

Assessment of In-Vehicle HMI Distraction

Karsten Klafter
Mark Vollrath
Tina Gehlert
Greta Große

On behalf of the German Insurers Accident Research (UDV)
Research Report No. 105

Assessment of In-Vehicle HMI Distraction

Edited by:

Department of Traffic and Engineering Psychology, TU Braunschweig



Karsten Klaffer, M.Sc.
Prof. Dr. Mark Vollrath

Project Management at UDV:

Dr. rer. nat. Tina Gehlert

Publisher

Gesamtverband der Deutschen Versicherungswirtschaft e.V.
(GDV - German Insurance Association)
Unfallforschung der Versicherer
(UDV - German Insurers Accident Research)
Wilhelmstraße 43 / 43 G, 10117 Berlin
Potfahc 08 02 64, 10002 Berlin
Tel. 030 / 20 20 - 50 00, Fax 030 / 20 20 - 60 00
www.gdv.de, berlin@gdv.de
www.udv.de, unfallforschung@gdv.de
www.youtube.com/unfallforschung

Editorial work

Dr. rer. nat. Tina Gehlert

Photo references

The usage rights of the other photographs and graphics in this brochure are owned by the UDV (German Insurers Accident Research).

Press date

07/2026

ISBN-No.:

978-3-948917-36-4

All editions

on UDV.de

Disclaimer

The present contents have been prepared with all due care and attention. Nevertheless, no guarantee is made in respect of the completeness, correctness, current relevance or appropriateness of the information or appraisals given.

Table of Contents

Table of Contents.....	1
1 Introduction	4
2 Types of Human-Machine Interfaces	6
2.1 Physical Controls.....	6
2.2 Touch-Based Interfaces	7
2.3 Speech-Based Interfaces	7
3 Vehicle Functions.....	9
3.1 Functions and Driving Tasks	9
3.2 Case One: Volkswagen Golf 8	10
3.2.1 Physical Controls and Touch-Based HMI	10
3.2.2 Speech-Based HMI.....	11
3.3 Case Two: Tesla Model 3.....	12
3.3.1 Physical Controls and Touch-Based HMI	12
3.3.2 Speech-Based HMI.....	13
3.4 Conclusion.....	14
3.5 Optimal Operability & Clustering of Functions.....	14
4 Human-Machine Interfaces: Considerations for an Optimal Design	16
4.1 Applicable to all HMI Types.....	16
4.2 Applicable to Visual-Manual HMIs: Physical Controls & Touch-Based HMIs.....	16
4.2.1 Guidelines and Standards.....	16
4.2.2 Research.....	18
4.3 Applicable to Speech-Based HMIs.....	20
4.3.1 Guidelines and Standards.....	20
4.3.2 Research.....	21
4.4 Modality Choice for Different Tasks.....	22
5 HMI Checklist.....	23
5.1 Checklist Structure and Assessment Process.....	23
5.2 Development and Description of Checklist Parts	24
5.2.1 Assessment Scope	24
5.2.2 Items and Quantitative Scoring (Module I and II b).....	25
5.2.3 Optimal HMI Implementation and Use Context (Module II a)	27
6 Empirical Validation: Method	30
6.1 Participants.....	30
6.2 Design and Measures.....	30
6.3 Materials	32
6.3.1 Lane Change Task.....	32
6.3.2 Generic Human Machine Interfaces	33

6.4 Secondary Tasks.....	35
6.5 Procedure.....	36
7 Results	37
7.1 Driving Performance Measures and Data Preparation.....	37
7.2 HMI Type and Task Complexity: Effects on Driving Performance.....	38
7.2.1 General Overview: Simple and Complex Task Operation via Physical Controls, the Touch-Based and Speech-Based Interface	38
7.2.2 Interaction Between HMI Type and Task Complexity.....	40
7.3 Interface Design Aspects (Implementation Quality): Effects on Driving Performance and Function Operation Quality	43
7.3.1 Measures, Data Analysis Strategy, Scoring System, and Aspect Sorting	43
7.3.2 Interface Design Aspects: Physical Controls	46
7.3.3 Interface Design Aspects: Touch-Based Interface.....	62
7.3.4 Interface Design Aspects: Speech-Based Interface	90
7.4 Summary	111
8 Comparison to Euro NCAP's General Vehicle Controls	115
8.1 General Vehicle Controls: Structure	115
8.2 General Vehicle Controls and HMI Checklist: Assessment Process and Scope	116
8.3 GVC Function Implementation: Enhancement Options.....	117
9 Conclusion	122
9.1 HMI Checklist: Towards Implementation Quality.....	122
9.2 Empirical Validation: Ranking of the Design Aspects.....	123
9.3 Future Directions	124
References.....	126
List of Figures	130
List of Tables.....	130
Appendix A: HMI Checklist	133
Appendix B: Operationalization of the HMI Checklist's Interface Design Aspects.....	173

Abstract

Modern vehicles are equipped with a wide range of functions to improve driving safety and the comfort of drivers and passengers alike. To operate the vehicle's functions, different types of human-machine interfaces (HMI) are used. In the past, physical controls, i.e., buttons and knobs have been employed to operate the vehicle's functions. Nowadays, these have been largely replaced by central touchscreens that offer a rich graphical interface and allow for the operation of many of the vehicle's functions. However, due to their visual nature and reliance on precise manual input with a lack of tactility, they are rather distracting while driving. To mitigate this issue, many manufacturers include speech-based interfaces in their vehicles to omit visual-manual distraction.

When implementing a function in a vehicle, two aspects must be considered: The task of the function itself and the type of HMI(s) used for its operation. The function tasks vary in their complexity and proximity to the driving task. The more a function is needed for safe driving, the more important optimal operability becomes, i.e., operation should be fast, be little prone to errors and lead to little distraction. Technically, any of the three types of interfaces can be used to operate a function: physical controls, touch-based, and speech-based interfaces. However, depending on the characteristics of a function, some types of HMI types may be better suited than others. Beyond assigning a function to a type of HMI, the quality of the implementation must be considered. Each HMI type has a high degree of freedom, for example, touchscreens differ in their size, position, and software design. The literature on implementation quality is rather fragmented, unspecific, and largely focuses on visual-manual interfaces (physical controls and touch-based interfaces) with speech-based interfaces missing. The aim of this project was to develop an HMI checklist evaluating important design aspects of each HMI type regarding possible distraction from driving, thus enabling an overall assessment of the implementation quality of a function.

The HMI checklist consists of two modules (shared and function-specific HMI features) and contains 51 items. On a general level, the use context of the function is specified and the suitable HMI types for its operation are derived. If the actual HMI types are appropriate, the implementation quality within each HMI type is assessed with the corresponding items. Each item refers to a design aspect that must be fulfilled for a good implementation quality. To this end, the items contain an explanation, testing procedure, and a five-point rating scale. After assessing all items, a sum score can be calculated, where higher sums indicate a larger deviation from the ideal implementation.

A key part of the checklist's development was the empirical validation of the design aspects across the three HMI types. For the validation, two operation tasks (simple and complex) were incorporated in a generic interface for each HMI type (physical controls, touch-based, and speech-based). Driving performance was measured via the Lane Change Task. Reference driving performance was set with the interface's reference version (RV) where all design aspects were implemented optimally. To determine the impact of a design aspect, it was violated in the interface and the effect on driving performance was compared to the RV.

For a general overview, the driving performance of the RV was compared across HMI types and both tasks (simple and complex). For the simple task, physical controls and speech-based operation led to the best driving performance. For the complex task, speech-based operation was best. Regarding implementation quality, the most important design aspects for physical controls are those affecting manual effort. Ideally, the control elements should require little effort by incorporating shallow travel distances, a high mechanical integrity, and by placing them close to the driver. For touch-based interfaces, the overall interaction effort should be kept minimal. To achieve this, tapping instead of sliding should be used and buttons should be aligned (e.g., in a geometrical grid) and be of appropriate size. Also, extra operation steps should be avoided to reduce the likelihood of errors. For speech-based interfaces, to keep distraction low, the interaction should not fall back to other HMI modalities. Also, the speech recognition should be reliable and not introduce errors that need to be corrected.

The HMI checklist focusses on the principle of implementation quality by highlighting the individual impact of the design aspects regarding low-distraction function operation. Future research could expand this by incorporating medium-complex function tasks, different driving situations, and varying levels of implementation quality instead of a best- and worst-case implementation.

In-Vehicle Human-Machine Interfaces: Checklist for Assessing Distraction and Operation Safety

1 Introduction

The automotive industry is a high-paced environment regarding its technological advancements. At first sight, this usually concerns the driving capabilities of vehicles to either enhance the driving experience itself or to optimize the vehicle's handling across different environments. More importantly, however, the implementation of technologies is often aimed at improving safety (e.g., by using anti-lock braking or electronic stability control). Nowadays, vehicles are highly computerized and use sensors that enable them to analyze their environment and perform driving maneuvers on their own (Ahmed et al., 2022). Thus, most production cars are equipped with advanced driver assistance systems, such as adaptive cruise control, lane keeping assistance, or automatic emergency braking. In the future, the aim is to offer self-driving capabilities so that a human driver is no longer needed (Society of Automotive Engineers, 2021).

For the foreseeable time, however, a driver will still be necessary for safe vehicle operation. Beyond the core driving task, the driver is responsible for managing essential safety functions such as headlights or windscreen wipers. While these safety-critical operations remain central, vehicles also employ a vast array of non-safety features designed to improve comfort of drivers and passengers alike. These comfort-oriented functions can potentially compromise driving safety depending on their implementation in the vehicle cockpit (e.g., Dingus et al., 2016).

To interact with the vehicle and its functions, in-vehicle human-machine interfaces (HMIs) are used. Like the driving capabilities, HMIs underwent a technological evolution (Meixner & Müller, 2017). The first HMIs were purely mechanical and only displayed driving-related information – typical examples are the speedometer and tachometer. Then, entertainment options were implemented, which started as simple radios and evolved into infotainment systems. Accordingly, the amount of information presented to the driver increased steadily. Modern HMIs still display driving- and vehicle-related information but are accompanied by driver assistance systems, navigation, and entertainment options. For a long time, buttons and knobs placed in the center console were used to control the vehicle's functions. Over the past decade, however, many of these mechanical input devices have been replaced by a central touchscreen with a rich graphical user interface.

The central touchscreen can be characterized as a rather complex HMI that is the result of two opposing trends. On the one hand, the amount of vehicle functions increased, while at the same time the number of dedicated devices to control them decreased. Thus, to reach a given function, drivers usually must pass through multiple menu layers and perform high-effort interaction gestures that go beyond a simple tap. For increased access, manufacturers might put an array of functions on a single screen to avoid having to go through many menus. But this solution imposes its own challenges as users now must visually search through the small elements on the screen. This is known as

a depth and breadth trade-off (Nicolas et al., 2024), representing just one challenge when designing the central touchscreen's interface.

Due to their visual nature and high interaction requirements, the switch from physical controls to a central touchscreen is a severe safety risk. Drivers need to look away from the road for extended periods of time (Grahm & Kujala, 2020). To reduce or even eliminate visual distraction, while still allowing access to complex functions, manufacturers increasingly incorporate speech-based interfaces as an additional input method. Although they mostly omit visual distraction, a speech-based interface may still lead to cognitive distraction which in turn can negatively affect driving performance. However, the effects are less severe when compared to visual-manual distraction (Zhang et al., 2023).

When implementing a function into the vehicle's HMIs two main questions arise. First, in what way should a function be implemented that ensures optimal interaction? Should the interaction happen via physical controls (e.g., buttons, knobs), via a touchscreen, or a speech-based interface? Research suggests that some interface types are more suitable for simple, single-step tasks, while others are better for more complex tasks with multiple steps. And secondly, with choosing an interface type, what elements make the interface low-distraction and easy to use? Both aspects (interface type choice and interface design elements) should be considered to minimize driver distraction.

The objective of this project was to develop a checklist that assesses whether a given function is safely implemented regarding the choice of interface type, the design of interface elements, and the interaction during function operation. The assessment tool includes criteria that are linked to low-distraction HMIs and that ensure safe function operation. Furthermore, the use context of the function is determined and then used to point towards the optimal HMI type (physical controls, touch-based, speech-based). The interface types and design criteria were subsequently validated in a driving simulator experiment regarding driver distraction using two tasks of different complexity. The results were used to validate and rank the design guidelines for low-distraction HMIs.

This report summarizes the project results. Chapters 1 and 2 introduce the three HMI types – physical controls, touch-based, and speech-based interfaces. Chapter 2 outlines their basic properties, interaction principles, and distraction dynamics. Chapter 3 illustrates how the HMI types are used for operating functions in two different car models. Chapter 4 reviews standards, guidelines, and research on HMI design with a special focus on driver distraction. The findings of chapters 1 – 4 were used to derive HMI checklist. In chapter 5, the HMI checklist is described in more detail, including its structure and application. Chapter 6 highlights the driving simulator experiment used for validating the HMI checklist. In chapter 7, the results are described. On a general level, the effects of HMI type and task complexity on driver distraction are outlined. Then, every design aspect per HMI type is analyzed. Chapter 8 compares the HMI checklist with Euro NCAP's General Vehicle Controls protocol that also assesses vehicles' HMI design. The final chapter 9 summarizes key project results and outlines further research.

2 Types of Human-Machine Interfaces

When the driver controls the movement of a vehicle, driving is considered the primary task. The operation of any function not directly affecting the vehicle's movement can therefore be classified as a secondary task (e.g., wiper controls), or depending on the function, as a tertiary, non-driving-related task (e.g., switching to a different song; Bubbs et al., 2015). To ensure the primary task can be carried out in a satisfactory quality, the interaction with a function must be low-distraction, i.e., easy to carry out, while also preventing operation errors. To achieve this, different HMIs can be used that are best described by their modality. Modality refers to the sensory channel used for the input and output between the driver and in-vehicle system (Karray et al., 2017).

Based on the idea of a main modality, the in-vehicle HMIs can be grouped into three different categories: physical controls, touch-based interfaces, and speech-based interfaces. These will be described in more detail in the following sections highlighting their main characteristics and the interaction process during function operation. For understanding the basic distraction mechanism, Wickens' (2002) Multiple Resource Theory will be applied. It suggests that attentional resources are split between different processing stages (perception, cognition, response), modalities (visual, auditory, tactile), and processing codes. When two tasks compete for the same resource, i.e., stage, modality, or code, performance decreases. Inversely, no interference occurs if two tasks use different resources. For driving, the visual resource is used at the perceptual stage and the manual resource at the response stage. If function operation competes for any of these resources, distraction will occur and driving performance decreases.

2.1 Physical Controls

Physical controls (e.g., buttons, knobs) are characterized by their physical-haptic nature. All stages of the operation (reaching the element, activating/adjusting the function, and if possible, even feeling the status of the function) can ideally be achieved without looking at the control element. The interaction enables instant haptic feedback as the movement of the control element itself can be felt. The elements are usually placed in various locations in the cockpit: on the steering wheel, on the steering column, on the door, overhead, or on the instrument panel and center stack (Ma et al., 2022; Robertson & Cox, 2024). But not only their placement is diverse, the elements themselves can take many different forms: buttons, stalks, levers, rotary switches, joysticks, and more. The physical movements from the driver are rather simplistic and come in the form of tapping, pushing, or turning and are usually done in a single, continuous movement.

From a function operation standpoint, there are two main problems with physical controls. First, they can only be used to control a limited number of functions. In most cases, one element only controls a single function (e.g., one button turns a single function like the rear defogger on or off). In other cases, multiple functions can be controlled via e.g., a rotary switch (adjusting the audio volume via turning and muting via pressing) although the variety is still limited when compared to a touch-based interface. Second, the process of controlling complex, multiple-step functions is cumbersome (e.g., inserting a phone number or an address via a joystick). Nevertheless, the simplicity of physical controls offers the benefit that the function and its operation can be easily understood. Additionally, the function can be reached directly, i.e., it is not hidden in a menu.

The interaction with physical controls can be analyzed using Wickens' (2002) Theory of Multiple Resources. The main resources regarding driving (primary task) are the visual resource at the perceptual stage and the manual resource at the response stage. For the visual-perceptual resource, two cases can apply when interacting with physical controls: Firstly, and ideally, blind operation is possible as the driver has learned the position and operation gestures of the function. In this case, only manual resources are needed, but no visual resources. Secondly, the driver needs to look at the element so that visual resources are required. However, once the information is perceived, no future glances are needed as the labeling will not change. Thus, both sufficient visual and manual resources should be available to execute the driving task.

Exceptions to this are programmable buttons with an integrated display and elements that control visual elements in a separate screen. In these cases, there will still be an increased visual demand. However, physical controls should still offer a benefit over touchscreens with their sparse use of the manual response resource: A press of a button or the turn of a knob with haptic feedback requires significantly less precision than a slide or zoom gesture on a touchscreen. Furthermore, the position of physical controls is fixed in comparison to virtual input elements on a touchscreen. As a result of the overall sparse use of resources, physical controls only cause small interferences with the driving task and are ideal for simple function operation tasks (Ma et al., 2022).

2.2 Touch-Based Interfaces

Touch-based interfaces (e.g., central touchscreen) are characterized by their rich visual interface and capability to process touch inputs. During all stages of the operation, there is almost no haptic feedback. Obviously, the driver can feel the screen, once the finger makes contact with its surface, however, the control element itself, e.g., a virtual button cannot be felt. Furthermore, turning on/off, or adjusting a function is usually not communicated via haptic feedback. To output the status of the function, visual feedback is used, sometimes accompanied by audio feedback. Research suggests that drivers prefer multimodal feedback (e.g., haptic, audio) over visual feedback alone (Pitts et al., 2012). The physical device is usually positioned in the middle of the center console and sometimes tilted towards the driver (Patel et al., 2022). Screen sizes vary, as does the orientation of the screen (vertical or horizontal). Compared to physical controls, there is far less variety in the physical appearance (a screen vs. a button, knob, lever, etc.) as well as in the positioning of the screen (center console vs. steering wheel, steering column, center console, etc.). However, there is more diversity regarding the actions that the driver must perform to interact with the screen: tapping, sliding, or pinching. Compared to physical controls, touch-based interactions are more involved because there is nothing to grab, hold onto, and there is no haptic feedback.

For function operation, there are two main advantages to using a central touchscreen. Firstly, many functions can be placed inside the graphical user interface, thus not requiring much space in the cockpit. Secondly, functions of any complexity can be implemented – from simple one-step interactions (e.g., turning the window defogger on or off) to complex multiple-step interactions (e.g., adding an additional stop to the navigation). As with physical controls, there is an inverse relationship between the number and complexity of functions and the simplicity of the HMI. For the touchscreen, this results in a highly complex interface that is not easy to use (at least when compared to physical controls). Many of the functions cannot be directly accessed because they are placed in deep menu hierarchies.

As before, the interaction with the touchscreen can be analyzed using Wickens' (2002) Theory of Multiple Resources. For driving, the visual resource at the perceptual stage and the manual resource at the response stage are required. As touchscreens are visual interfaces, they heavily rely on the visual-perceptual resource. Drivers need to visually search for the desired control element of the function, interact with it, and then visually check again if the interaction was a success. In other cases, the screen needs to be looked at continuously to reach the desired function status (e.g., switching the map to an optimal zoom level). Thus, the visual resource is repeatedly unavailable for the driving task when interacting with the touchscreen. To some extent, the same is true for the manual resource for the response. As the interface is flexible, even a tap could land anywhere on the screen, depending on where exactly the function was placed. Further, some actions require continuous tracking (sliding) or complex gestures (pitching). Moreover, these manual actions mostly require visual control since no haptic feedback is provided. Because of their high usage of resources, touchscreen-based interactions usually result in the worst driving performance when compared to physical controls or speech-based interfaces (Ma et al., 2022).

2.3 Speech-Based Interfaces

Speech-based interfaces, like their name suggests, rely on the auditory channel for input and output. In contrast to physical controls and touch-based interfaces, the operation does not rely on physical devices, at least under ideal circumstances. However, as Young & Zhang (2017) note, in-car speech technologies are rarely "pure" – usually a mixed-mode interface is used where visual information is displayed, e.g., in the central touchscreen. To activate the speech-based interface, drivers can either press a push-to-talk button or say a keyword. With the interface activated,

a command can be voiced to operate a vehicle function (e.g., “Turn up the music!”). At the end of the input, voice feedback, a chime and/or visual feedback are presented. If and what keyword is used differs between systems, the same goes for the chimes used.

From a function operation standpoint, the speech-based interface could theoretically be used to control most in-vehicle function. However, most manufacturers only provide access to a selection of functions. As the speech-based interface does not require physical space inside the cockpit, controls are only limited by the commands used. However, this highlights the first challenge when working with speech-based interfaces. Users must be informed about the scope of functions available via the speech-based interface and how to address them. Furthermore, the system should be able to recognize a multitude of commands, some of which do not explicitly state the function. Instead of “Increase the temperature!”, a driver might say “I am freezing.” or “Make it warmer!”. Another challenge lies in voice recognition: noises, dialects, mumbling, or conversations of passengers introduce noise the system has to account for when identifying the target phrase of the driver. To overcome these issues, neural networks may be used (Khdier et al., 2021; Lyashenko et al., 2021). Although basically any function could be accessed with speech-based interfaces, it could be counterproductive to do so: Especially single-step actions (“Turn the wipers on!”) take longer to activate via speech-based interfaces than via other means. This is especially critical for immediate-to-execute functions. Furthermore, as speech recognition can be erroneous, the target function might not even be activated which may result in a safety hazard (The driver can’t see the road ahead). Error handling presents another challenge, as drivers need to be informed about the error and how to proceed further. Finally, the speech-based interface can be rather complex when addressing functions can be done via numerous commands. This may also lead to errors when the system confuses commands.

Like with the previous HMIs, the speech-based interaction can be analyzed using Wickens’ (2002) Multiple Resource Theory. For driving, the visual (perception), cognitive, and manual (response) resource are required. Any parallel task sharing these resources with the driving task will result in a deteriorated driving performance. In contrast to physical controls and touch-based interfaces, the speech-based interface relies on the auditory resource for perception and vocal resource for response. Thus, driving and operating a function via the speech-based interface take place in two different modalities, visual and auditory, respectively. As a result, the interference between both tasks is minimized, although it cannot be completely eliminated. The reason lies in the use of the cognitive resource where a driver must both coordinate the use of the speech-based interface (e.g., remembering commands, reacting to the speech-based interfaces’ output, dealing with errors), while simultaneously understanding and anticipating the movement of other traffic participants (Endsley, 1995). If the speech-based interface is not “pure” (i.e., using visual feedback; Young & Zhang, 2017) the interference with the driving task will increase. Still, as the speech-based interface does not (strongly) rely on the visual-manual resources needed for driving, they are a lower distraction option for operating complex functions (Ma et al., 2022).

3 Vehicle Functions

The aim of this project was to develop a checklist that evaluates whether an HMI for operating a given function is low-distraction and safe-to-use, i.e., operation errors are unlikely to occur. The basis for our instrument was the checklist of Horn et al. (2023), which is largely based on the NHTSA guidelines for visual-manual distraction (National Highway Traffic Safety Administration, 2013, 2016). In the first two modules, named “devices” and “functions”, the NHTSA criteria for the assessed HMI are checked. The third module in the form of a decision tree was a new development and is used to define the use context of the function independent of the HMI implementation. It defines an acceptable threshold for the implementation of a function regarding the modality (physical controls, touch-based, speech-based) and menu depth (e.g., 2 steps).

Separating the function from its HMI implementation is a sensible idea because a function could theoretically be accessed through any of the three HMI types (physical controls, touch-, and speech-based). Obviously, not every function should be accessible through every HMI type. Some functions need to be operated instantly (e.g., turning off the high beam) where the speech-based interface or touchscreen would be too slow, thus physical controls are the best option. In contrast, complex and infrequently accessed functions like vehicle settings (e.g., setting the date and time) can be placed in a touchscreen menu without relying on physical controls.

We started our work with a focus on vehicle functions and their current implementations in popular vehicles. This enabled us to outline the scope of functions available to the driver. We later grouped these functions into clusters and defined optimal operability. This work was then used to enhance Horn et al.’s (2023) decision tree to better define the usage context of a function. This will be described as part of the development of our own checklist (see section 5.2.3).

3.1 Functions and Driving Tasks

By using a vehicle function, the driver tries to achieve a certain task. According to Bubbl et al. (2015), these tasks can be grouped into three different categories: primary, secondary, and tertiary tasks. Primary tasks are directly related to the driving task where the driver’s goal is to keep the car on track by using the steering wheel, throttle, and brake. Here, maneuvers are executed and a trajectory is being followed. Secondary tasks are closely related to the primary task as they enable the driver to safely fulfill the primary task. The use of the indicators, horn, or windscreen wipers fall into this category. Finally, tertiary tasks are not related to the driving task and contain comfort and infotainment functions. Most tasks that are employed as part of distraction research usually fall into this category like climate control, media, or phone usage. Obviously, for optimal operability, the primary and secondary task functions must be as easy to use as possible. Tertiary task functions should be easy to use and low-distraction as well, although they are less important for the driving task.

To further categorize the tertiary tasks and gain insight into their prevalence, Huemer & Vollrath’s (2011) study on driver secondary tasks can be utilized. Here, device usage inside the car was analyzed, where the most frequent tasks were audio-related (e.g., using the radio, changing the CD), followed by telecommunication tasks (although via mobile phone), followed by the handling of the climate control (e.g., adjusting the temperature). The data of the study, however, is from 2009 and the number of available functions in vehicles has greatly increased since then. Starkey & Charlton’s (2020) study is more up-to-date: They were interested in drivers’ usage of in-vehicle system applications and asked participants how often they showed 14 different behaviors related to in-vehicle information system usage. Although these behaviors are sometimes broad (e.g., “use of manufacturer display”) and thus do represent individual operation tasks, the data indicates that nearly all drivers engaged in audio-related tasks (“listen to music”, “select music on phone”), followed by navigation tasks (“listen to directions”, “look at map on navigation system”), and communication tasks (“talk on hands free”, “text/email/social media”). Reading, internet usage, and the reception of warnings were reported the least.

