sweden,” *Accident Analysis & Prevention*, vol. 211, p. 107 874, 2025.

[13] M. Turk, “Multimodal interaction: A review,” *Pattern recognition letters*, vol. 36, pp. 189–195, 2014.

[14] A. Badeau, C. Carman, M. Newman, J. Steenblik, M. Carlson, and T. Madsen, “Emergency department visits for electric scooter-related injuries after introduction of an urban rental program,” *The American journal of emergency medicine*, vol. 37, no. 8, pp. 1531–1533, 2019.

[15] S. Oviatt, “Multimodal interfaces,” *The human-computer interaction handbook*, pp. 439–458, 2007.

[16] A. Bergh, A. Mehic, D. Sandberg, and J. Wernberg, “E-scooters and traffic accidents: Evidence from staggered roll-out in swedish municipalities,”

[17] International Transport Forum, *Transport innovation for sustainable development*, Summit Report, Available at: `https://www.itf-oecd.org/itf-2020-summit-transport-innovation-sustainable-development`, 2020.

[18] G. Peci and S. Ali, “Usage patterns and environmental effect analysis of e-scooter sharing system: A case study in gothenburg, sweden,” 2022.

[19] Transportstyrelsen, *Elsparkcykel*, Retrieved from: `https://www.transportstyrelsen.se/elsparkcykel`, Accessed: 2021-12-31, 2021.

[20] K. Kamphuis and I. Van Schagen, “E-scooters in europe: Legal status, usage and safety: Results of a survey in fersi countries,” 2020.

[21] H. Stigson and M. Klingegård, *Kartläggning av olyckor med elsparkcyklar och hur olyckorna kan förhindras*, 2020.

[22] H. Bunt, *Multimodal Human-Computer Communication: Systems, Techniques, and Experiments*. Springer, 1998.

[23] B. Dumas, D. Lalanne, and S. Oviatt, “Multimodal interfaces: A survey of principles, models and frameworks,” in *Human machine interaction: Research results of the mmi program*, Springer, 2009, pp. 3–26.

[24] S. Oviatt, “Ten myths of multimodal interaction,” *Communications of the ACM*, vol. 42, no. 11, pp. 74–81, 1999.

[25] S. Oviatt, P. Cohen, L. Wu, *et al.*, “Designing the user interface for multimodal speech and pen-based gesture applications: State-of-the-art systems and future research directions,” *Human-computer interaction*, vol. 15, no. 4, pp. 263–322, 2000.

[26] Q. Ma, H. Yang, A. Mayhue, Y. Sun, Z. Huang, and Y. Ma, “E-scooter safety: The riding risk analysis based on mobile sensing data,” *Accident Analysis & Prevention*, vol. 151, p. 105 954, 2021.

[27] E. Dritsas, M. Trigka, C. Troussas, and P. Mylonas, “Multimodal interaction, interfaces, and communication: A survey,” *Multimodal Technologies and Interaction*, vol. 9, no. 1, p. 6, 2025.

[28] C. Akalya devi and D. Karthika Renuka, “Multimodal emotion recognition framework using a decision-level fusion and feature-level fusion approach,” *IETE Journal of Research*, vol. 69, no. 12, pp. 8909–8920, 2023.

[29] C. Zhang, Z. Yang, X. He, and L. Deng, “Multimodal intelligence: Representation learning, information fusion, and applications,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 3, pp. 478–493, 2020.

[30] B. Hu, L. Xu, J. Moon, N. J. Yadwadkar, and A. Akella, “Mosel: Inference serving using dynamic modality selection,” *arXiv preprint arXiv:2310.18481*, 2023.

[31] J. C. Kim, T. H. Laine, and C. Åhlund, “Multimodal interaction systems based on internet of things and augmented reality: A systematic literature review,” *Applied Sciences*, vol. 11, no. 4, p. 1738, 2021.

[32] I. Rakkolainen, A. Farooq, J. Kangas, *et al.*, “Technologies for multimodal interaction in extended realitya scoping review,” *Multimodal Technologies and Interaction*, vol. 5, no. 12, p. 81, 2021.

[33] R. Blanco-Gonzalo, O. Miguel-Hurtado, C. Lunerti, *et al.*, “Biometric systems interaction assessment: The state of the art,” *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 5, pp. 397–410, 2019.