The classification of Bubb et al. (2015) was used during the research of the sample vehicle's functions to assess whether they belong into the primary, secondary, or tertiary task category. The studies on the usage frequency of functions (Huemer & Vollrath, 2011; Starkey & Charlton, 2020) will later be used to classify the large number of tertiary tasks not relevant to the driving task. This will be done at the end of this chapter as part of the development of the function cluster for optimal operability.

3.2 Case One: Volkswagen Golf 8

The Volkswagen Golf 8 is a popular vehicle on German roads. In 2024, over 100.000 new cars were registered (Kraftfahrt-Bundesamt, n.d.). Thus, there are many drivers who will interact with the car's HMIs. Depending on the trim level, there will be differences between each individual vehicle, however, the general HMI concept is the same in every car. There are videos on the video platform YouTube that explain how basic functions can be operated. We used them as the basis of our research. We viewed ten videos from a YouTube channel dedicated to explaining vehicle functions (e.g., Fahrzeugfunktionen Einfach Erklärt, 2020) and an additional video explaining the speech-based interface (Schwabens - Checker, 2020). The provision of explanatory videos somewhat highlights the increased complexity of modern in-vehicle HMIs and the need to explain them towards novice vehicle owners.

3.2.1 Physical Controls and Touch-Based HMI

Table 1 shows the functions that rely on a visual-manual interaction (Wickens, 2002). Functions are clustered along two dimensions. First, they are grouped into sections following Bubb et al.'s (2015) proximity to the driving task. Secondly, functions are grouped into function groups to keep the reporting of results compact.

Table 1: Volkswagen Golf 8 – functions accessible via physical controls and touch-based HMIs

PRIMARY				
Function Group	Functions	Device	HMI Type	Steps
Accelerating	1	Pedal	Physical Controls	1
Braking	1	Pedal	Physical Controls	1
Steering	1	Steering Wheel	Physical Controls	1
SECONDARY				
Function Group	Functions	Device	HMI Type	Steps
Drive Mode	5	Lever	Physical Controls	1
Exterior Light Functions	3	Touch button	Touch	1
Defogger	2	Touch button	Touch	1
High Beam Control	2	Lever	Physical Controls	1
Indicators	2	Lever	Physical Controls	1
Wiper Controls	7	Lever	Physical Controls	1
Driving Assistance Control	6	Button	Physical Controls	At least 1
TERTIARY				
Function Group	Functions	Device	HMI Type	Steps
Windows	3	Rocker switch	Physical Controls	1
Mirrors	3	Rotary knob	Physical Controls	2
Interior Lights	3	Touch button	Physical Controls	1
Media Control	6	Button (SW)	Physical Controls	At least 1

TERTIARY				
Function Group	Functions	Device	HMI Type	Steps
Change View (Instrument Cluster)	1	Button (SW)	Physical Controls	At least 1
Start Speech-Based HMI	1	Button (SW)	Physical Controls	1
Quick Settings*	Variable (15)	Touchscreen	Touch	At Least 1
Functions on Home Screen*	Variable (5)	Touchscreen	Touch	At Least 1
Navigation*	29	Touchscreen	Touch	Avg.: 1.6
Media Control*	36	Touchscreen	Touch	Avg.: 1.1
Climate Control*	18	Touchscreen	Touch	Avg.: 1
Driver Assistance Settings*	13	Touchscreen	Touch	Avg.: 1.7
Vehicle Setup/Settings*	22	Touchscreen	Touch	Avg.: 2
...	...			
TOTAL	185			

Note. * = Operation steps counted within the function's menu. If the operation starts from the home screen at least one step must be added. Moreover, settings are counted as individual functions in the touchscreen menus.

3.2.2 Speech-Based HMI

Table 2 shows the functions available via the speech-based HMI that is based on a verbal-vocal interaction (Wickens, 2002). Like with the visual-manually accessible functions, function groups and the number of accessible functions are listed. Furthermore, the number of voice commands tested and if the speech-based interface provided or needed additional information are indicated.

Table 2: Volkswagen Golf 8 – functions accessible via the speech-based interface

SECONDARY			
Function Group	Functions	Commands	Inquiry?
Defogger	Rear Defogger	1	N
TERTIARY			
Function Group	Functions	Commands	Inquiry?
Climate Control	5	21	Y
Navigation	2	11	Y
Media Control	5	7	Y
Phone	3	3	N
Launching Apps	2	2	N
Sun Roof	1	2	N
...	...		
TOTAL	18		

3.3 Case Two: Tesla Model 3

Tesla was one of the first manufacturers to use a central touchscreen to access most of the vehicle's functions. Since then, all their models rely on the central touchscreen as the main interaction method while making minimal use of physical controls. In Germany, this sparks concerns as the use of a vehicle's touchscreen is only allowed when glance time remains short (Straßenverkehrs-Ordnung, 2013). If a driver loses control of the vehicle and causes an accident because they were looking away from the road for extended periods of time, this will be classified as a traffic offense (ADAC, 2020). Thus, Tesla's HMI concept remains controversial to this day. Furthermore, it is complex and deviates from what drivers are used to in other vehicles. As a result, like with the Volkswagen Golf 8, there can be videos found on YouTube explaining how to operate the vehicle's functions. We used these videos to better understand how the central touchscreen (e.g., Alex Sibila, 2023) and the speech-based interface (e.g., Technikfieber, 2023) are operated.

3.3.1 Physical Controls and Touch-Based HMI

Table 3 shows the functions that rely on a visual-manual interaction (Wickens, 2002). Functions are clustered along two dimensions. First, they are grouped in sections following Bubb et al.'s (2015) proximity to the driving task. Secondly, functions are grouped into function groups to keep the reporting of results compact

Table 3: Tesla Model 3 – functions accessible via physical controls and touch-based HMIs

PRIMARY				
Function Group	Functions	Device	HMI Type	Steps
Accelerating	1	Pedal	Physical Controls	1
Braking	1	Pedal	Physical Controls	1
Steering	1	Steering Wheel	Physical Controls	1
SECONDARY				
Function Group	Functions	Device	HMI Type	Steps
Drive Mode	3	Touchscreen	Touch	1
Exterior Light Functions	4	Touchscreen	Touch	2
High Beam Control	2	Touch button (SW)	Touch	1
Indicators	2	Touch button (SW)	Touch	2
Wiper Control	2	Touch button (SW)	Touch	1
	6	Touchscreen	Touch	2
Driving Assistance Control	2	Scroll wheel (SW)	Physical Controls	At least 1
Defogger	2	Touchscreen	Touch	At least 1
TERTIARY				
Function Group	Functions	Device	HMI Type	Steps
Start Speech-Based HMI	1	Touch button (SW)	Touch	1
Windows	2	Rocker switch	Physical Controls	1
Mirrors	2	Touchscreen	Touch	1 & 4
		Scroll wheel (SW)	Physical Controls	
Quick Settings*	9	Touchscreen	Touch	At least 1
Climate Control*	23	Touchscreen	Touch	At least 1
Navigation*	19	Touchscreen	Touch	Avg.: 2.7

TERTIARY				
Function Group	Functions	Device	HMI Type	Steps
Media Control*	7	Touchscreen	Touch	Avg.: 2.6
Phone*	4	Touchscreen	Touch	Avg.: 2.3
Settings & Vehicle Setup*	78	Touchscreen	Touch	Avg.: 1.1
...	...			
TOTAL	171			

Note. * = Operation steps counted within the function's menu. If the operation starts from the home screen at least one step must be added. Moreover, settings are counted as individual functions in the touchscreen menus.

3.3.2 Speech-Based HMI

Table 4 shows the functions available via the speech-based interface that is based on a verbal-vocal interaction (Wickens, 2002). Like with the visual-manually accessible functions, function groups and the number of accessible functions are listed. Furthermore, the number of voice commands tested and if the speech-based interface provided or needed additional information are indicated.

Table 4: Tesla Model 3 – functions accessible via the speech-based interface

SECONDARY			
Function Group	Functions	Commands	Inquiry?
Wiper Control	3	3	N
Defogger	1	1	N
TERTIARY			
Function Group	Functions	Commands	Inquiry?
Climate Control	15	15	N
Navigation	9	9	N
Settings & Vehicle Setup	13	13	N
Open Menus	3	3	N
Phone & Messaging	2	2	Y
Web browsing	3	3	N
Entertainment	1	1	N
...	...		
TOTAL	53		

3.4 Conclusion

Both vehicles, the Volkswagen Golf 8 and Tesla Model 3, offer well over 100 functions to drivers. Obviously, the videos did not show the use of every single function, but they still serve to demonstrate the “function inflation” in modern cars. Looking at the visual-manually accessible functions contained in Table 1 and Table 3, the following conclusions can be made:

- Primary task functions are exclusively accessed via physical controls with only one operation step.
- Secondary task functions are primarily accessed via a mix of physical controls and the touchscreen. The Volkswagen Golf 8 exclusively uses physical controls with one operation step. The Tesla Model 3 relies on the touchscreen for over half of the function groups, the rest of the functions can be operated via physical controls and often requires more than one operation step.
- Tertiary task functions are primarily implemented via the touchscreen and seldomly require just one operation step. Again, across function groups, the Tesla Model 3 uses far less physical control elements than the Volkswagen Golf 8.
- Regarding the number of functions, their number steadily increases the higher the task category (see Bubb et al., 2015).

For the results contained in Table 2 and Table 4, they can be summed up as follows:

- Both systems are flexible in regard to the voice commands they accept.
- In general, less functions are available via the speech-based HMI when compared to the visual-manual HMIs.
- Secondary task functions can only be sparsely accessed. In the Volkswagen Golf 8, only one function group could be accessed, in the Tesla Model 3, two function groups could be operated.
- The number of available tertiary functions is greater than the number of secondary functions.

As mentioned above, the Tesla Model 3 deviates from the Volkswagen Golf 8 with its stronger reliance on the touchscreen. A similar deviation can be seen in the functioning of the speech-based interface. While the Golf offers a “pure” speech-based interface (Young & Zhang, 2017) that can be operated hands-free via the use of a keyword and spoken feedback, the Tesla Model 3 omits both of these elements. Drivers must start the speech-based interface with physical controls and visually check for feedback on the screen. Although a “beep” is used to indicate the processing of the command, it does not sufficiently transport the activation status of the function.

3.5 Optimal Operability & Clustering of Functions

After reviewing the literature (see chapter 4), a function’s implementation should strive for *optimal operability*, i.e., operation should be easy (i.e., not many operation steps), efficient (i.e., instant activation/feedback) and safe (i.e., low chance of operation errors). As optimal operability differs between HMIs (as well as their distraction potential), a given function should use an appropriate HMI for its operation. To make the grouping of functions easier, we consolidated the functions of both vehicles (Volkswagen Golf 8 & Tesla Model 3) into larger function groups. These groups were then arranged along three dimensions to determine their optimal operability (see Figure 1):

1. Proximity to driving task (x-axis): Primary, secondary or tertiary task according to Bubb et al. (2015).
2. Safety relevance (y-axis, left): According to the decision tree in Horn et al. (2023).
3. Complexity & distraction potential (y-axis, right) based on Ma et al. (2022).

Furthermore, on the left-side y-axis, proximity to the driver and input method were added, and on the right-side y-axis, several HMI operation characteristics were added.

Optimal operability increases when the proximity to the driving task decreases and the safety relevance increases. Thus, optimal operability is the highest in the top left-hand corner of the cluster (see Figure 1). Accordingly, it's the lowest in the bottom right-hand corner. Overall, the cluster illustrates the same trend found in the vehicle functions tables above: Groups further away from the driving task contain a higher number of functions, i.e., only a few functions belong in the primary task class, while many belong to the tertiary driving task (see Tables 1 – 4).

For primary driving tasks, safety relevance is the highest, thus the corresponding functions are usually close to the driver and implemented via physical controls. At the same time, the interaction is simple, offering a tactile, blindly usable and latency-free operation via a single step. This is mostly true for the secondary tasks as well, which are still rather important for driving safety. Functions belonging to the tertiary task class, however, are not safety relevant itself and operated via touch-based interfaces requiring multiple-step input. These functions should still be somewhat optimally operable since they can still introduce driver distraction, but regarding the driving task, they are irrelevant. We suggest to optimize and prioritize the access of those functions that are used more frequently as stated in Huemer & Vollrath (2011) as well as Starkey & Charlton (2020).

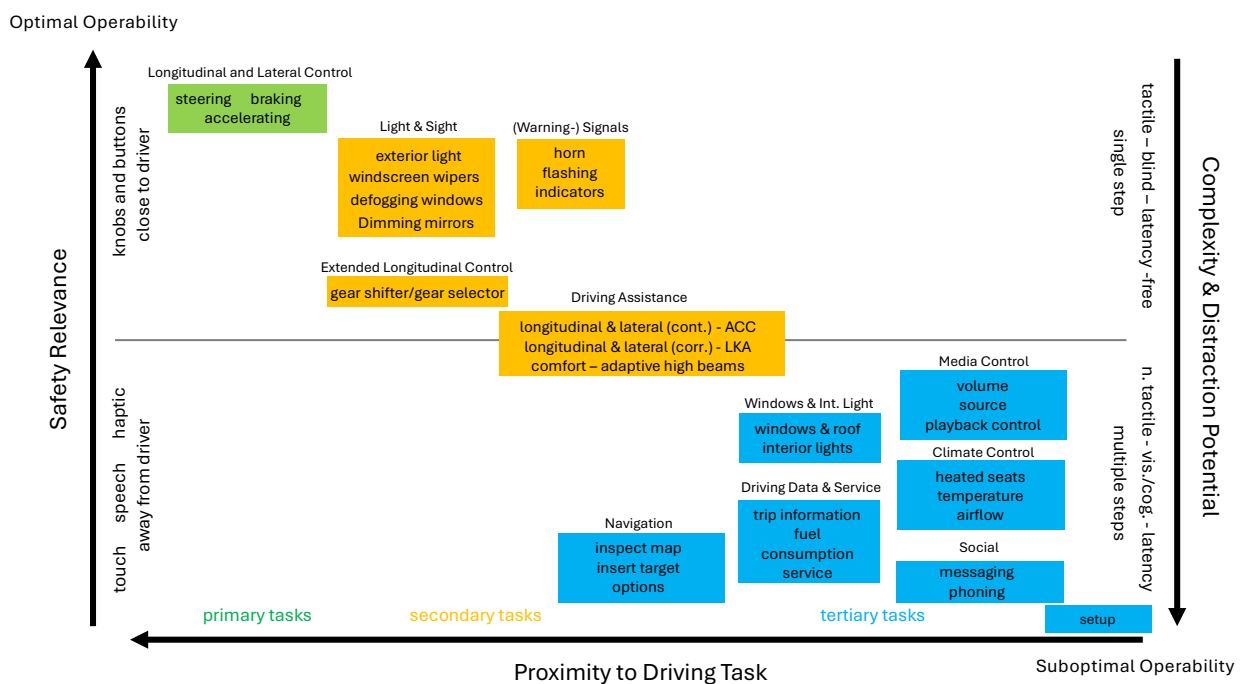


Figure 1: Function Cluster for Optimal Operability

4 Human-Machine Interfaces: Considerations for an Optimal Design

The design process of an automotive HMI represents a challenge where creative freedom is met with customer needs, market conditions, and safety considerations. Manufacturers must ensure that the HMI does not impede the primary driving task. As HMIs are complex, designers may use resources that point them towards relevant design aspects for an optimal, low-distraction function implementation. Furthermore, for a given function, the optimal HMI modality (physical controls, touch-based, or speech-based) should be chosen to ensure optimal operability.

The literature on HMI design can be split up into two groups that differ in their specificity and number of HMI aspects covered. The first group of literature is primarily made up of guidelines, norms, and reviews that are broad but not very specific. They cover many HMI interaction principles but lack details on how an actual, optimal implementation should look like. Thus, a guideline may point out that the system response must be “timely and clearly perceptible” (European Union, 2008, p. 21) without providing specific time frames (What exactly does “perceived as quite instantaneous” mean?). The second group of literature mostly contains study results that are much more specific but not broad. For example, the study of Liu et al. (2022) explored which parameters influence the click and sliding performance on the central touchscreen. Obviously, this is just one aspect of the interface design, and the interaction performance will also depend on other aspects like e.g., button shape or feedback. The latter group of individual studies also contains information on which HMI modality to choose with respect to the operation task’s complexity.

The following sections provide an overview of the literature that served as the foundation of the HMI checklist. Per HMI type, each section starts with the “broad literature” that will then be supplemented by individual study results. Descriptions will be kept at a more general level. The section closes with studies highlighting how the interface type and task complexity interact with each other.

4.1 Applicable to all HMI Types

Regardless of the HMI type used (physical controls, touch-, or speech-based), through each HMI a dialogue with the user is started that should adhere to certain rules. ISO 15005:2017 *Road vehicles — Ergonomic aspects of transportation and control systems — Dialogue management principles and compliance procedures* (International Organization for Standardization, 2017b) illustrates principles for the interaction between driver and the vehicle’s transport and information and control system (TICS) while the vehicle is in motion. The first principle explains that the dialogue must be compatible with the driving task – it should not require extended visual-manual effort to fulfill. Both hands should stay on the steering wheel, information should be presented timely, in small chunks, and follow standards the driver is familiar with. The second principle is concerned with the appropriateness of the dialogue for a TICS task. It mainly requires the dialogue to be consistent in the way information is presented and the interaction is structured. Also, the interaction must be controllable by the driver, i.e., the driver initiates the start and end of the dialogue. Finally, the third principle requires the dialogue to be appropriate for the driver. It should be easy to understand, to learn, and match the driver’s expectations.

4.2 Applicable to Visual-Manual HMIs: Physical Controls & Touch-Based HMIs

4.2.1 Guidelines and Standards

Several broad guidelines exist for in-vehicle HMIs that list criteria for achieving a low-distraction interface and interaction. These guidelines are mostly concerned with visual-manual interfaces and cover similar topics. Typical visual-manual guidelines are the following ones:

- European Statement of Principles (European Union, 2008)
- Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems (Alliance of Automobile Manufacturers, 2003)

- Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (National Highway Traffic Safety Administration, 2013)
- Design Guidelines for In-Vehicle Display Systems – Version 3.0 (Japan Automobile Manufacturers Association, 2004)

Since the guidelines are rather similar in the topics covered, it is sensible to summarize them. Such a summary can be found in Horn et al. (2023) who developed a checklist prototype that we used as the basis for the development of our HMI checklist. This checklist prototype is largely focused on visual-manual distraction and employs the visual-manual NHTSA guideline (National Highway Traffic Safety Administration, 2013). As part of their theory, Horn et al. (2023) provide a guideline criteria summary of the four visual-manual guidelines listed above. Table 5 contains the principles listed across five categories. The link between the checklist prototype and our HMI checklist will be described in more detail in chapter 5.

Table 5: Visual-Manual Guideline Criteria Summary – Adapted from Horn et al. (2023)

Category	Principle
Installation	→ Installation guidelines must be followed / Correct installation → System must not obstruct driver's field of view → System must not obstruct driving-relevant controls and displays → Controls must be positioned so that operation does not require excessive deviation from driving position → Displays with relevant information should be mounted as close as possible/practical to the line of sight → Minimal glare and reflections (regarding night driving)
Information Presentation	→ Tasks can be completed with brief, sequential glances and quickly (max. single glance duration / max. cumulative glance duration) → Standards (guidelines / empirical data) for characters, sounds, symbols, acronyms etc. are followed → Driving-relevant information is presented precisely and without delay → Information with high safety relevance is prioritized → Mute option available / no audio output with uncontrollable volume level
Interaction	→ Maximum one-handed operation → Complete hands-free operation possible for speech-based systems → Interruptibility / option to resume → Driver determines pace of interaction (delayed inputs) → System controls do not impair the use of driving-relevant vehicle controls → Operation of the system does not obstruct vision → Volume of all potentially distracting auditory information can be controlled → System feedback occurs promptly and is clearly recognizable ($\approx 250\text{ms}$) → Non-safety-relevant dynamic information can be disabled
System Response	→ Distracting visual information without relevance to driving task (e.g., TV) is blocked → System behavior does not interfere with displays/controls relevant to driving safety → System clearly distinguishes between functions suitable and unsuitable for use while driving → Functions not suitable for use while driving are blocked → Safety-relevant status messages / malfunctions are reported to the driver promptly and correctly
External Information About the System	→ Adequate documentation / safety instructions → Manuals / instructions are correct and kept simple → Instructions are in language / form that is regionally understood

Category	Principle
	→ Instructions clearly distinguish between functions suitable and unsuitable for use while driving → Product information accurately describes system functions → Product information clearly highlights required competencies / limitations regarding user groups → Representations of use (images, descriptions) are realistic and do not encourage unsafe behavior

For the interaction with visual-manual elements, two ISO norms have been identified that contain specifications regarding the positioning of physical controls and the central touchscreen. Especially relevant for physical controls is ISO 4040:2009 *Road vehicles – Location of hand controls, indicators and tell-tales in motor vehicles* (International Organization for Standardization, 2009) that subdivides the area around the steering wheel into zones that certain physical control elements should use. Examples are indicator or wiper controls. As a result, it is mainly concerned with the lever controls of secondary task functions. ISO 3958:1996 *Passenger cars – Driver hand-control reach* (International Organization for Standardization, 1996) is relevant for physical and touchscreen controls as it stresses the reachability of the HMIs. Both aspects are already covered in Horn et al.'s (2023) guideline summary via the installation category.

For the presentation of visual information via a display ISO 15008:2017 *Road vehicles – Ergonomic aspects of transport information and control systems – Specifications and test procedures for in-vehicle visual presentation* (International Organization for Standardization, 2017a) may be used. Here, the following criteria can be extracted that give further guidance on the visual information presentation:

- Viewing position of the display must be optimal for the driver
- Brightness range should cover all brightness conditions – thus, brightness must be adjustable
- Contrast should be optimal at any brightness level
- Appropriate color combinations should be used
- Appropriate fonts shall be used
- Reflections must be reduced without reducing contrast
- Stability of the image (no flickering or jitter), blinking for alerts is allowed

4.2.2 Research

The research on visual-manual in-vehicle HMIs is mainly centered around touchscreens while also being driven by the automation trend in the automotive industry. Although our objective was the development of a checklist focusing on low-distraction HMIs for manual driving, design guidelines for autonomous driving may still be applicable for manually driven vehicles.

Naujoks et al. (2019) established design guidelines for partially-automated vehicles where a driver is always required in case the automation cannot continue driving. Thus, relevant system states must be communicated to the driver via an HMI. Some of the 20 design principles can also be applied regarding low-distraction visual interfaces in manually driven vehicles:

- Guideline #1: Unintentional activation and deactivation should be prevented.
- Guideline #4: Visual interfaces used to communicate system states should be located at a suitable position and distance. High priority information should be presented close to the driver's expected line of sight.
- Guideline #5: HMI elements should be grouped together according to their function to support the perception of mode indicators.
- Guideline #6: Time-critical interactions with the system should not require continuous attention.

- Guideline #7: The visual interface should have a sufficient contrast in luminance and/or color between foreground and background.
- Guideline #9: Commonly accepted or standardized symbols should be used to communicate the automation mode. Use of non-standard symbols should be supplemented by additional text explanations or vocal phrases.
- Guideline #11: Messages should be conveyed using the language of the users (e.g., national language, avoidance of technical language, use of common syntax).
- Guideline #12: Text messages should be as short as possible but as long as necessary.
- Guideline #13: Not more than five colors should be consistently used to code system states (excluding white and black).

Guidelines #4, #6 and #7 can also be found in the guideline criteria summary by Horn et al. (2023) and/or in ISO 15008:2017. Guideline #1 adds an important aspect to the system behavior when a driver tries to operate a function. If some functions are erroneously operated, they might introduce a major distraction to the driver (e.g., accidentally opening the windows while driving at high speeds) or endanger others (e.g., accidentally engaging the high beams and blinding others). Guidelines #5, #12, and #13 suggest that the visual interface should be tidy instead of cluttered. Finally, guidelines #9 and #11 are partially covered by the guideline criteria summary but highlight the use of symbols and text the driver is familiar with.

Regarding touchscreen focused research, Ma and Gong (2024) list five factors that influence how distracting an interaction with the central touchscreen is and make suggestions towards an optimal design:

1. Operation steps (number of actions required for the operation of a function): single step
2. Click displacement (total distance of finger movement for all operation steps): < 20 mm
3. Rightmost position of the operation area: < 35° - 40°
4. Average touchpoint area (average area of all icons that need to be tapped): 700 mm²
5. Horizontal rotation area (rotation of the display towards the driver): > 7°

Grahn & Kujala (2020) investigated the effects of the touchscreen size, user interface design, and subtask boundaries (i.e., typing a number in chunks vs. in one go) on driving performance. A larger touchscreen slightly decreased the distraction potential. More importantly, however, was the use of an automotive design with large fonts, icons, and a generally less cluttered design. This supports the Guidelines #5, #12, and #13 from Naujoks et al. (2019). The division into more subtasks generally decreased the distraction potential. Thus, it is important that a function always waits for driver input and does not automatically continue, which aligns with ISO 15005:2017 and the interaction section in the guideline criteria summary.

Liu et al. (2022) researched how clicking and sliding gestures can be optimally executed on the central touchscreen. They suggest that the optimal button size is 14.4 mm with a “hot zone” size (clickable area of the button) of 13.2 mm x 13.2 mm for an accuracy of 80%. Furthermore, buttons should not be too close to the steering wheel as they get occluded by the arm, while at the same time they should not be too far to the right. For sliding gestures, the middle of the screen is best for the operation of sliding gestures. Furthermore, sliding from left to right and from top to bottom is easier than from right to left and from the bottom to the top, respectively. In comparison to Ma and Gong (2024), the touchpoint area seems to be smaller and the aspect of the touchscreen being too close to the steering as well as the investigation of sliding gestures are helpful additions for the central touchscreen’s interface design.