[34] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain computer interfaces, a review,” *sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[35] A. Brügger, K.-F. Richter, and S. I. Fabrikant, “How does navigation system behavior influence human behavior?” *Cognitive research: principles and implications*, vol. 4, no. 1, p. 5, 2019.

[36] P. Anderson, A. Chang, D. S. Chaplot, *et al.*, “On evaluation of embodied navigation agents,” *arXiv preprint arXiv:1807.06757*, 2018.

[37] Y. Zhu, R. Mottaghi, E. Kolve, *et al.*, “Target-driven visual navigation in indoor scenes using deep reinforcement learning,” in *2017 IEEE international conference on robotics and automation (ICRA)*, IEEE, 2017, pp. 3357–3364.

[38] T. Vainio, “Exploring cues and rhythm for designing multimodal tools to support mobile users in wayfinding,” in *CHI’09 Extended Abstracts on Human Factors in Computing Systems*, 2009, pp. 3715–3720.

[39] J. Ducasse, A. M. Brock, and C. Jouffrais, “Accessible interactive maps for visually impaired users,” *Mobility of Visually Impaired People: Fundamentals and ICT Assistive Technologies*, pp. 537–584, 2018.

[40] J. Baus, A. Krüger, and W. Wahlster, “A resource-adaptive mobile navigation system,” in *Proceedings of the 7th international conference on Intelligent user interfaces*, 2002, pp. 15–22.

[41] M. ucha, E. Drimlová, K. Reka, *et al.*, “E-scooter riders and pedestrians: Attitudes and interactions in five countries,” *Heliyon*, vol. 9, no. 4, 2023.

[42] E. Salinas, R. Cueva, and F. Paz, “A systematic review of user-centered design techniques,” in *Design, User Experience, and Usability. Interaction Design: 9th International Conference, DUXU 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I 22*, Springer, 2020, pp. 253–267.

[43] R. Gong and M. Hua, “Designing multimodal user interfaces for hybrid collaboration: A user-centered approach,” in *International Conference on Human-Computer Interaction*, Springer, 2023, pp. 67–82.

[44] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. Van Gerven, “Cognitive load measurement as a means to advance cognitive load theory,” in *Cognitive load theory*, Routledge, 2016, pp. 63–71.

[45] J. L. Plass, R. Moreno, and R. Brünken, “Cognitive load theory,” 2010.

[46] P. A. Kirschner, *Cognitive load theory: Implications of cognitive load theory on the design of learning*, 2002.

[47] L. M. Reeves, J. Lai, J. A. Larson, *et al.*, “Guidelines for multimodal user interface design,” *Communications of the ACM*, vol. 47, no. 1, pp. 57–59, 2004.

[48] M. Augstein and T. Neumayr, “A human-centered taxonomy of interaction modalities and devices,” *Interacting with Computers*, vol. 31, no. 1, pp. 27–58, 2019.

[49] F. G. Paas and J. J. Van Merriënboer, “Instructional control of cognitive load in the training of complex cognitive tasks,” *Educational psychology review*, vol. 6, pp. 351–371, 1994.

[50] J. Sweller, J. J. Van Merrienboer, and F. G. Paas, “Cognitive architecture and instructional design,” *Educational psychology review*, vol. 10, pp. 251–296, 1998.

[51] S. G. Hart, “Nasa-task load index (nasa-tlx); 20 years later,” in *Proceedings of the human factors and ergonomics society annual meeting*, Sage publications Sage CA: Los Angeles, CA, vol. 50, 2006, pp. 904–908.

[52] W. Gaver, “What should we expect from research through design?” In *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, pp. 937–946.

[53] A. M. Rugman and J. R. D’cruz, “The" double diamond" model of international competitiveness: The canadian experience,” *MIR: Management International Review*, pp. 17–39, 1993.

[54] Design Council, *The double diamond*, Online, Accessed: 2023-02-28, 2023. [Online]. Available: `https://www.designcouncil.org.uk/our-resources/the-double-diamond/`.

[55] A. Lucero, “Using affinity diagrams to evaluate interactive prototypes,” in *Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015, Proceedings, Part II 15*, Springer, 2015, pp. 231–248.

[56] V. Clarke and V. Braun, “Thematic analysis,” *The journal of positive psychology*, vol. 12, no. 3, pp. 297–298, 2017.

[57] M. R. Mehl, “Quantitative text analysis.,” 2006.