Nicolas et al. (2024) studied depth and breadth trade-offs for different tasks on a central touchscreen while driving. Regardless of the task, a broad menu where many elements are on a single page is generally preferable to an interface with a high depth menu. This indirectly supports Ma and Gong’s (2024) results as with a broad interface less interaction steps are needed. Of course, at the same time, the interface should not be too cluttered as with Naujoks et al.’s (2019) guidelines.

Finally, Ebel et al. (2023) used natural driving data to predict driver distraction. The data suggests that UI elements should be placed closely together, that static and easy-to-use elements should be chosen, and that lists should not be used. The results largely align with the findings of Ma and Gong (2024) as well as Nicolas et al. (2024).

4.3 Applicable to Speech-Based HMIs

4.3.1 Guidelines and Standards

While visual-manual HMIs have well-established guidelines (see section 4.2.1), speech-based systems in automotive contexts still lack these kinds of resources. This may be attributed to the fact that the cognitive distraction induced by speech-based systems is generally considered less severe than visual-manual distraction (Zhang et al., 2023). Furthermore, speech-based interfaces are still a rather new development in vehicles, especially for controlling functions with a low proximity to the driving task (Bubb et al., 2015).

However, there is one standard that can be used as a foundation for the design of speech-based HMIs: SAE J2988 *Guidelines for Speech Input and Audible Output in a Driver Vehicle Interface* (Society of Automotive Engineers, 2015) gives guidance on when to implement a speech-based interface, the dialogue structure, error handling, and system feedback. For the design of the audio signals, ISO 15006:2011 *Road vehicles – Ergonomic aspects of transport information and control systems – Specifications for in-vehicle auditory presentation* (International Organization for Standardization, 2011) can be used to specify the properties of the auditory channel. Both standards can be employed to derive a criteria summary that is similar to the one of Horn et al. (2023) for the visual-manual guidelines. The criteria summary can be found in Table 6.

Table 6: Speech-Based Standard Criteria Summary

Category	Principle
Audio Signal	→ System uses a well-audible frequency range → Auditory signals must stand out from background noise → Auditory levels must not be too high
Information Presentation	→ Information may be coded via chimes and/or speech → Not too many chimes shall be used → Prioritization of information with high relevance → Visual information may be presented for redundancy → Speech shall use simple vocabulary → Auditory information should be as short as possible
Interaction	→ Universal commands are used for common actions (e.g., “help”, “cancel”, “repeat”, “main menu”) → As little dialogue steps as possible shall be used → System should understand as many utterances as possible → A dialogue can be easily initiated either via a button on the steering wheel or by using a keyword → System indicates when it is ready to receive an input → System informs about the acceptance/rejection of an input → System informs about possible delays → System can be interrupted or the whole dialogue stopped during any stage of the interaction → Driver determines the pace of the interaction
System Response	→ System does not allow the operation of functions that are not suitable for the speech-based interface (e.g., functions affecting the movement of the vehicle) → System features adequate error correction and handling

Category	Principle
Additional Information About the System	→ System informs the user about speech-accessible functions and the corresponding vocabulary

4.3.2 Research

Speech-based options for in-vehicle HMIs are a popular topic in automotive research, especially as they enable low-distraction interactions – at least when compared to visual-manual HMIs. They are not distraction free and may induce cognitive distraction (see section 4.4), which should be reduced or omitted. To achieve this, Hua and Ng's (2010) study can be applied that lists the following four guiding principles:

- use a broad and shallow menu structure (no more than three menu levels)
- provide visual feedback and memory aids
- provide vocal shortcuts
- implement hard keys for often used tasks

The use of a broad and shallow menu structure resembles the depth and breadth trade-off from visual-manual interfaces (Nicolas et al., 2024). For speech-based interfaces, the use of a broad menu structure minimizes the number of dialogues needed, which is a principle that can be found in the interaction section in Table 6. Similarly, the provision of visual feedback aims to minimize the cognitive load. Of course, visual feedback may result in visual distraction, so the visual feedback should be implemented carefully. The provision of vocal shortcuts is also useful to reduce the number of dialogues needed. Finally, the authors recommend the use of physical controls for regularly performed tasks. This makes sense, as physical controls are often the optimal way to interact with simple functions (Ma et al., 2022). But this indirectly addresses the need to provide access to the function via other means (physical controls or touchscreen) in case the speech-based interface is not operational.

There are other collections of design principles for speech-based interfaces as well, either for the use in vehicles (e.g., Murad et al., 2018) or for speech-based interfaces in general (e.g., Suhm, 2003). However, they largely match the criteria listed in Table 6 and/or the guiding principles of Hua & Ng (2010).

A group of studies has investigated the effects of the speech style used by the speech-based interface. Especially in cases where a user has to gain situation awareness (see Endsley, 1995), an informative (simple and direct, more “commanding”) speech style is better than a conversational (more suggestive; Wang et al., 2022) one. Furthermore, reaction times improve when a voice with a higher level of assertiveness is implemented (Wong et al., 2019). This is presumably more relevant in cases belonging to second level tasks (Bubb et al., 2015), for example, when a driver needs to react within a few seconds (e.g., turning on the light). Beyond speech style, voice agents with different personas might be employed (Braun et al., 2019). However, it is unclear how this affects distraction.

Finally, the study of Eriksson et al. (2014) demonstrates that the speech-based interface should omit menu structures and allow direct access to a function via natural voice commands. Further, the speech-based interaction must be paced by the user to allow for compensatory strategies and to reduce cognitive load. A final aspect that might lead to high levels of cognitive distraction is when the speech-based interface does not react in a way expected by the user (Reichel et al., 2014).

4.4 Modality Choice for Different Tasks

The previous sections 4.1 – 4.3 described design principles for each HMI type that reduce its distraction potential and make the interaction easier. However, that does not answer the question of which of the three HMI types (physical controls, touch-based, or speech-based) should be used to access a function.

Ma et al. (2022) investigated the effects of secondary tasks on driving performance and distraction when being operated via different modalities. Ten secondary tasks that were classified into three classes were used: basic, medium, and advanced. The classification is based on two dimensions: complexity (number of operation steps) and difficulty (average interaction per step). The results reveal that steering wheel buttons and speech-based operation are most suitable for basic tasks (adjust volume, skip to next track, turn on recirculated air, accept incoming call). The speech-based interface is the most ideal for medium tasks (change media source, call contact from list, increase temperature, increase air volume). For advanced tasks, a combination of touch- and speech-based input should be used (type in a phone number and call, select a destination from the destination history). The study highlights that for simple tasks, physical controls enable blind operation, thus little visual distraction is induced. For medium tasks, the speech-based interface strikes the balance between good operation speed and low visual distraction. Advanced tasks, however, should never be operated via physical controls – here, the speech-based interface seems to be a good option to omit visual distraction although it lacks high operation speeds. Thus, the addition of a touchscreen interface, which is used in parallel, may accelerate the interaction, while conversely, introducing visual distraction. As a result, a middle ground must be found when using both interfaces for advanced tasks.

In a similar fashion, Zhang et al. (2023) compared three different input modalities: touchscreen-based, gesture-based, and speech-based. Although the difficulty of the secondary task was varied, this had no effect on driving performance. Instead, the effects could be largely attributed to the different HMI modalities. As expected, the touchscreen-based interaction led to the highest distraction and worst driving performance. For gesture-based input, which used in-air gestures, effects were less severe when compared to the touchscreen-based interaction, although the crash risk was still elevated. Finally, speech-based interaction affected the driving and visual performance the least. The authors suggest that the speech-based interface should be used as the primary input for most tasks and that gestures might be deployed for commonly accessed functions. The interaction with the touchscreen should be avoided if possible. The study follows the trend of Ma et al. (2022) but adds the option for gesture input.

In other studies, speech-based interfaces are the preferred option as well for operating multiple-step functions when compared to visual-manual interfaces (e.g., for texting: He et al., 2014; Kurtz et al., 2021; e.g., for address inputting: Zhong et al., 2022). However, even the speech-based interface can introduce significant levels of cognitive distraction that are detrimental to safe driving. Based on his study results, Strayer (2015) suggests that a red line should be drawn for certain tasks even if they are operated via a speech-based interface. If the cognitive distraction reaches a high enough level, it interferes with the driver's ability to allocate the attention towards relevant stimuli. This results in longer response times and worse driving performance. This claim can be further backed up by studies showing that drivers' gaze does not stay on the road when interacting with a speech-based agent (Zhang & Roberts, 2023). Note, however, that speech-based interaction may have improved in the past years. In the past, when interacting with speech-based agents (e.g., Apple's Siri) cognitive distraction was high due to many errors occurring (Strayer et al., 2017). In a more recent replication of this study, cognitive distraction levels were significantly lower (Loew et al., 2023). This might be attributed to technological advancements in speech-based interfaces.

In conclusion, depending on the complexity of the task, appropriate HMI types (physical controls, touch-based, speech-based) should be used to control certain functions. For highly complex tasks, the studies suggest that they always have a high potential for driver distraction – even when they are operated via the speech-based interface. Although beyond the scope of the checklist, it may be advisable to entirely restrict access to those functions while driving.

5 HMI Checklist

As stated in the introduction, the objective of this project was to design an HMI checklist that can be employed to assess the function implementation in the vehicle regarding its distraction potential and whether it is safe-to-use, i.e., operation errors are unlikely to occur.

The foundation of our work was the checklist of Horn et al. (2023), which was developed on the basis of the NHTSA-Guidelines. There, the checklist items are split into two modules for hardware and function aspects to ensure an efficient evaluation process. Further, the scope of functions that can be assessed was expanded by the inclusion of a third module in the form of a decision tree. As a result, functions close to the driving task (secondary tasks; Bubb et al., 2015) can be considered as well. Our newly developed checklist includes the same range of vehicle functions: It allows the assessment of the HMI implementation of secondary and tertiary tasks, but not primary tasks (i.e., steering, throttle, brake).

We enhanced the prototype of Horn et al. (2023) in three crucial ways. First, we extended the scope of the checklist so that it can be used to assess speech-based HMIs. At the same time, we further split the checklist so that each HMI type (physical controls, touch-based, and speech-based interfaces) can be evaluated individually. Secondly, we added a quantitative scoring system to generate an HMI operation score pointing towards issues in the HMI's design. Here, the deviation from an ideal HMI implementation is determined on an item level. Thirdly, the decision tree was revised to point towards the optimal HMI implementation for a function given its usage context. As a result, the ideal implementation for secondary (e.g., turning off the high beams) and tertiary functions (e.g., inputting a navigation address) can be determined. The ideal implementation is comprised of the HMI modality and the operation steps needed for the function operation.

The final HMI checklist can be found in Appendix A. It was validated in a driving simulator experiment described in chapter 6. The results (see chapter 7) revealed that the HMI types interact with task complexity, therefore confirming the basic assumption of the decision tree (see section 5.2.3). More importantly, the validation was used to ensure that the design aspects (see section 5.2.2) of each HMI modality do indeed affect driver distraction and function operation.

5.1 Checklist Structure and Assessment Process

The basic structure of the previous instrument (Horn et al., 2023) has been retained. However, the items in each module were split into three groups for each HMI type (physical controls, touch-based, speech-based). This was necessary as the characteristics of each HMI modality affecting distraction and safety-to-use differ from one another. The general checklist structure can be seen in Figure 2.

The assessment process starts with the definition of a function group or an individual function to be evaluated with the instrument. At this point, the HMI type/s through which the function can be operated is/are determined (physical controls, touch-based, speech-based).

Then, each utilized HMI type is evaluated using the items in module I *Shared HMI Features*. This first module is a follow-up of Horn et al.'s (2023) idea to only assess features of the HMI that are shared between functions once. Especially for the touch- and speech-based HMIs, this reduces the assessment effort as unnecessary repetitions are avoided. For example, every function implemented in the touchscreen uses the same hardware, therefore the physical touchscreen device only needs to be evaluated once. When each item in every employed modality is evaluated, a first HMI operation score can be calculated. A higher score points towards problems in the shared HMI features.

Next, and with module II, the focus is shifted away from the HMI towards the function. Module II a *Optimal Function Implementation* only looks at the function itself and the use context in which it is operated. From there, the optimal implementation can be derived that determines the suitable HMI modalities (physical controls, touch-based, speech-based) and operation steps for the function. If the function's current HMI implementation (sufficiently) matches the optimal implementation, module II b *HMI Function Interaction* and its items can be used to evaluate the HMI function implementation and the interaction with it. Like with module I, an HMI operation score is generated pointing towards problems in the function's HMI implementation.

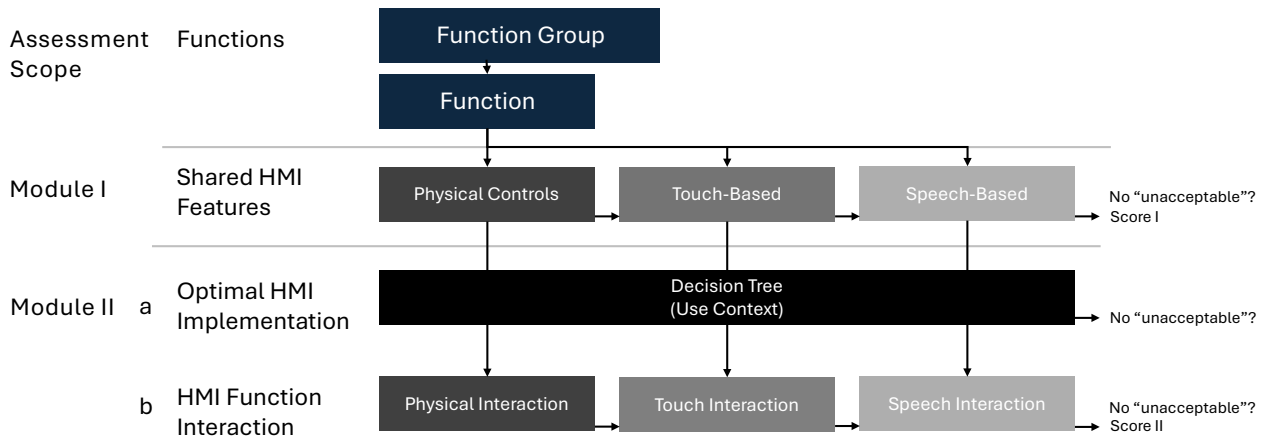


Figure 2: Checklist Structure

The basic scoring principle is shown on the right side of Figure 2. For each module, a rating of “unacceptable” must not be attained. In case this happens, the assessment of the HMI type (module I) or function (module II) must be stopped, and the interface problem should be fixed immediately. Only if no design aspect in module I and II b and the implementation of the function regarding its use context (module II a) reaches an “unacceptable”, the assessment can be continued, and valid scores be generated.

5.2 Development and Description of Checklist Parts

In comparison to the previous checklist of Horn et al. (2023), our instrument does not include a “preceding” part that includes a detailed instruction and description of the checklist’s scope and limitations. Both aspects are instead covered by this report. Important definitions and basic instructions, however, are included to establish the basic working principle of the instrument. Furthermore, the checking routine is described in detail at the item level. The instrument is aimed at professionals familiar with vehicle interface design, therefore, the goal was to extensively focus on the contents of safe function implementation.

5.2.1 Assessment Scope

The assessment scope is used to define the function(s) that shall be assessed regarding their implementation quality. It specifies that the checklist is intended for assessing HMIs that are part of the vehicle, but not HMI extensions like Apple CarPlay or Android Auto. This was done for two reasons: first, tethering options only include tertiary task functions. The focus of our checklist, however, is to also consider functions with a high proximity to the driving task belonging to the secondary task class. Second, tethered functions add another base layer to the vehicle’s HMI concept that works differently from the one that was implemented by the manufacturer. This would require another set of items for assessing the shared HMI features in the tethered interface. For example, a manufacturer’s speech-based interface has different base properties (e.g., recognition rates, speech output quality) compared to the speech-based controls of a connected mobile device.

To determine the function(s) to be evaluated, examples from different function groups are provided. These are directly taken from the function cluster for optimal operability (see Figure 1) that was derived from our vehicle function research (see sections 3.2. and 3.3). This was done to ensure a mutual understanding of what classifies as a function, thereby standardizing the HMI assessment process.

Finally, the HMI type(s) used by the function(s) are determined. For each HMI type, a detailed definition is provided. Here, a description of the physical devices involved (not for speech-based interfaces), the number of functions accessible, and details on the interaction are given (see chapter 3). Similarly to the provision of functions, this is a measure to create a common understanding of the terminology used, thus acting as a form of standardization.

5.2.2 Items and Quantitative Scoring (Module I and II b)

In contrast to the previous checklist developed by Horn et al. (2023), our instrument does not strictly adhere to a singular interface design resource (see chapter 4). Instead, we used the visual-manual guideline criteria summary (see Table 5) as a starting point to generate the items (design aspects) in module I and II b for the visual-manual interfaces (physical controls, touch-based). Aspects not already covered by the guidelines were derived from individual studies (see section 4.2.2). A similar approach was chosen to generate the items for assessing the speech-based HMI in module I and II b. The speech-based standard criteria summary (see Table 6) acted as a starting point with further aspects derived from individual study results (see section 4.3.2). For each item, the references can be found in the last row of the item body (see Appendix A). After validation, only items that affected driving performance or the operation process were included in the checklist. Items with no effect on either dimension were excluded (see chapter 7). The validated version of the checklist is comprised of 50 items.

The basic item structure is made up of eight elements (From top to bottom, see Figure 3):

- Item number and short title for easy orientation
- Item text describing the optimal HMI principle
- 5-point rating scale used for scoring
- Affected dimensions compared to a reference version; see section 7.3.1
- Weighting (value: 1.33; range: 1 – 1.9) indicating the significance of the design aspect; see section 7.3.1 for an explanation of the scoring system (the weighting factor is derived from the penalty score)
- Explanation to create a clear understanding
- Procedure that shall be used for the assessment
- References indicating the source of the item

M1-PC-4 – Reachability				
The control element has been installed in a position where it can be easily reached by the driver.				
Both hands on the steering wheel.	One hand on the steering wheel, arm not extended.	One hand on the steering wheel, arm slightly extended.	One hand on the steering wheel, arm widely extended.	Both hands necessary.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 1)	+ (0.05 × 4.5)	= 1.33
Explanation	The control element should be close to the driver so that minimal motor effort is needed.			
Testing	Put both hands on the steering wheel. Now operate the control element and pay attention to your hands and whether one hand needs to be removed from the steering wheel. If so, pay attention to how much you need to reach for the control element.			
References	Visual-Manual Guideline Criteria Summary – Installation, Interaction ISO 3958:1996 & ISO 15005:2017			

Figure 3: Item Structure

With guidelines only listing general principles, it often remains unclear how exactly a good implementation can be achieved. Hence, we used lengthier item texts and explanations to ensure that the HMI principle can be properly operationalized. Furthermore, the rating scale includes rating points that are described in detail, resembling a behaviorally anchored scale. In other words, this leaves little room for the interpretation of the principle itself and its optimal implementation.

For quantitative scoring, a five-point rating scale was created. It ranges from “0”, resembling an optimal HMI function implementation, to “unacceptable”, indicating that the implementation is a severe safety hazard. If an “unacceptable” occurs the assessment process must be stopped. Values ranging from “1” to “3” indicate deviations from the optimal implementation on varying levels. The cell “Weighting” is calculated based on the dimensions (driving performance and/or operation quality) that were negatively affected in comparison to a reference version during validation. For details on the validation process and the calculation of the weighting score, see section 7.3.1. The weighting factor indicates the importance of the design aspect regarding driver distraction. It acts as a multiplier for the rating on the five-point rating scale.

For some modalities in module I and II b, bonus items exist that use a different scoring system. They contain criteria that cannot achieve the status “unacceptable”. Here, the five-point rating scale ranges from “- 4” (principle fully respected by the HMI’s design) to “0” (principle not implemented). Thus, the bonus item can be used to lower the HMI operation score. The structure of a bonus item is shown in Figure 4.

M2-TB-9 – Shortcut				
The operation of the function can be shortened by skipping individual operation steps.				
Function can be operated via a shortcut with a single operation step.	Function can be operated via a shortcut with fewer operation steps.	Function accessible via physical controls.	Function accessible via speech-based HMI.	Function does not offer a shortcut.
-4 ☐	-3 ☐	-2 ☐	-1 ☐	0 ☐
Affects				
Weighting	1.0	$+ (0.1 \times 0)$	$+ (0.05 \times 0)$	= 1
Explanation	The operation of the function can be shortened by using a shortcut (touch-based HMI, physical controls, and/or voice-based HMI) which skips operation steps.			
Testing	First try to create a shortcut for the function on the home screen. Then check whether you can operate the function so that the final operation step can be accessed directly via this shortcut. Also check to what extent the function can also be accessed via physical controls and the voice-based HMI.			
References	Ebel et al. (2023)			

Figure 4: Example of a bonus item.

At the end of module I and II b, the item scores are summed up to an HMI operation score. The higher the score, the more problematic the HMI implementation of the function is. Insufficient implementations are associated with high levels of distraction and/or poor operation safety. The possible point range differs per HMI type as the number of items as well as the assigned weighting factors differ. For physical controls, a maximum of 45 points can be scored, for the touch-based HMI 78 points, and for the speech-based HMI 60 points. Please note that these are sums of the maximum scores across module I and II.

5.2.3 Optimal HMI Implementation and Use Context (Module II a)

A function itself is used to accomplish a certain task in the vehicle. This raises the question of which HMI type (physical controls, touch-based, speech-based) should be used for function operation. The decision tree in module II b is designed to determine the optimal implementation of a function given its use context. Optimal implementation refers to the HMI modality and operation steps that shall be used. Thereby, two objectives can be achieved. Firstly, from a safety perspective, optimal operability of the function can be ensured, especially if it is time-critical (e.g., turning off a malfunctioning driving assistant). Secondly, the limited space of easily accessible single-step interfaces (mostly physical controls, see Figure 1 for optimal operability towards the top left-hand corner) can be distributed towards those functions that are time-critical.

The decision tree was derived from of Horn et al. (2023) and redesigned to include operation errors and the distraction potential as well as safety hazards that result from it. Moreover, all HMI types/modalities (physical controls, touch-based, speech-based) are considered at every of the five implementation levels. Also, the evaluation process is now outlined in the decision tree itself. The decision tree is shown in Figure 5. The decision tree was in parts validated as the results (see chapter 7) demonstrate that low and high complexity tasks should be operated with a suitable HMI modality.

The determination of the optimal implementation and whether it is met by the actual implementation works in three steps. First, the optimal implementation itself is determined via answering the questions along the paths of the decision tree. There are four questions in total, two or three of which need to be answered depending on the path:

- Question 1: What is the cause for the operation of the function?
 - external factors (relating to the driving task): e.g., using turn signals due to infrastructure, turning off the cruise control due to an obstacle, turning on the windshield wipers due to rain
 - internal (voluntary by the driver): e.g., changing the song playing, adjusting the air conditioning temperature, answering a call
- Question 2: How much time is available for executing the function? When must it be completed?
 - instantly: e.g., braking, avoiding obstacles
 - a few seconds: e.g., setting the cruise control speed, turning on the windshield wipers, turning on lights
 - longer term: e.g., adding a gas station to the navigation, setting an alternative route to avoid traffic congestion
- Question 3: How frequently is the function used?
 - often: e.g., adjusting the radio volume, climate control operations
 - rarely: e.g., changing settings in configuration menu, pairing a new phone
- Question 4: How high is the risk for and how severe are the consequences of an operation error?
 - high risk: interference with lateral and longitudinal control (e.g., turning of driving assistance), strong distraction (e.g., accidentally turning the radio very loud)
 - low risk: slight distraction (e.g., accidentally setting the ventilation to full strength)

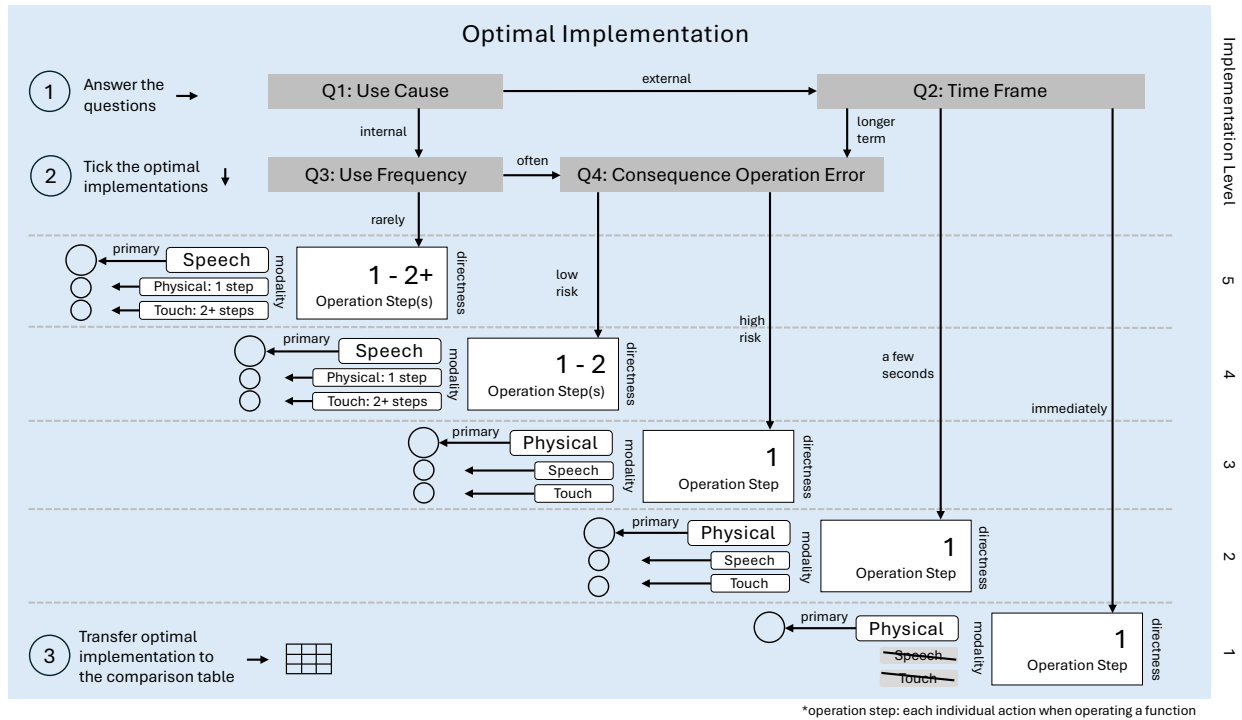


Figure 5: Decision tree for determining the optimal implementation

Table 7: Example Functions on Each Level of the Decision Tree

Level	Path	Functions
5	"internal" → "rarely"	→ navigation: insert and navigate to full address → phone: dial and call a phone number
4	"internal" → "often" → "low risk"	→ climate control: adjust airflow → media control: skip to next track
	"external" → "longer term" → "low risk"	→ navigation: route to gas station → climate control: turn on recirculation
3	"internal" → "often" → "high risk"	→ climate control: adjust fan speed → media control: adjust volume
	"external" → "longer term" → "high risk"	→ windows: open/close side windows → disable driving assistance
2	"external" → "a few seconds"	→ sight: clear windscreen → signals: Turn on indicator
1	"external" → "instantly"	→ warning signal: use horn → light: disable high beams

After answering the questions, the path in the decision tree ends at one of the five implementation levels. Here, the directness (number of acceptable operation steps) as well as the acceptable HMI modalities (physical controls, touch-based, and/or speech-based) can be derived. Both indicate the optimal implementation of the function. Table 7 contains example functions and paths for each implementation level. For example, the horn is a warning signal that is located at the first level of the decision tree. As a result, only physical controls are allowed but no other input modalities (touch-based, speech-based) and operation must happen in a single step. In contrast, inserting a navigation destination would lead to the fifth implementation level. Here, speech-based input should be used as the primary modality and touch-based input as the secondary modality since multiple steps are required for operation.

At the fifth level, multiple operation steps are allowed. The distinction between primary and secondary HMI modality can be found at every level at the decision tree. The idea is that the primary HMI modality should be used as the primary way to access a function. Secondary modalities either act as backup (for example when speech-based input is not feasible) or additional ways to control the function. As the first implementation level includes safety-relevant functions (e.g. horn, lights) no secondary modalities are allowed and only physical controls must be used. On the higher levels, secondary modalities are allowed and on levels four and five these should be suitable for the task's complexity. Functions with multiple steps (e.g., navigation address entry) should use touch-based input instead of physical controls as the secondary modality. Only if only one step is needed (e.g., dialing via a shortcut), physical controls are recommended as the secondary modality.

In the second step, the actual implementation of the function, i.e., through which HMI type and with how many operation steps it can be accessed, is compared to the required, optimal implementation. This is done in a comparison table that structures the decision process. First, the number of operation steps is checked. Only at levels 1 – 3, this can result in an “unacceptable” if more than one operation step is used. At levels 4 and 5 more steps are deemed “suboptimal”. Ideally, the same number of steps is used. Next, the use of the primary modality is checked. Again, at levels 1 – 3, the primary modality must be used whereas at levels 4 and 5 this is optional but not using the primary modality results in a “suboptimal”. Finally, at levels 4 and 5, if the primary modality is not used, the secondary modality must be suitable for the complexity of the function. Otherwise, the function implementation is “unacceptable”. Note that the final step of the comparison also includes a check that no forbidden secondary modalities are used for level 1 functions.

As part of module II b, a detailed definition of the used terms (“modality”, “directness”, and “operation steps”) is provided to create a common understanding. This approach is similar to the provision of definitions in the assessment scope (see section 5.2.1).

Within each level, the modalities are arranged in a hierarchy. The primary modality is listed first, followed by the two remaining modalities acting as secondary modalities. To create this arrangement, the studies listed in the chapter on modality choice (see section 4.4) were used. On levels 5 and 4, speech acts as the primary modality, which is followed by touch as the secondary modality for complex tasks (2+ steps) and physical controls for simple tasks (1 step). This follows the recommendation of Ma et al. (2022) to use a multimodal approach of speech- and touch-based HMIs for very complex tasks. Physical controls were not recommended for complex tasks. The validation process revealed this idea to be correct as speech- and touch-based interfaces are most suitable for complex tasks. From levels 3 to 1, physical controls are used as the primary modality as all functions are safety-relevant either due to their distraction potential (level 3) or proximity to the driving task requiring timely reactions (level 2 and 1). Furthermore, at levels 3 and 2, speech is used as the first-level secondary modality before touch as it allows for a function operation without using visual-manual resources. This was largely validated by the validation process as speech-based interfaces offer nearly distraction-free function operation. Physical controls and touch-based interfaces performed on a similar level in the validation, however, as our approach of physical controls was very specific, the results must be treated cautiously. Therefore, and based on assumptions of optimal operability (see section 3.5; Figure 1) as well as the literature on modality choice (section 4.4), physical controls are preferred over touch-based interfaces as the secondary modality. On level 1, touch- and speech-based interfaces must not be used because of the safety criticality of these functions.

6 Empirical Validation: Method

The checklist is intended to serve as an evaluation tool for in-vehicle HMIs used for the operation of functions both related and unrelated to the driving task. For this purpose, we developed items based on existing criteria collections that we deemed to be relevant for low-distraction and safe-to-use HMIs. Moreover, to enable a quantitative assessment, a five-point rating scale was designed for every item.

An empirical validation employing a driving simulator experiment and generic HMIs was carried out to accomplish three objectives. First, to verify that a given item, on a general level, is relevant regarding driver distraction and safe function operation. Second, for items that are found to impact driving performance, their relative effects were compared and a weighting factor representing their impact was assigned. The third objective lies in comparison of modalities (physical controls, touch-, and speech-based HMIs) with respect to the operation task. According to the literature (see chapter 4) for a simple task, driving performance should be best for physical controls and worst for the touch-based HMI. For complex tasks, however, the speech-based HMI should be the best, followed by the touchscreen and physical controls performing the worst.

The driving simulator study will be described in more detail in the following sections. This includes descriptions of the experimental design and materials that were used. The results of the study are described in chapter 7.

6.1 Participants

The final sample is comprised of $N = 130$ participants (74 male, 55 female, 1 not specified). On average, they were $M = 39.8$ years old ($SD = 19.0$ years; range: 18 – 81 years). They held their driver's license for $M = 22.5$ years ($SD = 18.9$ years). Most participants (30 %) reported an annual mileage between 12,000 and 30,000 km. Most of them described their driving style as rather calm (53.1 %) and rather considerate (73.1 %).

The experiment was conducted in 13 groups of $n = 10$ participants for an appropriate experimental length (see section 6.2). There were no differences in gender split between groups, however, there was a difference in age ($p = .046$). The average age of participant groups ranged from 30 to 47.4 years. To account for potential biases, countermeasures were taken in data analysis, which are described in section 7.1.

Participants were recruited via the participant pool of TU Braunschweig. The recruitment strategy aimed to ensure a diverse sample with respect to age distribution and prior (in-vehicle) HMI experience. All participants had to possess a valid driving license to take part in the study.

6.2 Design and Measures

The experiment incorporates a three-factorial mixed-design (HMI type (3) x operation task (2) x checklist items (on average 17 per modality)). The experimental design is shown in Table 8. In total, there are 51 HMI checklist items, comprising 12 items for physical controls, 23 items for touch-based interfaces, and 16 items for speech-based interfaces. Each item represents a different interface design aspect within an HMI type. However, participants did not experience all combinations of the three factors. Instead, only four items of a single HMI type were tested in group of $n = 10$ participants (except for one group where only three items were tested). As a result, 13 participant groups were needed to cover all 51 items. This was done to keep the length of the experiment at roughly 90 minutes and to prevent fatigue.

For each item, an optimal implementation (score of 0 on the five-point rating scale) was compared against the suboptimal implementation (score of 3 on the five-point rating scale). To achieve this, generic interfaces were deployed. The optimal implementation was tested in a reference version where all items (interface design elements) of the HMI modality were optimal. Thus, a single measurement simultaneously captured the performance level for

all items at the optimal level. For each suboptimal implementation, however, a separate measurement was required per item, with all remaining items held at the optimal level.

Table 8: Experimental Design

HMI Type	Group (n = 10)	Reference Version (Every item per HMI type optimally implemented)	Items tested (Suboptimal Implementation per individual item)
Physical Controls (12 items)	1	All 12 items optimal	M1-PC-1, M1-PC-3, M1-PC-4, M1-PC-5
	2		M1-PC-2, M1-PC-6, M2-PC-1, M2-PC-4
	3		M2-PC-2, M2-PC-3, M2-PC-5, M2-PC-6
Touch-Based HMI (23 items)	4	All 23 items optimal	M1-TB-3, M1-TB-5, M1-TB-10, M1-TB-11
	5		M1-TB-1, M1-TB-2, M1-TB-4, M1-TB-6
	6		M1-TB-7, M1-TB-8, M1-TB-9, M1-TB-12
	7		M1-TB-13, M2-TB-1, M2-TB-3, (no item code: Gesture Control)
	8		M2-TB-2, M2-TB-4, M2-TB-6, M2-TB-7
	9		M2-TB-5, M2-TB-8, M2-TB-9
Speed-Based HMI (16 items)	10	All 16 items optimal	M1-SB-3, M1-SB-5, M1-SB-6, M1-SB-7
	11		M1-SB-1, M1-SB-2, M1-SB-4, M2-SB-3
	12		M2-SB-2, M2-SB-5, M2-SB-6, M2-SB-7
	13		M2-SB-1, M2-SB-4, M2-SB-8, M2-SB-9

Note. Items referenced by their item code. For a full description see Appendix A. Group = Participant group.

Please note that the implementation of the design elements in the generic interfaces (see section 6.3.2) is just one specific variant of an interface design in a very specific simulator setting. Even with the checklist, there remains a high degree of freedom in implementing functions into HMIs. Therefore, the results need to be interpreted cautiously. For improved clarity, the specific operationalization of each tested design aspect is provided in Appendix B.

Two generic function operation tasks were employed that were derived from the decision tree (see section 5.2.3; Figure 5). The first task, indicating at a lane change, corresponds to the first level of the decision tree. The use cause is external and the time frame requires immediate operation. The second task, dialing a phone number and calling it, corresponds to the fifth level of the decision tree. The use cause is internal and the task is rarely executed. As a result, the first task is of low complexity and can be done in a single step whereas the second task is of high complexity requires multiple steps. Besides the two function tasks, the operation of a radio with physical controls was included to act as a benchmark.

Dependent measures included both driving and function operation performance indicators. Driving performance was measured via the Lane Change Task, a standardized measure of distraction (see section 6.3.1; International Organization for Standardization, 2010). It consists of the two sub-tasks of lane keeping and lane changing (see Huemer & Vollrath, 2010). For both, the mean deviation from an ideal trajectory was calculated. For lane keeping, the standard deviation from the ideal trajectory was additionally calculated. For the lane change, the reaction time was additionally measured from the point when the lane change sign became visible to the first steering input. Operation performance was measured by the experimenter who counted the number of tasks completed and the number of errors. After each trial, participants were asked to evaluate how well they could operate the function and as how distracting they perceived the operation process. The measures collected by the experimenter represent objective operation quality, the measures from the participant subjective operation quality. All dependent measures and how they were analyzed are described in more detail in section 7.3.1.

6.3 Materials

6.3.1 Lane Change Task

The Lane Change Task (International Organization for Standardization, 2010) is a standardized procedure designed to quantify the secondary task demand when interacting with in-vehicle HMIs. The task employs a simplified driving simulation in which participants drive along a three-lane road with no other traffic. They are instructed to keep a constant speed of 60 km/h, which cannot be exceeded. The lane changes are indicated by traffic signs positioned at both sides of the road. These appear at regular intervals (approximately 150 m apart) and become legible at a distance of 40 m. Once the driver has recognized the target lane from the sign, they must execute the lane change as quickly and accurately as possible.

Each trial consists of 18 randomized lane changes, combining different directions and lane change magnitudes (one or two lanes), and takes approximately 3 minutes to complete. The task can be split up into two sub-tasks: lane keeping segments between signs, similar to tracking tasks, and lane change segments that function as choice reaction tasks (see Huemer & Vollrath, 2010). To measure driving performance, the mean deviation from the ideal trajectory is calculated. For the sub-tasks, additional measures are collected. For the lane change, reaction time from the point when the lane change instruction becomes visible to the first steering input is measured. For lane keeping, the standard deviation from the ideal trajectory is computed.

While the LCT can be used to successfully differentiate between various types of distraction (e.g., Kurtz et al., 2021; Maciej & Vollrath, 2009), research has found several limitations regarding its validity. One concern is the considerable variation in baseline performance across different studies. For example, Young et al. (2011) reported persistently high mean deviations of 1.6 m. In other studies, the baseline performance is much lower with a mean deviation of only 0.64 m (Weir et al., 2007). Therefore, it is difficult to compare the results across different studies. However, this concern is of minor relevance here, as performance was not compared against other studies' baselines. Instead, participants' dual-task performance was compared against their own baseline via computing change scores (see section 7.3.1). Statistical control for differences between participant groups was additionally achieved through mixed models with random effects at the individual and group level.

Recent research by Morgenstern et al. (2023) indicates that the LCT's standard parameter (mean deviation) is less sensitive in discriminating between different types of distraction (visual-manual vs. cognitive) compared to newer methods such as the Box Task (Trommler et al., 2021) combined with the Detection Response Task (International Organization for Standardization, 2016). However, for the present study, this limitation is of minor relevance: the primary aim was not to separate distraction types, but to detect overall performance differences between HMI types and optimal and suboptimal HMI implementations. Moreover, the DRT itself represents an additional secondary task that could interfere with the HMI performance measures, and it explicitly excludes the measurement of primary driving task performance (such as lane keeping), which is a central outcome in the present study.

Furthermore, learning effects have been documented in several studies (Petzoldt et al., 2014), affecting both LCT performance and secondary task execution. These learning effects may negatively impact test results when participants are repeatedly recruited for LCT studies. However, in the present study each participant was only tested once, so carry-over learning effects across studies are not a concern. Within-session learning was mitigated by following the training recommendations of ISO 26022:2010: participants completed four LCT training trials (without a secondary task, thus driving only) and additionally practiced the secondary tasks (without driving) and the dual-task combination before data collection began (see section 6.5).

Despite these limitations, the LCT remains a well-established and appropriate tool for the purposes of the present study. Originally developed by Mattes (2003) within the Advanced Driver Attention Metrics (ADAM) project, it was explicitly designed to investigate how HMI distract the driver, a goal which aligns with the present study's aim of comparing the effects of HMI types and design aspects on driver distraction. Its adoption as ISO 26022:2010

(International Organization for Standardization, 2010) reflects methodological consensus on its suitability for HMI safety evaluation. Furthermore, the LCT has demonstrated sensitivity to distraction differences between in-vehicle systems across a range of studies (e.g., Burns et al., 2005; Harbluk et al., 2007; Maciej & Vollrath, 2009). Its traffic-free, controlled environment minimizes situational confounds and ensures that observed performance differences can be more accurately attributed to the secondary task rather than external factors. Initial validation of the LCT within the ADAM project further demonstrated that LCT performance correlates with lane position variability measured in real on-road driving ($r = .72$; Breuer et al., 2003), supporting the ecological validity of the measure. Finally, its low-cost, PC-based setup made it feasible to test the large number of HMI types and design aspects in this study (see section 6.2).

6.3.2 Generic Human Machine Interfaces

Three generic HMIs (physical controls, touch-based, speech-based) were created to simulate the operation of select vehicle functions. The foundation of each HMI is an app that was programmed in Apple Xcode (Apple Inc., 2024). The app ran on an iPad Air M2 with a screen size of 12.9 inches. The base version of each interface is the so-called reference version. Here, every interface design aspect is optimal, i.e., every checklist item can be rated with “0” (no deviation from the optimal implementation). The basic experimental setup was the same for every interface and is shown in Figure 6. The driving simulator was always equipped with buttons on the steering wheel and with the tablet placed on the table to the right of the driver. This was done to standardize the experimental setup across conditions.



Figure 6: Basic driving simulator setup used for all three generic HMIs (physical controls, touch-based, speech-based).

The physical controls and their visual interface are shown in Figure 7. To control the functions, the buttons on the steering wheel are used and visual and auditory feedback are provided by the iPad. The physical control elements are made up of an Xbox Adaptive Controller, to which buttons and triggers are connected. The simple function can be activated with a single button click while the more complex task requires a tracking operation that is controlled via the D-Pad and follow-up interactions with other buttons to fulfill the task.

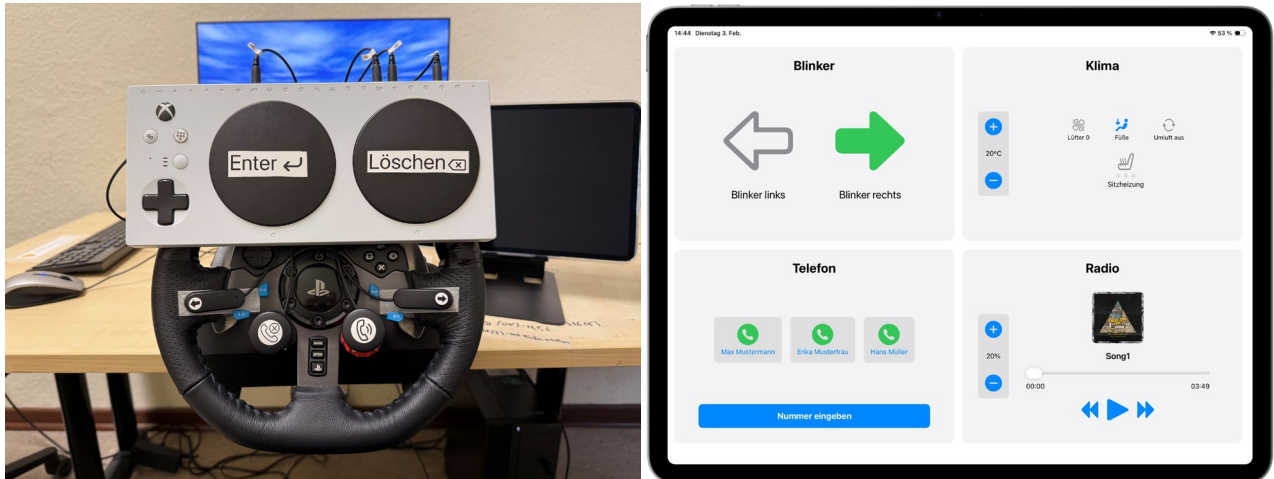


Figure 7: Physical controls: Physical steering wheel controls (left) and software interface (right).

The touch-based interface is shown in Figure 8. It uses a central home menu from which individual sub-menus can be accessed. Within each sub-menu, several functions of a function group can be controlled. For the simple task, a large virtual button must be pressed with a single tap. The more complex task requires repeated presses of small buttons in the correct order to achieve the task. Like with physical controls HMI, visual and auditory feedback are provided via the iPad.

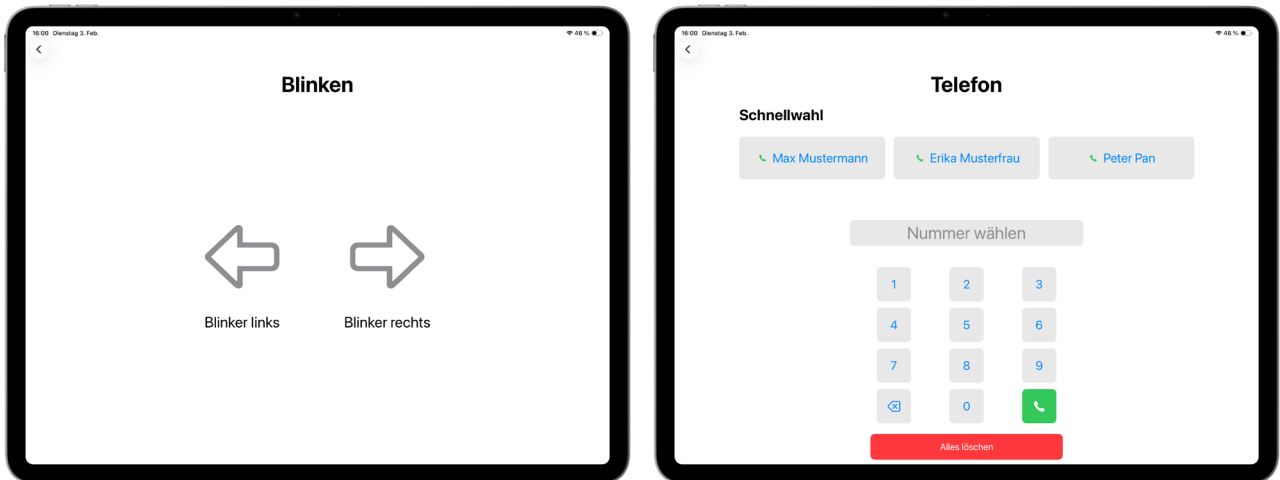


Figure 8: Touch-based interface: Touch interface used for the simple task (left) and complex task (right).

Finally, the speech-based interface is shown in Figure 9. The interface can be operated hands-free and is activated via a keyword. Once it recognizes the keyword, it indicates that it is ready to receive commands for function operation. Both the simple and complex functions are controlled by using a single command which then activates the target function. While the interface also provides visual and auditory feedback, the key feedback is spoken feedback indicating that the function has been successfully operated.

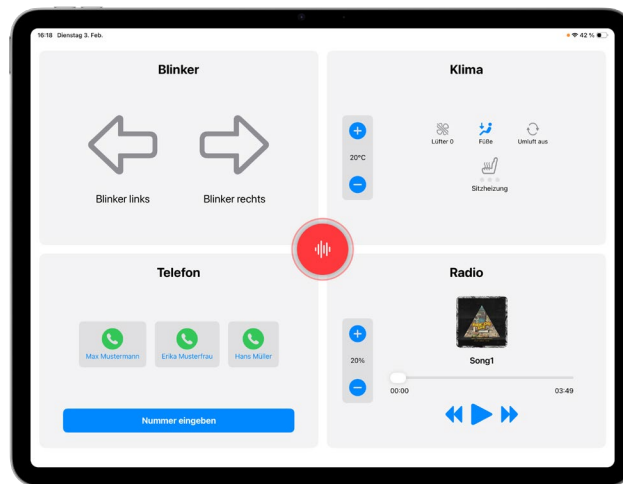


Figure 9: Speech-based interface: Software interface.

6.4 Secondary Tasks

As stated in the description of the experimental design (see section 6.2), two function operation tasks were employed. The first task was a simple task with an external use case that required immediate operation in a single step. The complex task had no external use case and required multiple operation steps. Both generic tasks had to be adapted so that they could be incorporated into all three generic interfaces (physical controls, touch-, and speech-based).

For the first task, it was important that the external cause was provided by the LCT driving simulation. As a result, the simple task was designed as an indicator task, which is fitting since the LCT centers around lane changes. Participants had to activate the indicator that corresponds to the lane change required by the LCT. While this is not representative of indicating in real life, it does represent the immediate single-step task nature well. For physical controls, buttons on the steering wheel needed to be pressed to activate the indicator. In the touch-based interface, large indicator buttons needed to be tapped. Finally, the speech-based interface used a direct voice command like “Indicate left!” to activate the indicators. All interfaces flashed the indicator three times. For feedback, the indicator arrow lit up green and made ticking sounds like in a real car. The implementation of the simple task in each interface type can be seen in more detail in the operationalization documentation (see Appendix B).

For the complex task, a phone number dialing and calling task was employed. Participants dialed in an 11-digit phone number, which was announced in four- and three-digit blocks by the experimenter. After dialing in the number, the call had to be started. When the connection was established, the call should be hung up. For physical controls, the dial pad was activated with a single button press. Then, the individual digits had to be selected with moving a selector across the dial pad with the D-pad. Once the correct digit had been reached, a button was pressed to input the number. When the number was fully typed in, the call was initiated with a separate button. Cancelling was also done with a separate button. In the touch-based interface, the dial pad was activated by pressing a touch button and the digits could then be individually tapped. Once the number was dialed in, the call was started with a calling-button. Finally, the speech-based interface required a single command where the participant stated they wanted to call the number which was announced by the experimenter. The implementation of the complex task in each interface type can be seen in more detail in the operationalization documentation (see Appendix B).

6.5 Procedure

After arriving in the lab, participants were informed about the experiment and were asked to fill out the informed consent form to take part in the study. Then, a demographic questionnaire was handed out containing questions regarding age, gender, socioeconomic background, and driving experience.

Next, participants were informed about the LCT and the function operation tasks. Also, the structure of the experiment was explained to them. In accordance to ISO 26022:2010 (International Organization for Standardization, 2010), before the actual measurements started, a practice block was held. Here, participants completed four training trials of the LCT without a secondary task. Then, only the secondary tasks were trained until they felt comfortable with function operation. For the generic interfaces, both the reference version as well as the interface design aspect test version were trained. Also, the operation of the radio was trained. Finally, three dual-task trials were completed where participants had to drive and operate the indicator, the phone (both in the reference version), and the radio.