[58] Y.-K. Lim, E. Stolterman, and J. Tenenberg, “The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 15, no. 2, pp. 1–27, 2008.

[59] B. Kuriakose, R. Shrestha, and F. E. Sandnes, “Multimodal navigation systems for users with visual impairmentsa review and analysis,” *Multimodal Technologies and Interaction*, vol. 4, no. 4, p. 73, 2020.

[60] D. Siroker and P. Koomen, *A/B testing: The most powerful way to turn clicks into customers*. John Wiley & Sons, 2015.

[61] J. Cuzick, "A wilcoxon-type test for trend," *Statistics in medicine*, vol. 4, no. 1, pp. 87–90, 1985.

[62] E. W. Weisstein, "Bonferroni correction," *https://mathworld. wolfram. com/*, 2004.

[63] J. Century and A. Cassata, "Implementation research: Finding common ground on what, how, why, where, and who," *Review of research in education*, vol. 40, no. 1, pp. 169–215, 2016.

[64] C. Rohrer, "When to use which user-experience research methods," *Nielsen Norman Group*, vol. 12, p. 21, 2014.

[65] A. Joshi, S. Kale, S. Chandel, and D. K. Pal, "Likert scale: Explored and explained," *British journal of applied science & technology*, vol. 7, no. 4, p. 396, 2015.

[66] O. C. Robinson, "Sampling in interview-based qualitative research: A theoretical and practical guide," *Qualitative research in psychology*, vol. 11, no. 1, pp. 25–41, 2014.

[67] J. Wang, H. Huang, Y. Li, H. Zhou, J. Liu, and Q. Xu, "Driving risk assessment based on naturalistic driving study and driver attitude questionnaire analysis," *Accident Analysis & Prevention*, vol. 145, p. 105 680, 2020.

[68] X. Zheng, H. Huang, J. Wang, X. Zhao, and Q. Xu, "Behavioral decision-making model of the intelligent vehicle based on driving risk assessment," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 7, pp. 820–837, 2021.

[69] C. Allen, M. Ewing, S. Keshmiri, *et al.*, "Multichannel sense-and-avoid radar for small uavs," in *Proceedings of IEEE/AIAA 32nd Digital Avionics Systems Conference (DASC)*, 2013.

[70] M. Sugarindra, M. Suryoputro, and A. Permana, "Mental workload measurement in operator control room using nasa-tlx," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, vol. 277, 2017, p. 012 022.

[71] Y. S. Lincoln and E. G. Guba, "Establishing dependability and confirmability in naturalistic inquiry through an audit.," 1982.

A

Appendix 1: Interview

Riding Experience:

1. How often have you been using e-scooters for transportation?
2. What is the typical distance of your e-scooter rides?
3. What are the main reasons you choose to ride an e-scooter instead of other forms of transportation? (for what purpose?)
4. How comfortable do you find the handlebars, and foot-rest area of the e-scooter?

Safety Perception:

1. Do you feel safe while riding an e-scooter? What factors contribute to your sense of safety (or lack thereof)?
2. Have you ever been in an accident or near-miss situation while riding an e-scooter? If so, could you describe what happened?
3. How important do you think it is for e-scooters to have safety features like lights, brakes, and helmets?

Navigation and Interaction:

1. When using your e-scooter, do you rely on a navigation system? If so, what type (e.g., smartphone app, built-in display)?
2. Have you used any special features on your e-scooter for navigation, such as turn-by-turn directions or location-based alerts? How useful were they?
3. How do you feel about the different types of warnings (auditory, visual, vibrotactile) that could be used on e-scooters? Which one do you find most effective in getting your attention in potentially dangerous situations?
4. If an e-scooter had a more advanced interaction system, like touch-screen controls or voice commands, would you use it? What functions would you like to see in such a system?

Future Improvements:

1. What do you think are the most important areas for improvement in e-scooter design or functionality?

B

Appendix 2: Survey

Introduction:

Thank you for taking the time to participate in this survey. We are two students from Chalmers who are doing our master thesis and we really need your feedback. Your insights will help us better understand the experiences, preferences, and needs of e-scooter riders. All responses will be kept strictly confidential.
Google Form link: `https://docs.google.com/forms/d/e/1FAIpQLSfbkCsI5bc9btsCYhx1H_l83FFG6YMPAvsSYK4BNuzbrUYxrw/viewform?usp=dialog`

Personal Information

1. ***Age**:
    - (a) 18 - 25
    - (b) 26 - 35
    - (c) 36 - 45
    - (d) 46 - 55
    - (e) 55+
2. ***Gender**:
    - (a) Male
    - (b) Female
    - (c) Non-binary / prefer not to state
3. **City/Area where you mainly ride e-scooters**: ____________________

Riding Experience

1. ***How often do you use e-scooters for transportation?**:
    - (a) hardly ever
    - (b) once a month

   - (c) once a week
   - (d) more than twice a week

2. ***What is the typical distance of your e - scooter rides?**:
   - (a) Less than 1 km
   - (b) 1 - 3 km
   - (c) 3 - 5 km
   - (d) More than 5 km

3. ***What are the main reasons you choose to ride an e-scooter instead of other forms of transportation? (Select all that apply) - Multiple Choice**:
   - (a) Convenience
   - (b) Cost-effectiveness
   - (c) Leisure
   - (d) Avoid traffic congestion
   - (e) Other (please specify)

Safety Perception

1. ***How safe do you feel while riding an e-scooter on a scale of 1 - 10 (1 = very unsafe, 10 = very safe)?**:
   1 2 3 4 5

2. ***What factors contribute to your sense of safety (or lack thereof) while riding an e-scooter? (Select all that apply)**:
   - (a) Traffic conditions
   - (b) Road conditions
   - (c) Other users' behavior
   - (d) Presence of dedicated bike lanes
   - (e) Other (please specify)

3. *** Have you ever been in an accident or near-miss situation while riding an e-scooter?**:
   - (a) Yes
   - (b) No
   - (c) If yes, please briefly describe what happened:

4. *** How important do you think it is for e-scooters to have safety features like lights, brakes, and helmets**:
    - (a) Extremely important
    - (b) Very important
    - (c) Somewhat important
    - (d) Not very important
    - (e) Not important at all
5. **How comfortable do you find the handlebars, and foot-rest area of the e-scooter on a scale of 1 - 10 (1 = very uncomfortable, 10 = very comfortable)?**:
    - Handlebars:
      1 2 3 4 5
    - Foot-rest area:
      1 2 3 4 5

Navigation and Interaction

1. ***To what extend do you want to put your phones inside the phone-holder?**
   1 2 3 4 5

Figure B.1: Phone holder

2. ***When using your e-scooter, do you rely on a navigation system?**:
    - (a) Yes
    - (b) No

(c) If yes, what type of navigation system do you use?

- Smartphone app(eg. google map)
- Built-in display on the e-scooter
- Other (please specify)

3. ***How do you feel about the different types of warnings that could be used on e-scooters (auditory, visual, vibrotactile)?**:

   (a) Auditory (beeping, voice prompts):

   - Extremely effective
   - Very effective
   - Somewhat effective
   - Not very effective
   - Not effective at all

   (b) Visual (lights, text messages):

   - Extremely effective
   - Very effective
   - Somewhat effective
   - Not very effective
   - Not effective at all

   (c) Vibrotactile (vibrations on the handlebars):

   - Extremely effective
   - Very effective
   - Somewhat effective
   - Not very effective
   - Not effective at all

4. ***If an e-scooter had a more advanced interaction system, like touch-screen controls or voice commands, would you use it?**:

   (a) Definitely Yes

   (b) Probably Yes

   (c) Neutral / Not Sure

   (d) Probably Not

   (e) Definitely No

(f) If yes, what functions would you like to see in such a system? (Select all that apply)

  - Adjusting speed
  - Changing riding modes
  - Checking battery status
  - Navigation direction
  - Other (please specify)

Future Improvements

1. **If in the future there is a helmet/handlebar that combines vibration and voice interaction, would you be willing to use it?**
    (a) Extremely willing to use
    (b) Very willing to use
    (c) Somewhat willing to use
    (d) Not very willing to use
    (e) Not willing to use at all
2. **What do you think are the most important areas for improvement in e-scooter design or functionality? (Select all that apply)**:
    (a) Longer battery life
    (b) More comfortable ride
    (c) Better safety features
    (d) More advanced navigation systems
    (e) Other (please specify)

Additional Comments

Do you have any other thoughts, suggestions, or experiences related to e-scooter riding that you would like to share?

Thank you again for your participation in this survey. Your responses are invaluable in helping us understand the e-scooter riding community better. If you have any questions about the survey, please feel free to contact us at yiongday@gmail.com