Upon completion of the training, the experimental block and data collection began. For a first baseline measurement, participants completed one LCT trial without any secondary task. Next, the LCT driving performance was measured while concurrently operating functions. In a randomized order, participants operated the simple or complex task either via the reference version of the interface or via one of the four design aspect test versions. Also, one trial included the operation of the radio. After the block of function operation trials, a second baseline measurement of the LCT was carried out. Thus, participants had to complete a total of 11 measurement trials. In each trial, driving performance and function operation performance were measured.

After completing the measurements, participants filled out a post-questionnaire regarding their experience with in-vehicle HMIs and function operation. Participants were then thanked for their participation and compensated with 15 € or study credit. The duration of the experiment was around 90 minutes.

7 Results

As part of the empirical evaluation of the HMI checklist, generic interfaces were used to operate a function of low and high complexity. The effect on driver distraction was measured via the changes in driving performance and function operation performance. The goal was to investigate which HMI type is best suited to operate the low and high complexity function. Furthermore, the effects of the design aspects within each HMI modality on driver distraction and function operation performance were measured.

The results are split into two sections. The first section focuses on the effect of the HMI modality on driving performance with respect to task complexity (see section 7.2). For this general overview, the reference version of the interfaces is used. The second part analyzes the effect of the design aspects on driving and operation performance in comparison to no distraction (baseline), radio operation, and the reference version (see section 7.3). Each interface design aspect is then discussed and its effect on the simple and complex tasks are described. Both driving performance and function operation performance (objective and subjective) will be evaluated. For this detailed discussion per aspect, only the comparison to the reference version is used.

7.1 Driving Performance Measures and Data Preparation

Three measures of the LCT are used to describe the driving performance. The mean deviation (MDev) measures the driver's lateral deviation from the ideal trajectory over the complete LCT trial. Thus, it includes both sub tasks of lane keeping and lane changing. For lane keeping only, the standard deviation of lateral position (SDLP) describes the lateral deviation from the ideal trajectory. For lane changing only, the reaction time (RT) was computed between the moment the lane change sign became legible and the first steering input greater than five degrees. MDev and SDLP are used for measuring visual-manual distraction, while RT is used for measuring cognitive distraction.

The experimental design (see section 6.2 for the detailed description) is nested so that every participant only experienced four of the 51 interface design aspects. Four interface design aspects were measured with a sub sample of $n = 10$ people. As a result, 13 groups were needed to cover all 51 aspects. This proposes a challenge where each individual aspect has a sample size of $n = 10$ and where the four aspects of the same group are within-subjects and the remaining aspects are between-subjects. Additionally, the sample size for the reference version is greater than for the individual aspects as every group of the same modality experienced the reference version. Therefore, the design is nested and unbalanced, which requires an appropriate modelling approach. Also, there were differences in age ($p = .046$) and baseline driving performance ($p = .026$) between the participant groups.

For the data analysis, first, change scores were computed to address the significant differences in baseline driving performance. Here, for every driving performance measure (MDev, SDLP, RT), the baseline driving performance was subtracted from the driving performance of each trial where functions were operated. This was done to account for differences in individual baseline performance. Second, a mixed model approach was implemented with random intercepts at the individual and group level to address both the nested and unbalanced data structure as well as significant differences between participant groups. The random intercept on the individual level was applied to account for potential differences in the participant's ability to operate functions both on its own (without driving) and concurrently while driving. As the age differed significantly between groups, we assumed that younger participants who grew up with digital technologies might perform better at operating functions solely and when driving in parallel. Also, with higher age, motor abilities decrease which could have affected our results. Third, a random intercept on the group level was realized to account for potential covariance structures between the four interface design aspects that were measured within each group. The change score and mixed model approach will be used for most of the following analysis, except for the overview in section 7.2.1.

7.2 HMI Type and Task Complexity: Effects on Driving Performance

The first part of the validation of the HMI checklist focused on the effect of HMI type (physical controls, touch-based, and speech-based) and task complexity (simple, complex) on driving performance. More importantly, we wanted to investigate how both aspects interact with each other so that the most suitable interface for low and high complexity tasks can be derived. For this general overview, the reference version of the interfaces was used where all design aspects are implemented optimally.

7.2.1 General Overview: Simple and Complex Task Operation via Physical Controls, the Touch-Based and Speech-Based Interface

For a descriptive overview, Figure 10 depicts the driving performance measures referring to lateral deviation (MDev, SDLP). The dotted lines indicate baseline driving performance. Driving performance is shown for all combinations of HMI type (physical controls, touch-based, speech-based) and task complexity (simple, complex). In both measures, driving performance differed between all conditions: MDev: $F(7, 384.27) = 60.73$, $p < .001$, $\eta^2p = .53$, SDLP: $F(7, 309.58) = 25.69$, $p < .001$, $\eta^2p = .37$. For MDev, operating the simple and complex task via the speech-based interface had only a small effect on driving performance, essentially staying at the baseline level. Next, there is a plateau of radio operation and the operation of the simple task via physical controls and the touch-based interface. Then, driving performance gets worse with operating the more complex task (increased MDev). Here, physical controls led to a greater decline in driving performance compared to the touch-based interface. Physical controls have larger confidence intervals compared to the other interface types. For SDLP, the driving performance was on the baseline level for the simple task for all three HMI types. However, the complex task led to a decline in lane keeping performance. Here, the speech-based interface is near the baseline, however, physical controls and the touch-based interface performed on similarly worse level.

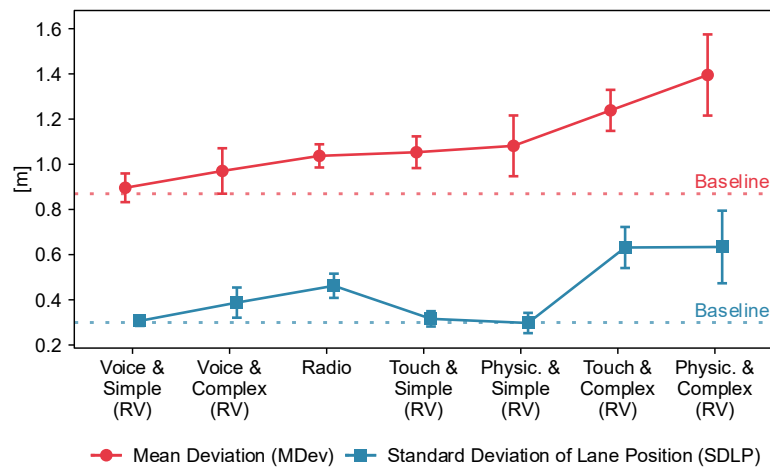


Figure 10: Lateral driving performance for every combination of task (simple, complex) and HMI type (physical controls, touch-based, speech-based). Baseline performance indicated by dotted line. Sorted in ascending order by MDev. RV = Reference version.

Two key factors explain the lateral deviations shown in Figure 10. First, for the tasks, not only did their complexity vary in the number of steps needed for completion, but also their temporal structure over the complete LCT trial. The simple task, which was designed as an indicating task, was done at every of the 18 lane changes in a trial. As result, per trial, participants only interacted with the function at very specific time points (at the beginning of the lane change). On the remainder of the trial, they could focus on keeping their lane. In contrast, radio operation and operating the complex task were done continuously both during the lane change and lane keeping phase. Therefore, the function operation frequency was much higher and both the lane keeping and lane changing phases were

affected. As a result, there is a large disparity between the tasks in SDLP, which only measures lateral deviation in the lane keeping phase. In the simple task condition, participants did not operate the function in this section of the trial. Hence, the deviation is on baseline level. This difference in task complexity and temporal structure can also be applied to MDev, which additionally covers lane changing. As the simple task required less steps and was done much less frequently, driving performance was better compared to the complex task. Second, the specific operationalization of the interfaces, especially of the physical controls, must be considered. As stated in the description of this type of interface, the location and control elements itself are diverse. As a result, the buttons and their placement on the steering wheel in our experiment are just one specific design of such an interface. Additionally, the LCT driving task is quite sensitive to steering inputs as they alter the deviation from the ideal trajectory. This is most likely why physical controls performed worse than the touch-based interface (complex task) or the radio and exhibit larger confidence intervals. Our experiment is concerned with concurrently performing a driving maneuver and operating functions via steering wheel controls. In real life, driving maneuvers are not constantly carried out, as is function operation via the steering wheel. Also, it is unlikely that driving and operating the function always happen concurrently. Instead, drivers are likely to do the tasks successively. Towards deviations from an ideal trajectory, in real life, these can be easily corrected and roads are usually wide enough to account for such errors. Another aspect to consider is the operation strategy of the participants as the confidence intervals for physical controls are broader than for touch- or speech-based interfaces. Participants differed in how they reached the element (leave hand on steering wheel an operating with thumbs vs. taking hand off steering wheel and operating with finger). As a result, our specific implementation of physical controls (steering wheel buttons, constantly operating performing driving maneuvers with no room for error) must be carefully transferred to other physical controls on, for example, the dashboard or center stack.

Figure 11 provides an overview of the reaction time (RT) for all combinations of interface type (physical controls, touch-, and speech-based) and task complexity (simple, complex). The dotted line indicates baseline reaction time. All conditions influenced RT, $F(7, 375.43) = 32.40, p < .001, \eta^2p = .38$. Speech-based operation is on baseline-level. This is followed by the radio. Touch-based operation of both tasks and the physically controlled simple task are on a plateau. Finally, the physical operation of the complex task led to the worst reaction times. As with the lateral deviation measures (MDev, SDLP), physical controls led to larger confidence intervals. RT was included to measure reaction time in the sense of cognitive processing time. However, as it was measured via the first steering input greater than five degrees after the lane change sign became legible, it can be affected by any factor that prolongs the steering input, only one of which cognitive distraction. Cognitive distraction would only be appropriate to consider for the speech-based interface. However, as RT is on baseline level, there appears to be no noticeable distraction. Physical controls, on the other hand, strongly interfered with steering input. This is presumably because steering and indicating (simple task) had to be done concurrently, as soon as the lane change sign became visible. Hence, the prolonged time to the first steering input is probably more of a manual interference than a cognitive distraction.

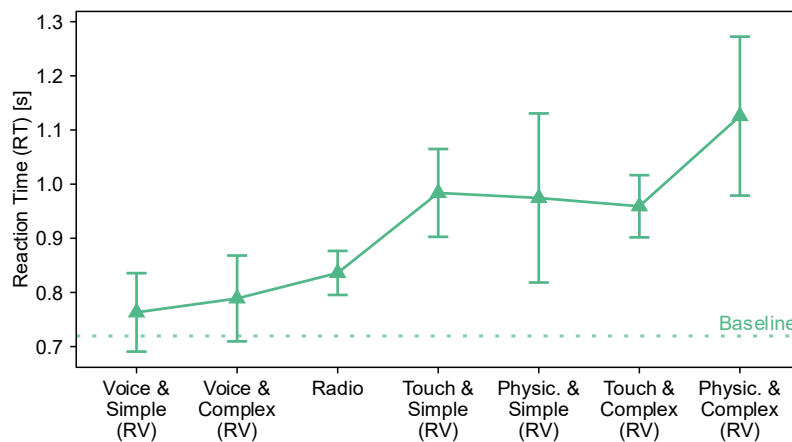


Figure 11: Reaction time for every combination of task (simple, complex) and HMI type (physical controls, touch-based, speech-based). Baseline performance indicated by dotted line. Sorted in ascending order by MDev. RV = Reference version.

7.2.2 Interaction Between HMI Type and Task Complexity

As seen in the overview, both the HMI type as well as task complexity affect driving performance and therefore driver distraction. Starting with the following analysis, the change score and mixed model approach will be employed (see section 7.1).

Figure 12 depicts the MDev interaction diagram for both tasks (simple, complex) and all three HMI types (physical controls, touch-based, speech-based). An ANOVA revealed significant main effects of modality, $F(2, 10.08) = 6.06$, $p = .019$, $\eta^2p = .55$, task complexity, $F(1, 127.58) = 61.45$, $p < .001$, $\eta^2p = .41$, and an interaction thereof, $F(2, 127.30) = 8.81$, $p > .001$, $\eta^2p = .15$, on the change score of MDev. Next, contrasts were calculated between all pairs of HMI modalities for each task. For the simple task, there were no differences between any pair of modalities (all $p > .05$). For the complex task, there were differences between the speech-based and the touch-based interface ($p = .006$) as well as to physical controls ($p < .001$). Physical controls and the touch-based interface did not differ from one another ($p = .122$). As stated previously, MDev represents the deviation from the ideal trajectory for both the lane change and lane keeping phases.

For the simple task, which required just one step and was performed only 18 times throughout a single trial, all HMI modalities resulted in the same level of distraction. Thus, participants were able to concentrate on driving for the majority of the trial and were rarely distracted by operating the simple function. It is therefore not surprising that there are no differences in MDev. As expected, the differences become more noticeable as the task becomes more complex (multiple operation steps) and has a higher operation frequency (continuous instead of 18 times per trial). For the complex task, the speech-based interface led to the best driving performance, followed by the touch-based interface and physical controls, which performed the same. This is somewhat surprising since we expected that physical controls would perform worse than the touch-based interface (see section 4.4). The results indicate that the attention allocation and manual input required appear to be similar between both interfaces. One could argue that physical controls and their visual interface require more visual tracking between three places (road, buttons, screen) whereas the touchscreen only has two visual areas of interest (road, screen). However, the results show that in turn the manual input was more demanding for the touchscreen (reaching for the screen and tapping precisely) compared to the buttons (both hands stay on the steering wheel, buttons themselves can be felt, press does not have to be precise).

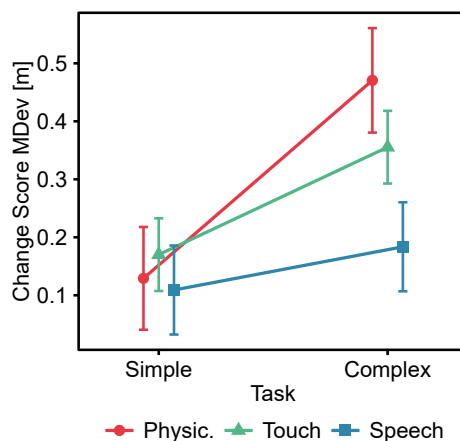


Figure 12: Interaction diagram for task (simple, complex) and HMI type (physical controls, touch-based, speech-based): Mean deviation (MDev) change score.

Figure 13 shows the SDLP interaction diagram for both tasks (simple, complex) and all three HMI modalities (physical controls, touch-based, speech-based). Just like for MDev, an ANOVA showed main effects for modality, $F(2, 10.37) = 7.30$, $p = .010$, $\eta^2p = .58$, task complexity, $F(1, 130.31) = 73.38$, $p < .001$, $\eta^2p = .36$, and an interaction of both factors, $F(2, 129.54) = 7.69$, $p < .001$, $\eta^2p = .11$, on the change score of SDLP. Contrasts revealed no differences between modalities for the simple task (all $p = 1.000$). As for the complex task, however, there were differences between the speech-based and touch-based interface ($p < .001$), and the speech-based interface and physical controls ($p < .001$). There was no difference between physical controls and the touch-based interface ($p = 1.000$).

As the SDLP only measures the deviation in lane keeping phases, it is fully within expectation that there are no differences between modalities for the simple task. During lane keeping, participants were not operating the simple task and could focus on following the ideal trajectory. Hence, driving performance is at baseline level. In contrast, operation of the complex task exhibits a clear difference between modalities. As with MDev, the speech-based interface resulted in the best driving performance, physical controls and the touch-based interface led to significantly more distraction. In both the MDev and SDLP interaction diagrams, the interaction of modality and task can be seen in the slopes of the graphs. For the speech-based interface, the slope from the simple to the complex task is rather shallow, indicating that speech-based function operation not really affected by the level of task complexity. On the other hand, the touch-based interface and physical controls are more sensitive to an increase in task complexity.

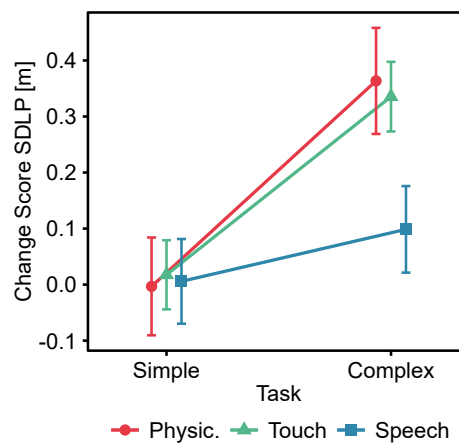


Figure 13: Interaction diagram for task (simple, complex) and HMI type (physical controls, touch-based, speech-based): Standard deviation of lane position (SDLP) change score.

Finally, Figure 14 depicts the RT interaction diagram for both tasks (simple, complex) and every HMI type (physical controls, touch-based, speech-based). An ANOVA revealed a main effect of modality, $F(2, 9.94) = 8.88$, $p = .006$, $\eta^2p = .64$, and task complexity, $F(1, 114.61) = 8.86$, $p = .004$, $\eta^2p = .07$, and an interaction of both factors, $F(2, 114.44) = 6.43$, $p = .002$, $\eta^2p = .10$, on the change score of RT. The interaction can be clearly seen in the diagram: When switching from the simple to the complex task, only physical controls exhibit prolonged RT. Contrasts only show a difference between the touch- and speech-based interface for the simple task ($p = .007$) and only between physical controls and the speech-based interface for the complex task ($p < .001$). As a result, there is no clear pattern between the interface types in RT when operating tasks of varying complexity. From a descriptive perspective, the speech-based interface performs the best and is unaffected by task complexity. Physical controls, on the other hand, lead to longer RT when task complexity increases. This is not surprising since the MDev and SDLP also exhibited the worst results (at least descriptively) for physical controls. Although RT is often used to capture cognitive distraction, in our study this can only be applied to the speech-based interface. For the visual-manual interfaces, the reaction time in the LCT is mainly affected by manual interference that delays the first steering input, thus not depicting cognitive distraction.

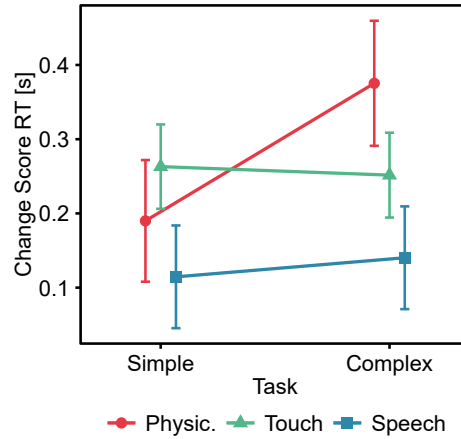


Figure 14: Interaction diagram for task (simple, complex) and HMI type (physical controls, touch-based, speech-based): Reaction time (RT) change score.

To sum up, the speech-based interface stands out positively in all three driving performance indicators (MDev, SDLP, RT). It consistently led to the least deviation from the ideal trajectory in both the lane change and lane keeping phases for the complex task. It also led to the quickest steering responses (RT) at the beginning of the lane change phase in both tasks. Physical controls and the touch-based interface led to worse RT for both tasks and worse lane deviation (MDev, SDLP) for the complex task. Although the results can be used to identify the best interface for the complex task, which is the speech-based one, it cannot differentiate between the modalities in the simple task. For one, this is due to the design only requiring the simple task to be performed at very specific time points during the whole trial. As a result, the MDev represents predominantly driving on its own and only somewhat driving and function operation in parallel. For another, the simple task competed with the initiation of the lane change. As result, the conclusion that for simple tasks all interfaces are equally suitable only applies for this very specific LCT driving context. However, this context does not resemble real life driving scenarios adequately. We strongly advise considering the results of other studies clearly indicating that physical controls or speech-based interfaces should be used for the operation of simple functions.

7.3 Interface Design Aspects (Implementation Quality): Effects on Driving Performance and Function Operation Quality

The interface design aspects refer to design principles that should be respected when implementing a function via the corresponding HMI type. Therefore, they are closely linked to the implementation quality of a given function. For the validation, the effect of a suboptimal implementation on driving performance and function operation quality was measured. Both driving performance and operation quality are made up of several dimensions (measures). To evaluate a given design aspect, all variables must be taken into account across both operation tasks (simple and complex). To this end, a data aggregation and analysis strategy was developed, which will be described in the next section. Thereafter, the individual design aspects per HMI modality are analyzed in comparison to three reference distraction levels: baseline (no distraction), radio (acceptable distraction), and reference version (reference level of distraction for the interface type). Finally, each design aspect and the effect of its suboptimal implementation are discussed. Driving performance and operation quality (subjective and objective) will be compared against the reference version of the interface, in which the design aspect was implemented optimally. In other words, the suboptimal implementation is directly compared against its optimal counterpart.

7.3.1 Measures, Data Analysis Strategy, Scoring System, and Aspect Sorting

The effect of the interface design aspects on driving performance and operation quality was evaluated. Driving performance was measured with the very same LCT dimensions used in the general overview (see section 7.2). Mean deviation (MDev) and standard deviation of lane position (SDLP) describe the deviation from an ideal trajectory, reaction time (RT) measures the reaction time between the lane change sign becoming legible and the driver's first steering input greater than five degrees.

Operation quality was measured both subjectively and objectively. For subjective operation quality, participants were asked how well they could operate the function via the interface (ease of use; EoU) and as how distracting they perceived the operation process (perceived distraction; Distr). Subjective operation quality was measured for both the simple and complex task. Objective operation quality was measured by the experimenter. For the simple task (indicating), the number of missed signals (Misses; Miss) and errors (Error) were counted. For the complex task (phone number dialing), the number of inserted numbers (N) and errors (Error) were counted.

To capture the effect of a design aspect's suboptimal implementation, contrasts were computed between the interface design aspect's test trial and three distraction reference levels: baseline (no distraction), radio (acceptable distraction), and reference version (reference distraction for the interface type). Contrasts were calculated for driving performance (MDev, SDLP, RT) and operation quality (simple task: EoU, Distr, Miss, Error; complex task: EoU, Distr, N, Error). To describe the effect more precisely, both the 80% and 95% confidence interval for the differences were computed. Three outcomes are possible, depending on the position of the difference's zero relative to the confidence intervals:

- No difference (Zero both in 80% and 95% confidence interval, $p > .20$)
- Possible difference (Zero not in 80% but in 95% confidence interval, $.05 < p < .20$)
- Significant difference (Zero neither in 80% nor 95% confidence interval, $p < .05$)

Next, a scoring system was applied that translates the dual-contrast outcomes into a contrast value. The contrast value also indicates the relative position of the interface design aspect's effect in relation to the distraction reference. Negative values indicate deterioration in driving performance and operation quality, positive values an improvement.

- No difference: ± 0 points
- Possible difference: ± 1 point
- Significant difference: ± 2 points

Finally, the contrast values for the driving performance and function operation measures were summed up. The resulting driving performance and function operation sums were then used to generate two scores, which we named penalty and net score. The *net score* (1) inverses both sums so that performance improvements result in negative values and detriments in positive values. Furthermore, driving performance is double-weighted compared to operation quality. As both tasks, simple and complex, have their own performance sums for driving and function operation, the score averages are used for the calculation. The net score ranges between -0.9 and +0.9 points.

$$Net = 0.10 \cdot (-Sum_{DP, Avg.}) + 0.05 \cdot (-Sum_{OP, Avg.}) \quad (1)$$

Sum_{DP Avg.} = Average of driving performance sums of both tasks

Sum_{OP Avg.} = Average of operation performance sums of both tasks

Building upon the net score, a *penalty score* (2) was derived. The penalty score is used as the weighting factor in the checklist. Instead of performance sum averages, the penalty score is based on the worst driving performance and operation performance sum of the two tasks. Also, it only takes negative values into account, i.e., only performance detriments are considered. Finally, the penalty score has a base value of 1 so that the multiplier never reduces the ticked value on the item's five-point rating scale. The penalty score ranges between 1 and 1.9 points.

$$Penalty = 1 + 0.10 \cdot \max(0, -Sum_{DP, WC}) + 0.05 \cdot \max(0, -Sum_{OP, WC}) \quad (2)$$

Sum_{DP WC} = Worst case of driving performance sums of both tasks

Sum_{OP WC} = Worst case of operation performance sums of both tasks

Finally, the driving performance sum average (Sum_{DP, Avg.}), part of the net score (1), was standardized to range between 1 (worse in all driving performance metrics) and -1 (improvement in all driving performance metrics). This standardized driving performance sum average (DP Avg.) will be used as the primary measure for the aspect overview discussion (e.g., section 7.3.2.1 for physical controls). Here, a comparison to all three distraction reference levels (baseline, radio, reference version) is made based on DP Avg. Also, a detailed comparison to the reference version containing the net and penalty score as well as the affected tasks is provided. This more detailed comparison is not made to the baseline and radio operation. This is due to the operation quality not being included in the baseline distraction level (no function operation) and the radio distraction level (only subjective but no objective operation quality). Both would be needed for calculating the penalty and net score.

For easier orientation, an overview of the scores is provided in Table 9. For each score, it is described what it represents, i.e., if it is concerned with driving/operation performance of an aspect's suboptimal implementation. The table then also sums up how the score is calculated and what it is ultimately used for. Regarding the checklist, the penalty score is the most important since it is used for deriving the weighting factor for each item.

Table 9: Score Overview

Score	Represents	Calculation	Used for
DP Avg. (Standardized Driving Performance Sum Average)	Relative effect of the suboptimal implementation of a design aspect compared to <i>all three reference levels</i> (baseline, radio, reference version): Driving performance	Standardized <i>averages</i> of driving performance contrast sums across both tasks (simple & complex)	Primary measure in the aspect overviews for each HMI modality (sections 7.3.2.1, 7.3.3.1, 7.3.4.1)
Net Score	Relative effect of the suboptimal implementation of a design aspect compared to <i>the reference version</i> : Driving performance & operation quality	Double-weighted driving performance + operation quality; uses <i>averages</i> across both tasks; detriments = positive values	Aspect overviews and detailed aspect discussions (sections 7.3.2 – 7.3.4)
Penalty Score	Worst-case impact of the suboptimal implementation of a design aspect compared to <i>the reference version</i> : Driving performance & operation quality	Based on net score logic, but uses <i>worst-case task</i> instead of averages; only detriments count; base value of 1 ensures the multiplier never reduces a checklist rating	Aspect overviews and detailed aspect discussions (sections 7.3.2 – 7.3.4) Weighting factor in checklist

Note. Net and penalty scores were only computed relative to the reference version, as operation quality is not available for baseline and radio reference levels.

7.3.2 Interface Design Aspects: Physical Controls

The analysis of the physical controls' interface design aspects is split into two parts: The first section provides an overview of the design aspects tested and compares their relative effect on driving performance. The second part analyzes every design aspect in detail. Here, effects on driving performance and operation quality (objective and subjective) are discussed as well as implications for function implementation. The design aspects are described in more detail in the checklist (see Appendix A). The specific operationalization of each design aspect in the experiment is described in Appendix B.

7.3.2.1 Overview and Comparison to All Three Distraction Reference Levels

The complete list of tested design aspects for physical controls is summarized in Table 10. The design aspects are arranged in two ways in relation to the reference version of the interface. First, they are sorted by their effect on driving performance via the standardized driving performance sum average (DP Avg. – Baseline, see section 7.3.1). Aspects with a more negative effect on driving performance are listed first. Second, if two aspects have the same effect on driving performance, they are additionally sorted by net score with the more negatively affecting aspect listed first. Based on their relative effect on driving performance, the aspects can be divided into two groups: Those which have a negative effect on driving performance and those with no effect compared to the reference version. For easier orientation, the driving performance comparison is colored: worse driving performance is marked orange and no difference grey. The aspects with the most negative impact on driving performance are listed at the top of the table.

Every design aspect was tested via suboptimally implementing it in the reference version of the interface. The global analysis (see section 7.2.1) has already revealed that for both tasks physical controls lead to a worse driving performance compared to the baseline (no distraction). For the complex task, the driving performance is worse than radio operation (acceptable distraction). Therefore, it is not surprising that the operation of both tasks via the function aspects' test interfaces introduced even more distraction when compared to the baseline and radio operation (see DP average). For all aspects, the driving performance differences to the baseline are greater than to radio operation, which is expected. When comparing both groups of aspects (negative effect on driving performance vs. no negative effect), viewing (and operating) distances to the right (M1-PC-1) had a more negative impact than increased distances downwards (M1-PC-5). This indicates that visual and manual effort not only depend on the distance itself but also on directionality. Further, removing auditory feedback (M2-PC-5) or changing the labeling of the control elements (M1/2-PC-6) had no real effect, which is likely due to the participants having learned the tasks well. For real-life function operation, which not happens in a constantly repeating manner like in our experiment, we would expect these principles to have an effect.

In comparison with the reference interface, there are eight design aspects that resulted in worse driving performance. In most cases, only one of the tasks (simple or complex) was affected regarding driving performance. For operation performance, usually both tasks were affected. For the simple task, there are MDev and RT differences while there are no effects on SDLP. For the complex task, there are no differences for MDev but for SDLP and RT. This is somewhat expected since the simple task was only executed during the lane change and not the lane keeping phase, the latter of which is depicted by SDLP. However, it is surprising that the complex task only exhibits differences for SDLP and therefore lane keeping but not MDev, which additionally covers lane changes. The results may indicate a compensation strategy, that when the complex task's operation got harder, participants stopped operation during critical lane change parts of the trial. This is reflected by the decrease in the number of inserted phone numbers without elevated error rates. Four design aspects had no effect on driving performance when compared to the reference interface. Still, two of the aspects had negative effects on operation quality, whereas the remaining two did not. Overall, the majority of design aspects affected driving performance when they were implemented in a suboptimal manner.

The following section will highlight the effects of the individual design aspects in more detail. This is especially important since the global results revealed that complex function operation is quite challenging with physical controls compared to operating a simple function. Therefore, specific design aspects may be more important regarding the simple or complex task. Another aspect to consider beyond task complexity is the specific implementation of the physical controls in our validation experiment. Our results are mainly valid for steering wheel controls. We did not test physical controls on the steering column or dashboard. Likewise, we did not test function operation for a task with a medium complexity. As a result, the findings need to be transferred cautiously to other physical controls and the operation of other kinds of tasks.

Table 10: Validation Results for the Design Aspects of Physical Controls

Checklist				DP Avg. [-1 (worse),+1 (better)]			Comparison with reference version				
Item	Description	Optimal	Suboptimal	Base.	Radio	Ref.	Net DP	Net OQ	Net Score	Affected Task DP	Affected Task OQ
Negative effect on driving performance relative to reference version											
M2-PC-1	Susceptibility to Operation Errors	Resistance and position optimal	Resistance unfavorable and hard to reach	-0.83	-0.75	-0.33	0.2	0.15	0.35	Both	Both
M1-PC-1	Viewing Distance to the Right	0° – 15°	45° – 60°	-0.83	-0.67	-0.25	0.15	0.17	0.32	Complex	Both
M1-PC-2	Blind Operation	Haptics when reaching, during operation, setting can be felt	Haptics only when reaching	-0.83	-0.75	-0.25	0.15	0.06	0.21	Both	Simple
	Note: Only the simple task's control elements were altered.										
M1-PC-3	Glare	No glare, matte surface	Reflective surface	-0.75	-0.33	-0.17	0.1	0.04	0.14	Complex	Complex
M2-PC-2	Expectation Conformity	Fully in line with expectations	Other position	-0.83	-0.58	-0.08	0.05	0.24	0.29	Simple	Both
M1-PC-4	Reachability	Both hands on the steering wheel	One hand on the steering wheel, arm widely extended	-0.83	-0.58	-0.08	0.05	0.19	0.24	Simple	Both
M2-PC-3	Correction of Operation Errors	One action, immediately	Multiple actions, slowly	-0.83	-0.42	-0.08	0.05	0.08	0.12	Complex	Both
M2-PC-4	Risk of Operation Errors	No safety risk	Strong distraction	-0.83	-0.42	-0.08	0.05	0	0.05	Complex	–
No effect on driving performance relative to reference version											
M1-PC-5	Viewing Distance Downwards	0° – 15°	30° – 45°	-0.83	-0.42	0	0	0.09	0.09	–	Both

Checklist				DP Avg. [-1 (worse),+1 (better)]			Comparison with reference version				
Item	Description	Optimal	Suboptimal	Base.	Radio	Ref.	Net DP	Net OQ	Net Score	Affected Task DP	Affected Task OQ
M2-PC-5	Additional Feedback	Haptic, acoustic, visual	Haptic and only acoustic or visual	-0.83	-0.33	0	0	0.04	0.04	–	Complex
M2-PC-6	Comprehensibility	Symbols and text distinct	Text and/or symbols unclear	-0.75	-0.25	0	0	0	0	–	–
M1-PC-6	Legibility	High-contrast, distinct, appropriate size	Low-contrast, ambiguous, small	-0.83	-0.33	0	0	0	0	–	–

Note. Sorted in descending order by negative relative effect on driving performance compared to the reference version (DP Avg. – Ref.). DP = Driving Performance, OQ = Operation Quality. DP Avg. = Standardized driving performance sum average derived from net score.

7.3.2.2 Design Aspects in Detail

The effect of each design aspect's suboptimal implementation will be discussed in detail in the following section. The operationalization during the validation will be highlighted as well as the specific driving performance and operation quality metrics affected. Also, the differences between the simple and complex task will be explored. The aspects will be discussed in order of effect on driving performance as shown in Table 10, from top to bottom. For illustration see Appendix B.

The aspect with the most negative impact on driving performance was M2-PC-1 *Susceptibility to Operation Errors*, which is concerned with the likelihood of operation errors occurring (see Table 11). In our validation, the control elements were moved away from the steering wheel and laid loosely on the table to the right of the participants. As a result, the elements had to be precisely grabbed and the interaction required increased motor effort. We are aware that this aspect also concerns reachability and one-handed operability when compared to the optimal implementation on the steering wheel. For the simple task, MDev and RT were negatively affected as was subjective operation quality (EoUS, Distr). In comparison, the complex task only exhibited a slight degradation in SDLP and both subjective (EoU) and qualitative (N) operation quality measures.

As the simple task was slightly more affected regarding lane keeping performance, especially time-critical function operation should rely on control elements that are close to the driver and easy to grab. The complex and continuous task seems to be affected less, however, the decrease in the number of numbers inserted could indicate a compensation strategy. This is likely since driving performance was not really degraded and the global analysis already revealed the high effort for physically operating a complex task.

Table 11: Physical Controls - Susceptibility to Operation Errors

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-PC-1		Susceptibility to Operation Errors		Resistance and position optimal		Resistance unfavorable and hard to reach		Controls on the steering wheel vs. controller/button loosely on table	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-2	EoU	-2	MDev	-1	EoU	-2		
SDLP	0	Distr	-2	SDLP	0	Distr	0		
RT	-1	Miss	0	RT	0	N	-2		
		Error	0			Error	0		
Sum	-3	Sum	-4	Sum	-1	Sum	-4	1.45	0.35
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The aspect with the second-most negative impact on driving performance was M1-PC-1 *Viewing Distance to the Right* (see Table 12). It was tested via placing the control elements on a stable board to the right of the participants. In contrast to M2-PC-1 *Susceptibility to Operation Errors*, the control elements had a fixed position on the board and were faced towards the driver. Interestingly, there were no effects on driving performance for the simple task. However, driving performance was affected for the complex task (SDLP, RT). For the simple task, only subjective operation quality was affected (EoU, Distr). For the complex task, all subjective (EoU, Distr) and objective (N, Error) operation quality measures were degraded. As M1-PC-1 is an easier variant of M2-PC-1 with a lower level of increased motor effort, the results align with those of M2-PC-1: With a somewhat easier operation, the simple task performed the same as on the baseline level. The complex task now exhibits effects since participants likely started function operation again, which was too hard with M2-PC-1. Basically the “non-operation threshold” was not exceeded so that operation took place.

The results can be applied to the existing operation via physical controls in the center console, indicating that simple task execution is just as demanding as with steering wheel controls (ideal implementation in our validation). However, for complex tasks, the operation with center console elements appears to be rather difficult. Here, visual attention must be shared between the control elements, the screen, and the road. Therefore, the results indicate that from a driving performance perspective, simple task execution via the dashboard is acceptable when compared to steering wheel controls. Still, participants reported increased distraction and more difficult operability (EoU). Therefore, we recommend the use of steering wheel controls if possible.

Table 12: Physical Controls - Viewing Distance to the Right

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-PC-1		Viewing Distance to the Right		0° – 15°		45° – 60°		Controls on steering wheel vs. right on controls holder	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	-2	MDev	0	EoU	-1		
SDLP	0	Distr	-2	SDLP	-2	Distr	-1		
RT	0	Miss	0	RT	-1	N	-2		
		Error	0			Error	-1		
Sum	0	Sum	-4	Sum	-3	Sum	-5	1.49	0.32
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The aspect M1-PC-2 *Blind Operation* is concerned with the tactile-haptic nature of the control element (see Table 13). Ideally, the actuation of the function can be felt. For the validation, control elements with increased motor effort and no tactility were used on the steering wheel. Although these elements were only used for the operation of the simple task, we also measured the operation of the complex task, whose control elements were not altered. This was done since the new control elements were bigger and we suspected they might interfere with the complex task's operation. For the simple task, there was a deterioration of driving performance (MDev) and subjective operation quality (EoU, Distr). For the complex task, only SDLP was slightly affected while there was no effect on operation quality. The effect on driving performance is likely due to the increased motor effort and not so much due to the lack of tactility. Because the operation element required longer pressing distances and the stretching of thumbs, this likely led to the degraded MDev. Although the complex task was not relying on the altered control element itself, it might have interfered with lane keeping performance. Still, as only SDLP is slightly affected, driving performance essentially remained unaffected.

The results point toward control elements requiring low motor effort with shallow travel distances for low-distraction function operation. Also, they should be designed in a compact manner. Finally, the actuation should be felt via tactile feedback to communicate the successful input.

Table 13: Physical Controls - Blind Operation

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-PC-2		Blind Operation		Haptics when reaching, during operation, setting can be felt		Haptics only when reaching		Buttons with click and resistance vs. trigger without haptic feedback	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-2	EoU	-2	MDev	0	EoU	0		
SDLP	0	Distr	-1	SDLP	-1	Distr	0		
RT	0	Miss	0	RT	0	N	0		
		Error	0			Error	0		
Sum	-2	Sum	-3	Sum	-1	Sum	0	1.31	0.21
Negative effect on driving performance									
Note: Only the simple task's control elements were altered.									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The reflectivity of the physical control element is covered with the design aspect M1-PC-3 *Glare* (see Table 14). For the validation, glare was simulated by a torchlight, of which the edge of the light beam hit the participant's eyes. Only the complex task was affected regarding driving performance (RT) and objective operation quality (N, Error). There were no effects on driving performance nor on operation quality for the simple task. Given the nature of the complex task that requires constant operation and, more importantly, allocating the visual attention between the control element, screen, and driving environment, the glare seems to have somewhat interfered with the attention allocation process. While lane keeping (MDev, SDLP) was not affected, RT was negatively impacted as perhaps lane changes were missed. But this would have also led to a worse lane keeping performance. It is also likely that participants applied a different steering strategy when blinded in their periphery, i.e., more cautious steering. The light source also decreased objective operation quality to some extent. However, this could be a compensation effect as lane keeping was not affected.

The results can be interpreted twofold: First, when drivers do not have to look at the control element, function operation is not affected by the control element's glare. Therefore, control elements do not have to be necessarily matte. However, in cases where visual attention is shared, the glare seems to interfere with the attention allocation process, thereby affecting function operation and driving. This could be relevant when a control element unrelated to the operation of a specific function produces glare and affects function operation. This would imply that all control elements should incorporate matte surfaces.

Table 14: Physical Controls - Glare

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-PC-3		Glare		No glare, matte surface		Reflective surface		No illumination vs. flashlight directed at driver	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	0	Miss	0	RT	-2	N	-1		
		Error	0			Error	-1		
Sum	0	Sum	0	Sum	-2	Sum	-2	1.27	0.14
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The aspect M2-PC-2 *Expectation Conformity* refers to the location of the control elements and the operation principle following established standards (see Table 15). In the validation, the location of the indicator operation buttons (simple task) was swapped on the steering wheel. For the phone task (complex task), the controller was placed on the table to the left of the steering wheel. Technically, the complex task hereby also increased the operation distance and motor effort. Driving performance was affected to some degree for the simple task (MDev) but not for the complex task. The effects on operation quality were more severe: For both tasks, the subjective (EoU, Distr) and objective measures (Error for both tasks, N for complex) were negatively impacted. Although the manual operation effort increased more so for the complex task, it is curious that this had no effect on driving performance. In this context, it may also appear odd that the effect on operation quality was the same for the simple and complex task. However, while the complex task has had an increase in motor effort, the simple task switched to an unusual operation logic (swapping of buttons). This is likely the reason operation quality was affected similarly.

The results demonstrate that a change in operation logic and an increase in motor effort clearly affect operation performance. Therefore, for easy operation, established function logic must be adhered to. Additionally, manual effort must be kept low by placing buttons close to the driver as discussed with aspect M1-PC-1 *Viewing Distance to the Right*.

Table 15: Physical Controls - Expectation Conformity

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-PC-2		Expectation Conformity		Fully in line with expectations		Other position		All controls on the steering wheel vs. indicator buttons swapped and controller on table	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-1	EoU	-2	MDev	0	EoU	-2		
SDLP	0	Distr	-2	SDLP	0	Distr	-2		
RT	0	Miss	0	RT	0	N	-2		
		Error	-2			Error	-1		
Sum	-1	Sum	-6	Sum	0	Sum	-7	1.36	0.29
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The aspect M1-PC-4 *Reachability* refers to the proximity of the control elements to the driver's hands (see Table 16). It is also included to some extent in the aspects M2-PC-1 *Susceptibility to Operation Errors*, M1-PC-1 *Viewing Distance to the Right*, M2-PC-2 *Expectation Conformity*, and M1-PC-5 *Viewing Distance Downwards*. During the validation, the control elements were placed on a stable platform and farther away to the front of the driver. They were placed to the right of the monitor of the LCT driving simulation to keep the viewing distances between driving environment, control elements, and screen low. This had nearly no effect on driving performance (only SDLP for the simple task). Subjective operation quality was negatively affected for both tasks, while for the complex task, objective operation quality (N) was additionally degraded.

The results show that reachability towards the front of the driver does not have such an adverse effect as increased operation and viewing distances to the right. As a result, operation elements should be placed close to the driver and their line of sight.

Table 16: Physical Controls - Reachability

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-PC-4		Reachability		Both hands on the steering wheel		One hand on the steering wheel, arm widely extended		Both hands on the steering wheel vs. arm widely extended (60 cm)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	-2	MDev	0	EoU	-2		
SDLP	0	Distr	-2	SDLP	0	Distr	-2		
RT	-1	Miss	0	RT	0	N	-2		
		Error	0			Error	0		
Sum	-1	Sum	-4	Sum	0	Sum	-6	1.33	0.24
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The aspect M2-PC-3 *Correction of Operation Errors* outlines how operation errors can be corrected (see Table 17). In the validation, participants had to cancel a simulated malfunction via pressing a cancellation button three times. The malfunction was simulated by displaying a flashing screen, alternating in green and red color, and playing error sounds. There was essentially no effect on driving performance (only SDLP for the complex task). Operation quality was somewhat affected: Distr for both tasks, N only for the complex task. Although the cancellation made the interaction more complex with three button presses on the steering wheel, this quick multiple-step interaction does not seem to affect driver distraction.

The results therefore indicate that quick, repeated inputs, such as scrolling through a list, are not decisive for driver distraction. Still, operation steps should be kept to a minimum. This is supported by the next aspect, M2-PC-4 *Risk of Operation Errors*, which resulted in a little less driver distraction by requiring only one button press for cancellation instead of three presses in quick succession.

Table 17: Physical Controls - Correction of Operation Errors

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-PC-3		Correction of Operation Errors		One action, immediately		Multiple actions, slowly		Direct cancellation possible vs. pressing cancel button 3 times	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	-2	SDLP	-1	Distr	-1		
RT	0	Miss	0	RT	0	N	-1		
		Error	0			Error	0		
Sum	0	Sum	-2	Sum	-1	Sum	-2	1.18	0.12
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

With the aspect M2-PC-4 *Risk of Operation Errors*, the effect of an erroneous function operation was tested (see Table 18). The function operation triggered a salient flash screen (red and green colors) and error sounds that had to be corrected with a single button press. As with M2-PC-3 *Correction of Operation Errors*, the button for cancellation was placed on the steering wheel. Compared to M2-PC-3, there was the same minor effect on driving performance for the complex task (SDLP), however, operation quality remained completely unaffected. In comparison to M2-PC-3, the results indicate that fewer operation steps are better. Nonetheless, a single added operation step only has a marginal effect on driver distraction for the complex task. This is somewhat curious as one would expect the added step to have an effect on the simple single-step function instead of the complex one that already has multiple steps.

Although the results may implicate no effect of a single added operation step and an operation error, we advise requiring function operation to have as little steps as possible (no unnecessary menu layers) and to reduce the likelihood of operation errors.

Table 18: Physical Controls - Risk of Operation Errors

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-PC-4		Risk of Operation Errors		No safety risk		Strong distraction		Normal vs. flashscreen + error sound (salient)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	-1	Distr	0		
RT	0	Miss	0	RT		N	0		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	-1	Sum	0	1.1	0.05
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

With the design aspect M1-PC-5 *Viewing Distance Downwards*, the downward positioning of the control element is described (see Table 19). For the validation, the control elements were placed on a stable platform and put on a chair to the right of the participant. This was done to mimic button placement on the center console. There were no effects on driving performance. For the simple task, only one of the objective (Miss) and one of the subjective operation (Distr) quality measures were affected. The complex task was less affected with only a subtle difference in the number of inserted numbers (N). The results are interesting, especially when M1-PC-1 *Viewing Distance to the Right* was the second-most distracting design aspect. Here, drivers had to reach for the control element and look to the right. With M1-PC-5, however, reaching the control element was easier and the arm could also be rested on the leg. Therefore, the improved results could stem from the reduced motor effort. Also, perhaps splitting the visual attention was easier as the distance between the control elements, the interface's screen, and driving simulation was lower for M1-PC-5 compared to M1-PC-1.

The results indicate that placing buttons on the center stack is not very problematic. However, operation distance was low in our validation. With higher distances to the right (M1-PC-1) or forward (M1-PC-4), it can be expected that driver distraction will increase. So, center stack positioning of physical controls must be done carefully.

Table 19: Physical Controls - Viewing Distance Downwards

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-PC-5		Viewing Distance Downwards		0° – 15°		30° – 45°		Controls on steering wheel vs. on control holder placed on chair	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	-2	SDLP	0	Distr	0		
RT	0	Miss	-2	RT	0	N	-1		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	0	Sum	-1	1.15	0.09
No effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The aspect M2-PC-5 *Additional Feedback* refers to the feedback dimensions used during function operation (see Table 20). Feedback can be provided via haptic, visual, and acoustic channels. For the validation, the acoustic feedback was omitted. This had no effects on driving performance for both tasks. Objective operation quality was only affected for the complex task (N). Although the interface omitted additional acoustic feedback, there was still acoustic feedback provided via the clicking of the physical control elements. For simple tasks, the results indicate that this acoustic feedback is either sufficient or not needed at all. For the complex task, the omission of acoustic feedback resulted in a lower number of phone numbers inserted. This is probably due to the phone call nature of the task: once a number was fully put in, participants could listen to how the call was established and when they could cancel it. Also, they could clearly hear when they successfully moved the cursor on the dial pad and put in a digit. Without the acoustic feedback, the continuous operation was interrupted. Still, function operation overall does not appear to have gotten harder. This is reflected by the lack of an effect on driving performance and the other function operation measures.

Overall, the results indicate that additional acoustic feedback is not really important for physically controlled functions in regard to driver distraction.

Table 20: Physical Controls - Additional Feedback

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-PC-5		Additional Feedback		Haptic, acoustic, visual		Haptic and only acoustic or visual		App with audio feedback vs. app muted	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	-2		
		Error	0			Error			
Sum	0	Sum	0	Sum	0	Sum	-2	1.07	0.04
No effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The aspect M2-PC-6 *Comprehensibility* is concerned with the meaning of the icons and text used on the physical control elements (see Table 21). For the validation, uncommon symbols replaced the symbols of the reference interface. For neither task were there any effects on driving performance and operation quality. This was somewhat expected since participants trained function operation with the interface beforehand and knew the button assignment. Therefore, when switching the symbols for a single trial, they already knew the button assignment. However, the symbols and icons might be relevant in situations where a driver is new to a vehicle, or a rarely used function needs to be operated.

Although our validation demonstrates that a change of symbols (for a new vehicle revision, for example) does not really have an effect as long as the function assignment is well learned, the above-mentioned edge case of unfamiliar function operation would likely benefit from easy to understand button labeling.

Table 21: Physical Controls - Comprehensibility

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-PC-6		Comprehensibility		Symbols and text distinct		Text and/or symbols unclear		Symbols from instruction vs. ambiguous text and geometric shapes	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	0		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	0	Sum	0	1	0
No effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The aspect M1-PC-6 *Legibility* refers to how well the button labeling can be perceived, visually (see Table 22). M2-PC-6 *Comprehensibility*, in contrast, is more concerned with the meaning of the labeling. For the validation, the button labels of the reference version were exchanged for labels that were harder to read, i.e., low-contrast, small fonts and symbols. This had no effect on driving performance or on operation quality for both tasks. As with M2-PC-6, this is likely due to the participants being familiar with the button assignments so that they did not need to look at the buttons. Therefore, the exchanged button labeling had no effect.

As mentioned previously, button labelling is presumably more relevant for non-regularly used functions. As a result, easy to read labeling should be employed so that in these cases, low-distraction function operation can be enabled. With clear labeling that is easy to understand, the visual search for a button in the vehicle cockpit can be achieved quicker.

Table 22: Physical Controls - Legibility

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-PC-6		Legibility		High-contrast, distinct, appropriate size		Low-contrast, ambiguous, small		Large, high-contrast, black-and-white vs. light grey, too small	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	0		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	0	Sum	0	1	0
No effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

7.3.3 Interface Design Aspects: Touch-Based Interface

The analysis of the touch-based interface design aspects is split into two parts: The first section provides an overview of the design aspects tested and compares their relative effect on driving performance. The second part analyzes every design aspect in detail. Here, effects on driving performance and operation quality (objective and subjective) are discussed as well as implications for function implementation. The design aspects are described in more detail in the checklist (see Appendix A). The specific operationalization of each design aspect in the experiment is described in Appendix B.

7.3.3.1 Overview and Comparison to All Three Distraction Reference Levels

Table 23 contains the list of all aspects that were tested for the touch-based interface. The design aspects are arranged in two ways in relation to the reference interface. First, they are sorted by their relative effect on driving performance (standardized driving performance sum average: DP Avg. – Baseline, see section 7.3.1). Second, if two or more aspects have the same effect on driving performance, they are sorted by net score. For both sorting principles, the aspect with the more negative effect on driving performance and operation quality is listed first. Four groups of aspects can be derived: Those with a negative relative effect on driving performance, those with no negative effect, those with a possible positive effect (but likely not effect), and finally, those who improve driving performance relative to the reference interface. For easy orientation, the driving performance comparisons are colored: negative relative effects on driving performance are marked orange, no effects are marked grey, and improvements are marked green. The most distracting interface design aspects are listed at the top of the table.

Every design aspect was tested via suboptimally implementing it in the reference version of the interface. The global analysis (see section 7.2.1) has already shown that touch-based function operation results in driving performance worse than the baseline (no distraction; both tasks) and radio operation (acceptable distraction; only complex task). Therefore, it is within expectation that function operation via the design aspect's test interface led to greater distraction than both distraction reference levels. An exception to this is the bonus aspects listed at the bottom of the table. The shortened function operation (M2-TB-9 *Shortcut*) led to driving performance that is better than the reference version but worse than the radio operation. The last two aspects (M1-TB-12 *Deactivation*, M1-TB-13 *Blocking of Distracting Functions*) required no function operation and tested the effect of simply displaying a static or dynamic interface. As a result, it was expected that driving performance would stay near or at the baseline level. When comparing the three groups of design aspects (negative effect, no effect, positive effect), there are discrepancies. Like for physical controls, the effect of viewing (and operating) distances depends on the directionality: only when increasing them downwards (M1-TB-5) do they have a negative effect but not when increasing them rightwards (M1-TB-10). Additionally, blocking smaller parts of the driving simulation (M1-TB-3) had a more negative impact than blocking greater parts (M1-TB-11), which is somewhat surprising. But with blocking greater parts of the simulation, the touch-based interface perhaps was easier to see and operate for the drivers. Finally, changes in the visual structure of the interface (M1-TB-7,8,9) had no effect. As with the physical controls, this is likely due to participants being very familiar with the interface.

Of the remaining 20 design aspects, 12 led to a worse driving performance compared to the reference version. Operation performance was usually affected for both tasks (simple and complex) or for the complex task only. Regarding driving performance, the simple task alone seems to have been slightly more affected (five aspects) than the complex task alone (three aspects) or both tasks together (four aspects). For the simple task, MDev and RT were clearly affected whereas the effects are more subtle for the complex task. Here, SDLP was also affected although the differences compared to the reference version are not as distinct. The deviation in SDLP differences between both tasks make sense as the simple task was not executed during the lane keeping phases that SDLP depicts. Regarding operation quality, mainly subjective measures exhibit differences to the reference version for the simple task. The complex task, on the other hand, exhibits differences mainly in objective measures of operation quality. Next, there are six aspects that did not differ from baseline driving performance. There are some differences in subjective and objective operation quality measures. Finally, there are five aspects that led to improved driving performance compared to the reference interface. For the first two items (M1-TB-10 *Viewing Distance to the Right*,

M1-TB-11 *View on the Road Not Blocked*) an improvement was not expected. However, the remaining three aspects (M2-TB-9 *Shortcut*, M1-TB-12 *Deactivation*, M1-TB-13 *Blocking of Distracting Functions*) were specifically expected to improve driving performance as discussed in the previous paragraph.

In the following section, each interface aspect will be discussed in detail. The global results revealed that touch-based function operation introduces high levels of visual-manual distraction, especially at high task complexity. Furthermore, they are the predominant interface used in modern vehicles. Therefore, touch-based interfaces must be designed in a way that minimizes driver distraction. As with physical controls, our validation study only tested very specific implementations and violations of the design principles. Also, only a simple and complex but no medium complexity task were tested. The results must therefore be transferred cautiously to other touch-based interfaces and functions.

Table 23: Validation Results for the Design Aspects of the Touch-Based Interface

Checklist				DP Avg. [-1 (worse),+1 (better)]			Comparison with reference version				
Item	Description	Optimal	Suboptimal	Base.	Radio	Ref.	Net DP	Net OQ	Net Score	Affected Task DP	Affected Task OQ
Negative effect on driving performance relative to reference version											
M1-TB-1	Input Precision	Every touch input registered	Often not registered	-1	-0.67	-0.5	0.3	0.21	0.51	Simple	Both
M2-TB-1	Susceptibility to Operation Errors	Operation logic and buttons good	Operation logic and buttons unfavorable	-0.83	-0.5	-0.5	0.3	0.15	0.45	Both	Both
M2-TB-2	Expectation Conformity	Contents and operation logic follow conventions	Strong deviation from conventions	-0.83	-0.58	-0.33	0.2	0.21	0.41	Simple	Both
M1-TB-2	Reachability	One hand on the steering wheel, arm not extended	One hand on the steering wheel, arm widely extended	-0.83	-0.58	-0.33	0.2	0.15	0.35	Simple	Both
---	Gesture Control	Menus and functions	No gesture control	-0.33	-0.17	-0.33	0.2	0.04	0.24	Simple	Simple
	Note: Only simple task tested. Bonus item: Optimal implementation tested. Excluded from checklist after validation.										
M2-TB-3	Risk of Operation Errors	No safety risk	Strong distraction	-0.83	-0.42	-0.17	0.1	0.04	0.14	Both	Phone
M2-TB-4	Correction of Operation Errors	One action, immediately	Multiple actions, slowly	-0.92	-0.5	-0.17	0.1	0.04	0.14	Both	Both
M1-TB-3	Driving-related Information Always Visible	Always visible, display separate from touchscreen	Information sometimes covered by touchscreen	-0.83	-0.58	-0.17	0.1	0.04	0.14	Both	Both

Checklist				DP Avg. [-1 (worse),+1 (better)]			Comparison with reference version				
Item	Description	Optimal	Suboptimal	Base.	Radio	Ref.	Net DP	Net OQ	Net Score	Affected Task DP	Affected Task OQ
M2-TB-5	Interaction Speed	Waits indefinitely	Waits briefly, returns, gives instructions	-0.5	-0.33	-0.17	0.1	0.04	0.14	Complex	Complex
	Note: Only complex task tested.										
M2-TB-6	Additional Feedback	On, off, adjust, error – acoustic and visual	Visual feedback only	-0.83	-0.33	-0.08	0.05	0.08	0.12	Simple	Complex
M1-TB-4	Reflection	Matte, strong light absorption	Glossy, anti-reflective coating	-0.83	-0.5	-0.08	0.05	0.06	0.11	Complex	Simple
M1-TB-5	Viewing Distance Downwards	0° – 15°	30° – 45°	-0.83	-0.5	-0.08	0.05	0.06	0.11	Complex	Complex
No effect on driving performance relative to reference version											
M1-TB-6	Tilt	Strong horizontal and vertical tilt	No tilt	-0.83	-0.42	0	0	0.08	0.08	–	Both
M2-TB-7	Legibility	Contents large and high-contrast	Contents too small/large, low-contrast	-0.83	-0.33	0	0	0.04	0.04	–	Both
M1-TB-7	Brightness and Contrast	Optimal, adjustable and automatic	Insufficient, adjustable, no automatic regulation	-0.83	-0.42	0	0	0.04	0.04	–	Complex
M1-TB-8	Attentional Guidance	Interaction principles and guidance optimal	No communication, guidance distracting	-0.83	-0.58	0	0	0.04	0.04	–	Complex
M1-TB-9	Minimalism	All content short and concise	Contents rather long	-0.83	-0.33	0	0	0.02	0.02	–	Simple

Checklist				DP Avg. [-1 (worse),+1 (better)]			Comparison with reference version				
Item	Description	Optimal	Suboptimal	Base.	Radio	Ref.	Net DP	Net OQ	Net Score	Affected Task DP	Affected Task OQ
M2-TB-8	In-process Abortion	Possible at any point	Only at some points and only cumbersome	-0.5	-0.33	0	0	0	0	–	–
	Note: Only complex task tested.										
Possible positive effect on driving performance relative to reference version (likely no effect)											
M1-TB-10	Viewing Distance to the Right	0° – 15°	45° – 60°	-0.83	-0.25	0.08	-0.05	0.02	-0.03	Complex	Complex
M1-TB-11	View on the Road Not Blocked	Clear field of view	Peripheral field of view blocked	-0.83	-0.25	0.08	-0.05	0	-0.05	Both	–
Bonus/Special Items: Positive effect on driving performance relative to reference version											
M2-TB-9	Shortcut	Executable with one interaction	No shortcuts	-0.42	-0.17	0.25	-0.15	-0.06	-0.21	Complex	Complex
	Note: Only simple task tested. Bonus item: Optimal implementation tested.										
M1-TB-12	Deactivation	Yes, can be switched off	Standby interface	0.17	1	0.83	-0.5	0	-0.5	–	–
	Note: No function operation. Effect of static interface tested.										
M1-TB-13	Blocking of Distracting Functions	Dynamic/moving content	No blocking	0	0.92	0.83	-0.5	0	-0.5	–	–
	Note: No function operation. Effect of dynamic interface tested.										

Note. Sorted in descending order by negative relative effect on driving performance compared to the reference version (DP Avg. – Ref.). DP = Driving Performance, OQ = Operation Quality. DP Avg. = Standardized driving performance sum average derived from net score.

7.3.3.2 Design Aspects in Detail

The effect of each design aspect's suboptimal implementation will be discussed in detail in the following section. The operationalization during the validation will be highlighted as well as the specific driving performance and operation quality metrics affected. Also, the differences between the simple and complex task will be explored. The aspects will be discussed in order of effect on driving performance as shown in Table 23.

The aspect M1-TB-1 *Input Precision* had the most negative effect on driving performance and operation quality combined, resulting in the highest net score (see Table 24). The aspect refers to the touch-based interface's ability to reliably detect touch inputs. For the validation experiment, the interface was modified to only process touch inputs with a likelihood of 33%. As a result, multiple presses were now required to operate the functions. Driving performance was strongly negatively affected only for the simple task (MDev, SDLP, RT) but not for the complex task. However, both tasks exhibited clear negative effects both in subjective (EoU, Distr) and objective (Miss, N) operation quality. The results indicate that time-critical single-step function operation suffers when touch-inputs are not handled correctly. That the complex task remained unaffected regarding driving performance may be attributed to the continuous operation nature of the task. Here, drivers had to monitor their inputs and repeatedly press buttons so that additional non-responses of the interface could be more easily handled by pressing a button again. Still, as only the number of inserted numbers (N) was affected but not the error rate (Error), this could point towards a compensation strategy as the task got too difficult. On the other hand, the single task in its optimal implementation only required a single glance and one-action motor movement. The introduction of an unresponsive interface therefore increased the visual-manual effort. Regarding operation errors, the unresponsive touch-input introduced a random error as each touch-input was processed with a chance of 33%. Therefore, participants could not anticipate whether the current operation step would be handled or require additional presses. Looking at the result, this uncertain anticipation likely impacted the simple task more severely as it usually only requires one step and is time-critical. But as mentioned previously, the disparity between both tasks may also be due to a compensation effect for the complex task.

The results strongly imply that touch-based interfaces must reliably register touch inputs, especially when the function only requires a single step, is time-critical, and only used rarely instead of constantly.

Table 24: Touch-Based Interface - Input Precision

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-1		Input Precision		Every touch input registered		Often not registered		Every input registered vs. with only 33% probability	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-2	EoU	-2	MDev	0	EoU	-2		
SDLP	-2	Distr	-2	SDLP	0	Distr	-1		
RT	-2	Miss	-2	RT	0	No	-2		
		Error	0			Error	0		
Sum	-6	Sum	-6	Sum	0	Sum	-5	1.83	0.51
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The aspect M2-TB-1 *Susceptibility to Operation Errors* (see Table 25) had the same negative effect on driving performance as the previously described aspect M1-TB-1 *Input Precision*. However, the impact on operation quality was not as severe so that the net score is lower. The design aspect describes how prone the interface is for operation errors. For the validation, the risk of errors was increased by using smaller buttons and arranging them in an unusual and switched-up constellation for both tasks. Driving performance was negatively impacted for both tasks (simple task: MDev, RT; complex task: RT). Regarding the simple task's operation performance, there were negative effects in all measures of subjective (EoU, Distr) and objective (Miss, Error) operation quality. For the complex task, only objective operation quality was affected (N). As with M1-TB-1, the violation of the design principle M2-TB-1 had a greater negative impact on the simple task compared to the complex task. This is likely due to how the task difficulty was increased for both tasks: For the simple task, the sides of the indicator buttons (left and right) were switched, and they were not arranged on the same horizontal level anymore. For the complex task, the dial pad kept the geometric shape (so that the buttons were aligned unlike the indicator task) and the buttons were arranged in a descending order. Therefore, we argue that the task difficulty increased more for the simple task, especially since its difficulty in the reference version is easier than the complex's task. Finally, participants may also have exhibited compensation behavior for the complex task as the number of inserted numbers (N) increased while the error rate (Error) remained unaffected

The results clearly show that the alignment and sizing of the control elements as well as the operation logic have a clear impact on driver distraction. To mitigate errors, control elements should be aligned and of optimal size. Furthermore, established and easy-to-understand operation logic should be adhered to.

Table 25: Touch-Based Interface - Susceptibility to Operation Errors

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-TB-1		Susceptibility to Operation Errors		Operation logic and buttons good		Operation logic and buttons unfavorable		Buttons optimal vs. smaller, unusually arranged, swapped	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-2	EoU	-1	MDev	0	EoU	0		
SDLP	0	Distr	-2	SDLP	0	Distr	0		
RT	-2	Miss	-1	RT	-2	N	-2		
		Error	-2			Error	0		
Sum	-4	Sum	-6	Sum	-2	Sum	-2	1.62	0.45
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

With the design aspect M2-TB-2 *Expectation Conformity*, the interface's operation logic is required follow established standards (see Table 26). For the validation, the simple task now incorporated a sliding gesture instead of tapping for operation. The complex task now employed a 4x3 dial pad instead of the common 3x4 dial pad. As with the previous aspects, the driving performance and function operation effects were more severe for the simple task. Here, driving performance (MDev, RT) as well as subjective (EoU, Distr) and objective (Miss) operation quality were negatively impacted. On the other hand, there was no effect on driving performance for the complex task, however, both operation quality dimensions (subjective: EoU, Distr; objective: N) were negatively impacted. The lack of effects on driving performance for the complex task may be due to compensation: Participants stopped operation of the complex task because it got too hard. This is likely since the number of inserted numbers (N) decreased while the error rate (Error) and driving performance (MDev, SDLP, RT) remained unaffected. However, the disparity between both tasks could again be attributed to the increase in operation difficulty. The simple task now required sliding over a distance of a few centimeters whereas the complex task still used tapping for input. Therefore, the simple task now required higher visual-manual effort for a single operation action than the complex task.

If the results are interpreted this way, they indicate that function operation should incorporate simple tapping instead of sliding as this requires less visual-manual effort.

Table 26: Touch-Based Interface - Expectation Conformity

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-TB-2		Expectation Conformity		Contents and operation logic follow conventions		Strong deviation from conventions		Conventional operation logic vs. indicator: sliders/circles instead of arrows, phone: 3×4 dial pad	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-2	EoU	-2	MDev	0	EoU	-2		
SDLP	0	Distr	-2	SDLP	0	Distr	-1		
RT	-2	Miss	-2	RT	0	N	-2		
		Error	0			Error	0		
Sum	-4	Sum	-6	Sum	0	Sum	-5	1.62	0.41
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-TB-2 *Reachability* describes how far a driver must reach for the touch-based interface (see Table 27). For validation, the touch-based interface was placed farther away from the participant and placed right next to the LCT driving simulation's monitor. For operation, the right arm now had to be widely extended. Driving performance was negatively impacted only for the simple task (MDev, RT) but not for the complex task. Only subjective operation quality was affected for the simple task (EoU, Distr) while for the complex task, both dimensions were affected (subjective: EoU; objective: N, Error). As with all previous aspects, the effects for the simple task are more severe than for the complex task. As discussed previously, this might be due to compensation strategies, the continuous nature of the complex task, or the greater difficulty increase for the simple task. A compensation strategy is unlikely this time since this time the error rate (Error) is elevated. The most probable explanation is the jump in task difficulty: the simple task was rather easy with a single tap without reaching across the table in its reference version. With increased operation distance, the arm had to be widely extended and that concurrently when the lane change sign occurred. Precise reaching and tapping while at the same time changing lanes require a high level of hand-arm coordination. Also, as there was no function operation between lane changes, some participants switched hands between the steering wheel and the touch-based interface. On the other hand, when interacting with the complex task, the arm remained in the same stretched position and was either rested on the table or lifted up again. Therefore, the distance travelled by the right hand was lower than for the simple task, which could be used for understanding the different effects on driving performance between both tasks.

This way, the results may demonstrate the importance of the distance between the steering wheel and the touch-based interface, especially when drivers must repeatedly reach for the control element. Therefore, the touch-based interface should be kept close to the driver.

Table 27: Touch-Based Interface - Reachability

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-2		Reachability		One hand on the steering wheel, arm not extended		One hand on the steering wheel, arm widely extended		One hand on the steering wheel vs. arm widely extended (55 cm)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-2	EoU	-2	MDev	0	EoU	-1		
SDLP	0	Distr	-2	SDLP	0	Distr	0		
RT	-2	Miss	0	RT	0	N	-2		
		Error	0			Error	-1		
Sum	-4	Sum	-4	Sum	0	Sum	-4	1.55	0.35
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect *Gesture Control* is a bonus item, i.e., it is not an essential design principle of the checklist. It explores additional input methods besides touch-based function operation (see Table 28). It was only tested for the simple indicator task but not for the complex dialing task. Instead of tapping buttons, participants waved in front of the tablet in the direction of the lane change to activate the indicator. Gesture control had a negative impact on driving performance (MDev, RT) and subjective operation quality (Distr) compared to touch-based operation. The results clearly demonstrate that gesture-based function operation is not suitable for simple, single-step tasks. The results cannot be used to draw conclusions for more complex tasks. However, as the manual effort increases (the arm and hand have to be lifted into the air – somewhat similar to M1-TB-2 *Reachability*), it can be expected that there would be no positive effect regarding driver distraction.

Therefore, using gesture control as an additional input method besides the established options is not really a feasible option to mitigate driver distraction. Also, the item was removed from the checklist as it does not contribute to low-distraction function operation (for the full checklist: see Appendix B).

Table 28: Touch-Based Interface - Gesture Control

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
---		Gesture Control		Menus and functions		No gesture control		Index finger gestures (swipe left/right) vs. none	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-2	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	-2	SDLP	0	Distr	0		
RT	-2	Miss	0	RT	0	No	0		
		Error	0			Error	0		
Sum	-4	Sum	-2	Sum	0	Sum	0	1.47	0.24
Negative effect on driving performance									
Note: Only simple task tested. Bonus item: Optimal implementation tested. Excluded from checklist after validation.									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-TB-3 *Risk of Operation Errors* refers to the consequences of an erroneous function operation regarding driver distraction (see Table 29). For validation, a simulated error was triggered during function operation. The error was made salient by incorporating a flashing screen, alternating in green and red colors, and by playing back error sounds. The error could be corrected instantly by pressing a button. Driving performance was only negatively affected for the simple task (MDev, RT) but not for the complex task. Effects on operation performance only occurred for the complex task (objective: N, Error). Notably, there was a slight improvement in SDLP for the complex task. As with the previous aspects, the driving performance detriments only appeared for the simple but not for the complex task. The difference may be attributed to the single- and multiple-step nature of the tasks regarding error handling. The simple task only occurred 18 times per trial and ideally required one operation step. With error handling, however, the operation required twice as many steps and took longer. Also, two button positions had to be visually selected and then pressed. Without error handling, only the indicator arrow needed to be pressed. On the other hand, the complex task was continuous throughout the whole trial and already required visual search and repeated button presses for inserting the digits. Therefore, it is sensible that an additional step in the operation does not result in much greater driver distraction. Another aspect to discuss is that participants were familiar with the occurrence of the operation error. This contrasts with real life operation errors, which most often appear unexpected. As a result, the error could be easily anticipated and handled. Errors that could not be easily anticipated were somewhat tested with M1-TB-1 *Input Precision* and were also used to explain the high levels of driver distraction of M1-TB-1.

Although the results cannot be applied to unexpected error handling, they demonstrate how regularly occurring in-between steps of function operation are handled. The effect for the simple function indicates that additional operation steps and button placements that shift in position between menus should be avoided.

Table 29: Touch-Based Interface - Risk of Operation Errors

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-TB-3		Risk of Operation Errors		No safety risk		Strong distraction		Normal vs. flashscreen + error sound (salient)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-2	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	1	Distr	0		
RT	-1	Miss	0	RT	0	N	-1		
		Error	0			Error	-1		
Sum	-3	Sum	0	Sum	1	Sum	-2	1.38	0.14
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-TB-4 *Correction of Operation Errors* (see Table 30) is based on the design aspect M2-TB-3 *Risk of Operation Errors*. The same simulated errors (flashing screen with red and green colors as well as error sound playback) were used. However, instead of one button press, three presses were needed to cancel the error. The results are similar to M2-TB-3: A negative effect on driver distraction can only be seen for the simple task (MDev, SDLP) but not for the complex task. Here, driving performance again improved slightly (SDLP). Operation quality was only slightly affected for both the simple (subjective: Distr) and complex task (objective: N). Although the complexity for cancelling the error slightly increased (three presses on same button), it did not have any substantial effects on driving performance or operation quality, especially compared to M2-TB-3 where only one press was required. With the effect being essentially the same, the results indicate that pressing a button three times in quick succession requires the same effort as pressing it once. However, as discussed with M2-TB-3 for the simple task, the results would presumably be different if three different positions on the screen had to be tapped instead of pressing on the same position three times. Also as discussed with M2-TB-3, participants were familiar with error handling and could anticipate it.

The results demonstrate that for regular occurring operation errors, pressing a button in quick succession requires the same effort as a single press. The results may be applied to interfaces where a value is changed with quick button presses (e.g., climate temperature, scrolling through a list). In such cases, quick repeated button presses likely do not result in increased driver distraction.

Table 30: Touch-Based Interface - Correction of Operation Errors

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-TB-4		Correction of Operation Errors		One action, immediately		Multiple actions, slowly		Direct cancellation possible vs. pressing cancel button 3 times	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-2	EoU	0	MDev	0	EoU	0		
SDLP	-1	Distr	-1	SDLP	1	Distr	0		
RT	0	Miss	0	RT	0	No	-1		
		Error	0			Error	0		
Sum	-3	Sum	-1	Sum	1	Sum	-1	1.34	0.14
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-TB-3 *Driving-related Information Always Visible* describes that driving-related information (e.g., speed, assistance status) is always visible for the driver (see Table 31). For validation, the design principle was violated by placing the touch-based interface in such a way that it covered just the speedometer of the LCT. The driving environment remained largely visible. There were only slight negative effects on driving performance for both the simple and complex task (both RT). Likewise, there were only minor effects on operation quality (simple task: EoU; complex task: N). There are several limitations to the validation approach of this design principle. First, the LCT does not require the driver to constantly check their speed as it is limited to the required 60 km/h. Second, with moving the interface, not only was the visibility of the driving-related information altered but also the viewing distance between the driving simulation and the touch-based interface and the reachability. Together with lack of a distinct effect, the validation cannot be used to draw conclusions on the visibility on driving-related information.

For real driving, however, this principle is essential as drivers need to be in-the-loop and stay informed about the current driving status. Therefore, in cases where driving-related information is presented in the touch-based interface, it should always be visible (e.g., separate areas for displaying such information).

Table 31: Touch-Based Interface - Driving-related Information Always Visible

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-3		Driving-related Information Always Visible		Always visible, display separate from touchscreen		Information sometimes covered by touchscreen		Display separate vs. covered by touchscreen	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	-1	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	-1	Miss	0	RT	-1	N	-1		
		Error	0			Error	0		
Sum	-1	Sum	-1	Sum	-1	Sum	-1	1.14	0.14
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-TB-5 *Interaction Speed* specifies how functions requiring multiple steps pace their interaction with the driver (see Table 32). More specifically, the interaction speed should be defined by the driver and not the function. For the validation, the design principle was tested with the complex task only. Here, the phone number had to be put in completely in under three seconds, otherwise the input would be cleared, and drivers had to start inserting the digits again. Driving performance was negatively affected (MDev) as was subjective operation performance (Distr). Interestingly, objective operation performance remained unaffected as the number of numbers put in (N) did not decrease. The results point towards drivers focusing on inputting the numbers and deprioritizing the driving task.

Therefore, the results indicate that multiple-step function operation should wait for driver input. Otherwise, drivers will more intensively focus on function operation instead of driving, resulting in higher levels of driver distraction.

Table 32: Touch-Based Interface - Interaction Speed

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-TB-5		Interaction Speed		Waits indefinitely		Waits briefly, returns, gives instructions		Waits indefinitely vs. reset after 3s (re-enter number)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	-2	EoU	0		
SDLP	0	Distr	0	SDLP	0	Distr	-2		
RT	0	Miss	0	RT	0	N	0		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	-2	Sum	-2	1.27	0.14
Negative effect on driving performance									
Note: Only complex task tested.									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-TB-6 *Additional Feedback* refers to functions incorporating multiple channels of feedback accompanying the operation process (see Table 33). Ideally, feedback beyond the visual channel should be used to reduce glances away from the road. For validation, an interface without acoustic feedback was used, thereby violating the design principle. This had only a slight negative effect on driving performance for the simple task (MDev) and no effect for the complex task. Overall, operation quality was not affected for the simple task (sum of zero), however, there were clear effects for the complex task (subjective: EoU, Distr; objective: N). Regarding driver distraction, only the simple task was affected but due to the lack of differences in qualitative operation quality, this effect is most likely random instead of being indicative of a driving performance detriment. With operation quality only being impacted for the complex task, the provision of extra feedback seems to be helpful for distinguishing between the steps in a multiple-step function operation. Obviously, this distinction is not important when the function is single-step. There are some considerations to make: First, the simple task was an indicator task. In real life, the operation of such an important function should never omit auditory feedback, especially as it would require visual effort to check whether the function has been successfully activated. Second, in our experiment, participants were very familiar with the operation of the functions as they regularly interacted with them. The function usage likely reached an overlearned state. As a result, omitting the auditory feedback did not really alter the visual-manual resource usage in the function operation process. The results therefore should not be used to overly-omit auditory feedback as function usage in real-life is done at much lower rates.

Overall, the results demonstrate that, from an ergonomic perspective, auditory feedback is important for operating multiple-step functions. Although there was no clear effect on driving performance (likely because function operation was overlearned), we do not suggest omitting auditory feedback. Especially as it allows for the communication of information without visual effort, thereby mitigating visual-manual distraction.

Table 33: Touch-Based Interface - Additional Feedback

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-TB-6		Additional Feedback		On, off, adjust, error – acoustic and visual		Visual feedback only		Additional feedback vs. tablet muted	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-1	EoU	-1	MDev	0	EoU	-2		
SDLP	0	Distr	1	SDLP	0	Distr	-1		
RT	0	Miss	0	RT	0	N	-1		
		Error	0			Error	0		
Sum	-1	Sum	0	Sum		Sum	-4	1.25	0.12
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-TB-4 *Reflection* specifies the reflective properties of the touchscreen (see Table 34). For validation, reflections towards the drivers were simulated by slightly blinding participants with the outer light beam of a torchlight. This had no distinct effect on driving performance (only SDLP for the complex task). Operation quality was slightly affected: Subjective measures for the simple task (EoU, Distr) and subjective as well as objective measures for the complex task (EoU, N). The results indicate that when reflected light hits the driver's eyes, it does not really result in increased driver distraction while operating functions. However, the validation did not incorporate full screen reflections that would likely more negatively impact the readability of the screen. Also, function operation was well learned so that participants did not perform an exhaustive visual search for control elements. Therefore, the results need to be treated carefully.

Screens should still reduce glare and reflections to make it easy for drivers to quickly find their desired control elements.

Table 34: Touch-Based Interface - Reflection

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-4		Reflection		Matte, strong light absorption		Glossy, anti-reflective coating		No illumination vs. flashlight from the side	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	-1	MDev	0	EoU	2		
SDLP	0	Distr	-2	SDLP	-1	Distr	0		
RT	0	Miss	0	RT	0	N	-2		
		Error	0			Error	0		
Sum		Sum	-3	Sum	-1	Sum	0	1.21	0.11
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-TB-5 *Viewing Distance Downwards* refers to the vertical placement of the touchscreen (see Table 35). In the validation, the design principle was violated by placing the touchscreen at a lower position representing a center stack placement. This had only a slight negative effect on driving performance for one measure for the complex task (SDLP). Otherwise, there were no effects on driving performance for the simple task. Operation quality was not impacted for the simple task, however, both subjective (EoU) and objective (N) operation quality were negatively affected for the complex task. The results indicate that the vertical placement of the touchscreen does not substantially influence driver distraction beyond the base level (reference version). Still, the effects point towards the complex task being impacted to a greater degree than the simple task regarding operation quality. Also, as the number of inserted numbers (N) decreased, this could point towards a possible compensation effect. Perhaps, the complex task got too difficult so that drivers focused on driving instead of function operation.

Overall, the touchscreen's vertical placement should be close to the driver's line of sight to reduce the viewing distances between the road and the touchscreen.

Table 35: Touch-Based Interface - Viewing Distance Downwards

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-5		Viewing Distance Downwards		0° – 15°		30° – 45°		Tablet next to steering wheel vs. lower at center stack height	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	-1		
SDLP	0	Distr	0	SDLP	-1	Distr	0		
RT	0	Miss	0	RT	0	N	-2		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	-1	Sum	-3	1.21	0.11
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design characteristic M1-TB-6 *Tilt* demands that the touchscreen should be tilted towards the driver to reduce visual-manual effort (see Table 36). For the validation, the horizontal and vertical tilt towards the driver was removed. To this end, the screen was aligned with the table edge of the experimental setup and tilted upright. The violation of the design principle did not result in increased driver distraction. For the simple task, subjective operation quality was somewhat affected (EoU, Distr), and for the complex task, objective operation quality (N) was impacted. As there was no clear effect on driver distraction, the design principle does not appear to be essential in the context of driver distraction. However, operation quality was affected and here, for the complex task, the number of inserted numbers decreased. This result indicates increased task difficulty but it could also point towards compensation where drivers focused on driving instead of function operation.

Both considerations suggest that the screen should still be tilted towards the driver, although tilt is not a main contributor to driver distraction.

Table 36: Touch-Based Interface - Tilt

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-6		Tilt		Strong horizontal and vertical tilt		No tilt		Tilted horizontally/vertically towards driver vs. no tilt	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	-1	MDev	0	EoU	0		
SDLP	0	Distr	-1	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	-2		
		Error	0			Error	0		
Sum	0	Sum	-2	Sum	0	Sum	-2	1.07	0.08
No effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-TB-7 *Legibility* refers to the design of the touch-based interface's contents that affect legibility (see Table 37). For validation, legibility was reduced by making all buttons and text smaller and in a light-grey color. The violation of the design principle had no effect on driving performance or on objective operation quality for both tasks. Only subjective operation quality was slightly affected (EoU). For the interpretation of the results, it is important to consider that participants were familiar with the interface and had learned to operate both tasks well. Therefore, it is expected that switching to a low-contrast interface did not increase visual demand by a great magnitude.

However, this might be different if a driver has to operate a function that they are not familiar with. In such cases, it could be expected that a low-contrast interface leads to greater visual demand and increased driver distraction. Therefore, we recommend using a high-contrast and easy-to-read interface design.

Table 37: Touch-Based Interface - Legibility

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-TB-7		Legibility		Contents large and high-contrast		Contents too small/large, low-contrast		Large, high-contrast vs. icons/text light grey, very small	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	-1	MDev	0	EoU	-1		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	0		
		Error	0			Error	0		
Sum	0	Sum	-1	Sum	0	Sum	-1	1.04	0.04
No effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-TB-7 *Brightness and Contrast* refers to the brightness and contrast abilities of the touchscreen (see Table 38). For the validation, the brightness was set to a very low setting so that reflections became visible on the screen. This had no effect on driving performance. Only objective operation quality was affected for the complex task (Error). As with the previous aspect M2-TB-7 *Legibility*, it has to be noted that participants were very familiar with the function operation process. Therefore, it was expected that reducing the screen brightness over a single trial would not have a major impact. This might be different when the function menu is only used rarely.

The results have two implications: First, if drivers are familiar with the interface and the brightness is too low for a short amount of time, it does not greatly increase visual effort. This might happen when driving through dark areas (e.g., under a bridge) so that the interface dims its brightness. When entering bright areas again it might be too dark for a limited time. Second, brightness should be kept optimal if the function's operation is not well learned as this presumably leads to increased visual demand and therefore driver distraction.

Table 38: Touch-Based Interface - Brightness and Contrast

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-7		Brightness and Contrast		Optimal, adjustable and automatic		Insufficient, adjustable, no automatic regulation		Optimal (80% brightness) vs. dark (20% brightness)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	0		
		Error	0			Error	-2		
Sum	0	Sum	0	Sum	0	Sum	-2	1.07	0.04
No effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-TB-8 *Attentional Guidance* refers to how attention is guided within the interface (see Table 39). It should clearly communicate what buttons are relevant for function operation by their sizing, placement, and coloring. For the validation, the attentional guidance was disrupted by inserting colored text and geometric objects acting as distractors. This had no effect on driving performance for both tasks. Objective operation quality was negatively impacted for the complex task (N).

As with the previous aspects M2-TB-7 *Legibility* and M1-TB-8 *Brightness and Contrast*, function operation was well learned so that minimal effects were expected. For less frequently used functions, results are likely different. Therefore, we advise incorporating a visual distinct interface that guides attention towards the relevant elements of function operation.

Table 39: Touch-Based Interface - Attentional Guidance

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-8		Attentional Guidance		Interaction principles and guidance optimal		No communication, guidance distracting		Concise interface vs. salient distractors (colorful)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	-2		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	0	Sum	-2	1.07	0.04
No effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-TB-9 *Minimalism* demands that an appropriate amount of content is placed on the screen (see Table 40). Ideally, the interface is not cluttered. For the validation, distractor text and objects were randomly placed on the screen. These were the same distractors as used for the validation of M1-TB-8 *Attentional Guidance*, however, their opacity was increased and shades of light grey instead of colors were used. For the complex task, there was an effect on driving performance, however, it amounts to zero (SDLP worse, RT better). Operation quality was only slightly affected for the simple task (Miss). As with all previous perceptive design aspects (M2-TB-7 *Legibility*, M1-TB-7 *Brightness and Contrast*, M1-TB-8 *Attentional Guidance*) function usage was well learned and is therefore not indicative of more sparse function interactions like in real life.

Together with the effects on driving performance for the complex task, we strongly advise to not overload the visual interface with text or other elements.

Table 40: Touch-Based Interface - Minimalism

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-9		Minimalism		All content short and concise		Contents rather long		Concise interface vs. vs. non-salient distractors (grey)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	-2	Distr	0		
RT	0	Miss	-1	RT	2	N	0		
		Error	0			Error	0		
Sum	0	Sum	-1	Sum	0	Sum	0	1.04	0.02
No effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The aspect M2-TB-8 *In-process Abortion* describes that a function operation process with multiple steps can be aborted at any point (see Table 41). For the validation of the principle only the complex task was tested. Here, the number had to be fully inserted before any digit or the full number could be deleted. There were no effects on driving performance and operation quality. Most likely this is due to participants rarely making mistakes during the operation process. With errors occurring very rarely, the alteration of this design principle had no effects.

However, function operation in real life is not that well learned so that errors will always occur at some points. Having to deal with error handling would increase operation steps (see M2-TB-3 *Risk of Operation Errors* & M2-TB-4 *Correction of Operation Errors*) that would most likely negatively impact driving performance. Therefore, we strongly advise that function operation can be aborted at any point.

Table 41: Touch-Based Interface - In-process Abortion

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-TB-8		In-process Abortion		Possible at any point		Only at some points and only cumbersome		Partial/full deletion possible vs. at least 11 digits required	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	0		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	0	Sum	0	1	0
No effect on driving performance									
Note: Only complex task tested.									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-TB-10 *Viewing Distance to the Right* describes how far to the right the driver must view and reach for the interface (see Table 42). For the validation, the interface was placed on a stand and put to the right of the participant. The screen was the same height as when placed on the table (reference version) and was tilted towards the driver. This had no negative effect on driving performance. For the complex task, there was a slight positive effect on driving performance (MDev). Regarding operation quality, there was no effect for the simple task. For the complex task, there was a slight positive effect on subjective operation quality (EoU) but a negative effect on objective operation quality (N). For the complex task's operation, the results largely indicate that drivers engaged less with the interface (N) so that this might have led to the slight improvement in driving performance (MDev).

The results indicate that increasing the viewing distance to the right does not considerably increase driver distraction beyond the reference level. Still, we advise keeping the interface close to the driver's field of view.

Table 42: Touch-Based Interface - Viewing Distance to the Right

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-10		Viewing Distance to the Right		0° – 15°		45° – 60°		Tablet next to steering wheel vs. to the right of the participant	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	1	EoU	1		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	-2		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	1	Sum	-1	1.04	-0.03
Possible positive effect on driving performance (likely no effect)									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-TB-11 *View on the Road Not Blocked* demands that the touchscreen does not block the view on the road (see Table 43). For the validation, the tablet was placed in such a way that the bottom right-hand corner of the LCT's screen was obstructed. There was only a slight negative effect on driving performance for the simple task (RT) but a positive effect for the complex task (MDev). Also, for the complex task, there were a positive (EoU) as well as a negative effect (N) on operation quality. Overall, the results point towards marginal but not really distinct effect. The results are somewhat comparable to M1-TB-3 *Driving-related Information Always Visible*. There, the negative effect on driving performance was more distinct and only negative. The deviation could be attributed to several factors: First, the bottom right half of the LCT was not visible, which might have impacted lane tracking and maneuvering performance. This might be somewhat visible in the simple task's driving performance. Second, with the changed screen placement, viewing and reaching distances were altered. Although participants reported a better EoU for the complex task, the number of inserted numbers (N) decreased. As the driving performance increased at the same time, this most likely suggests participants were compensating by focusing on driving instead of operating the function.

In light of these considerations, we advise to not placing the screen in a way that is too close to the driver's line of sight where it blocks the view on the road.

Table 43: Touch-Based Interface - View on the Road Not Blocked

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-11		View on the Road Not Blocked		Clear field of view		Peripheral field of view blocked		Tablet to the right of steering wheel vs. in front of monitor (view blocked)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	2	EoU	1		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	-1	Miss	0	RT	0	N	-1		
		Error	0			Error	0		
Sum	-1	Sum	0	Sum	2	Sum	0	1.1	-0.05
Possible positive effect on driving performance (likely no effect)									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-TB-9 *Shortcut* is a bonus item, i.e., it is not an essential design principle of the checklist. It refers to an option to shorten the operation of a complex function (see Table 44). For the validation, the operation of the complex function was altered by calling contacts via a shortcut instead of inserting a full number. As expected, there were positive effects on driving performance (MDev, SDLP) and subjective operation quality (EoU, Distr). The character of the function essentially changed from a multiple-step to a single-step function.

The results indicate that including shortcuts to skip operation steps is a sensible option for reducing driver distraction.

Table 44: Touch-Based Interface - Shortcut

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-TB-9		Shortcut		Executable with one interaction		No shortcuts		Call Max via shortcut vs. enter full number	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	1	EoU	2		
SDLP	0	Distr	0	SDLP	2	Distr	1		
RT	0	Miss	0	RT	0	No	0		
		Error				Error			
Sum	0	Sum	0	Sum	3	Sum	3	1	-0.21
Positive effect on driving performance									
Note: Only simple task tested. Bonus item: Optimal implementation tested.									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-TB-12 *Deactivation* demands that the touch-based interface can ideally be turned off (see Table 45). For the validation, the suboptimal version (standby interface) was tested by displaying the tablet's homescreen. Participants operated neither function and were instructed to focus on driving. The no-operation trials were measured for both task conditions to allow for comparisons to each respective reference version. Driving performance was clearly better than with operation (MDev, SDLP, RT), which was expected. In the aspect overview (see Table 23), driving performance is clearly better than radio operation and even a little better than baseline driving. This indicates drivers actively ignored the interface and focused more intensively on driving. This could also be due to learning effects as the baseline was measured before (learning experience only from the learning block) and after all aspect measurement trials (learning throughout all measurement trials). The aspect M1-TB-12's trial was therefore measured somewhere in between so that it could possibly depict the ongoing learning progress leading to improved driving performance, at least to some degree.

The results indicate that displaying a standby interface is not really a distraction source compared to the screen being completely turned off (baseline).

Table 45: Touch-Based Interface - Deactivation

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-12		Deactivation		Yes, can be switched off		Standby interface		Baseline vs. standby interface (home screen)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	2	EoU		MDev	2	EoU			
SDLP	0	Distr		SDLP	2	Distr			
RT	2	Miss		RT	2	N			
		Error				Error			
Sum	4	Sum	---	Sum	6	Sum	---	1	-0.5
Positive effect on driving performance									
Note: No function operation. Effect of static interface tested.									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-TB-13 *Blocking of Distracting Functions* refers to whether the interface contains distracting elements (see Table 46). For the validation, we were interested in the effect of dynamic content on driver distraction while not operating a function. On the tablet, a video was played and participants were instructed to focus on driving. As with M1-TB-12 *Deactivation*, this was measured in both function operation task conditions to allow for a comparison to the respective reference versions. The results are similar to M1-TB-12: Driving performance is better when no function is operated (MDev, SDLP, RT). In the aspect overview (see Table 23), driving performance is better than radio operation and as good as baseline driving. Compared to M1-TB-12 that used a static interface, driving performance did not improve as much. Therefore, the dynamic stimuli induced stronger distraction than the static interface.

This result clearly indicates that when drivers are not operating a function, the screen should display static content (or be turned off) instead of displaying dynamic and salient stimuli.

Table 46: Touch-Based Interface - Blocking of Distracting Functions

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-TB-13		Blocking of Distracting Functions		Dynamic/moving content		No blocking		Baseline vs. dynamic interface (video without sound)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	2	EoU		MDev	2	EoU			
SDLP	0	Distr		SDLP	2	Distr			
RT	2	Miss		RT	2	N			
		Error				Error			
Sum	4	Sum	---	Sum	6	Sum	---	1	-0.5
Positive effect on driving performance									
Note: No function operation. Effect of static interface tested.									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

7.3.4 Interface Design Aspects: Speech-Based Interface

The analysis of the speech-based design aspects is split into two parts: The first section provides an overview of the design aspects tested and compares their relative effect on driving performance. The second part analyzes every design aspect in detail. Here, effects on driving performance and operation quality (objective and subjective) are discussed as well as implications for function implementation. The design aspects are described in more detail in the checklist (see Appendix A). The specific operationalization of each design aspect in the experiment is described in Appendix B.

7.3.4.1 Overview and Comparison to All Three Distraction Reference Levels

Table 47 provides an overview of all design aspects that were tested for the speech-based interface. The design aspects are sorted in two ways in relation to the reference interface. First, they are arranged by their relative effect on driving performance (standardized driving performance sum average: DP Avg. – Baseline, see section 7.3.1). Second, if two or more aspects have the same effect on driving performance, they are sorted by net score. For both sorting principles, the aspect with the more negative effect on driving performance and operation quality is listed first. Three groups of aspects can be derived: Those with a negative relative effect on driving performance, those with no effect, and finally, those who might improve driving performance relative to the reference interface but likely have no effect. For easy orientation, the driving performance comparison measures (DP Avg.) are colored: negative relative effects on driving performance are marked orange, no effects are marked grey, and improvements are marked green. The most distracting interface design aspects are listed at the top of the table.

All design aspects were tested via a suboptimal implementation in the reference version of the interface. In the global analysis (see section 7.2.1), the speech-based interface had the least negative effect on driving performance in all three LCT measures (MDev, SDLP, RT) across both tasks. The driving performance was nearly the same as in the baseline (no distraction) and even better than radio operation (acceptable distraction). Therefore, it is expected that most aspects led to distraction levels that are slightly worse than the baseline but better than the radio. When analyzing all three groups of design aspects regarding their effect on driving performance (negative effect, no effect, positive effect), it is surprising that the volume of the speech-based interface (M1-SB-7) had a positive effect on driving performance. However, it is a well-known effect that when listening to auditory information becomes more involved, drivers tend to compensate and tend to focus on the road ahead (cognitive tunneling; Victor et al., 2005). In the LCT, this perhaps leads to improved driving performance. Regarding the quality of the speech output, it is interesting that properties such as intelligibility (M1-SB-5) and chimes (M1-SB-6) had no effect while only the length of the speech output (M2-SB-6) did. This indicates that length of speech-based information is more important than its structural properties.

In comparison to the reference version, eleven of the 16 design aspects led to a worse driving performance. For nearly all aspects, there were also negative effects on operation quality. Regarding driving performance, the simple and complex tasks seem to have been equally affected across all aspects. For the simple task, driving performance detriments can be seen in MDev and RT, whereas for the complex task, SDLP is additionally affected. The deviation in SDLP makes perfect sense as the simple task was not executed during lane keeping phases. The magnitude of effects for all three driving performance measures appears to be equal. For both tasks, subjective operation quality was impacted to a greater extent compared to objective operation quality. Three aspects did not differ in their effect on driving performance compared to the baseline. For two of the aspects, there were some positive deviations (improvements) in subjective and objective operation quality. Finally, there is one aspect that led to better driving performance when compared to the reference version. However, since the effect is only small and only concerns one measure (MDev), the aspect most likely is on the same distraction level as the reference version of the interface.

In the following section, each interface design aspect will be discussed in detail. The global results revealed that speech-based function operation results in comparatively low levels of driver distraction. Furthermore, this effect is consistent across the simple and complex task. Given this promising result, it is important to ensure the design is

optimal by mitigating design issues that are likely to lead to driver distraction. Even with the low general level of driver distraction, the minimum of driver distraction should be the goal. As with both previously discussed interface types, our validation tested very specific implementations and violations of the speech-based interface's design principles. Also, only a simple and complex but no medium complexity task were tested. The results must therefore be transferred cautiously to other speech-based interfaces and function operation processes.

Table 47: Validation Results for the Design Aspects of the Speech-Based Interface

Checklist				DP Avg. [-1 (worse),+1 (better)]			Comparison with reference version				
Item	Description	Optimal	Suboptimal	Base.	Radio	Ref.	Net DP	Net OQ	Net Score	Affected Task DP	Affected Task OQ
Negative effect on driving performance relative to reference version											
M2-SB-1	Fallback	Correct according to decision tree	Incorrect according to decision tree	-0.83	-0.67	-0.83	0.5	0.19	0.69	Both	Both
Note: Incorrect fallback tested (simple task: touch-based, complex task: physical controls).											
M2-SB-2	Correction of Operation Errors	One command, barge-in, immediately	Multiple commands, wait for assistant, no feedback	-0.83	-0.25	-0.42	0.25	0.09	0.34	Complex	Both
M1-SB-1	Deactivation Option & Alternative Operation Option	Speech-based HMI deactivatable, fallback correct (decision tree)	Speech-based HMI not deactivatable, incorrect fallback (decision tree)	-0.83	-0.17	-0.42	0.25	-0.04	0.21	Complex	Both
Note: Correct fallback tested (simple task: physical controls, complex task: touch-based).											
M1-SB-2	Quality of the Speech Recognition	Every voice input recognized	Prone to dialects and background noise	-0.83	-0.25	-0.33	0.2	0.13	0.33	Both	Both
M2-SB-3	Susceptibility to Operation Errors	Function reliably operable	Unreliable, other functions are activated	-0.92	-0.33	-0.33	0.2	0.09	0.29	Complex	Complex
M1-SB-3	Hands-free	Initial activation via key phrase	Push-to-talk button before each input on the steering wheel	-0.83	-0.25	-0.33	0.2	0.02	0.22	Simple	Simple

Checklist				DP Avg. [-1 (worse),+1 (better)]			Comparison with reference version				
Item	Description	Optimal	Suboptimal	Base.	Radio	Ref.	Net DP	Net OQ	Net Score	Affected Task DP	Affected Task OQ
Possible positive effect on driving performance relative to reference version (likely no effect)											
M1-SB-7	Volume	Default volume appropriate, adjustable, automatic	Default volume too loud/quiet, adjustable, no automatic	-0.58	0.25	0.08	-0.05	0.02	-0.03	Complex	Complex

Note. Sorted in descending order by negative relative effect on driving performance compared to the reference version (DP Avg. – Ref.). DP = Driving Performance, OQ = Operation Quality. DP Avg. = Standardized driving performance sum average derived from net score.

7.3.4.2 Design Aspects in Detail

The effect of each design aspect's suboptimal implementation will be discussed in detail in the following section. The operationalization during the validation will be highlighted as well as the specific driving performance and operation quality metrics affected. Also, the differences between the simple and complex task will be explored. The aspects will be discussed in order of effect on driving performance as shown in Table 47.

The design aspect M2-SB-1 *Fallback* addresses when speech-based function operation needs to switch to another HMI modality (see Table 48). This might be necessary when the speech-based interface is not available or outside circumstances (e.g., noise) make speech-based operation impossible. For the validation, incorrect fallback HMI modalities were used (see decision tree, see section 5.2.3). The simple task had to be operated via the reference version of the touch-based interface. The complex task needed to be operated via the reference version of the physical controls. Therefore, the test of this interface design aspect is a direct comparison of modalities that resembles the comparison in the global analysis. As expected, driving performance was affected considerably, both for the simple task (MDev, SDLP) as well as for the complex task (MDev, SDLP, RT). Furthermore, subjective operation quality was fully affected for both tasks (EoU, Distr). Objective operation quality was only somewhat negatively affected (Error). The results clearly demonstrate that speech-based operation is superior regarding driver distraction and operation quality. It was expected that the simple and complex tasks would be equally negatively affected as incorrect HMI modalities were used for their operation. As seen in the global results (see section 7.2), operating the complex task with physical controls led to the highest levels of driver distraction. Regarding the simple task, driving performance did not differ between HMI modalities in the global analysis, however, in this design aspect's test it can be clearly seen that the touch-based operation is less suitable for the operation of the simple task. The correct HMI modalities were tested with the design aspect M1-SB-1 *Deactivation Option & Alternative Operation Option*.

The results clearly demonstrate that incorrect HMI modalities should not be used as a fallback option for the speech-based interface. For simple tasks, physical controls are the correct choice, while touch-based interfaces are the best choice for complex tasks. We advise against omitting fallback altogether as there will always be cases where the speech-based interface is not available or suitable.

Table 48: Speech-Based Interface - Fallback

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-SB-1		Fallback		Correct according to decision tree		Incorrect according to decision tree		Incorrect fallback: indicator via touch, phone via physical controls	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-2	EoU	-2	MDev	-2	EoU	-2		
SDLP	0	Distr	-2	SDLP	-2	Distr	-2		
RT	-2	Miss	0	RT	-2	N	1		
		Error	-1			Error	-2		
Sum	-4	Sum	-5	Sum	-6	Sum	-5	1.79	0.69
Negative effect on driving performance									
Note: Incorrect fallback tested (simple task: touch-based, complex task: physical controls).									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-SB-2 *Correction of Operation Errors* is concerned with how an erroneous function operation can be cancelled (see Table 49). The operation error was simulated by displaying a flashing screen with alternating red and green colors and by playing back error sounds. For the validation, the cancellation process was cumbersome as the speech-based interface did not directly cancel the function but instead asked “Are you sure?”. Participants needed to answer with “Yes” to confirm the cancellation. Alternatively, they could cancel the error by pressing a button on the touchscreen three times. The simple task was essentially unaffected as there were no effects on driving performance and only a marginal effect in a single operation quality measure (Distr). In contrast, the complex task was more severely affected, both regarding driving performance (MDev, SDLP, RT) as well as operation quality (subjective: EoU, Distr). It makes perfect sense that the effect is only visible for the complex task. For one, it was a continuous operation task compared to the simple task that was only performed at each of the 18 lane changes per trial. For another, the task was more error prone as drivers needed to remember and put in a phone number. As there was no effect on objective operation quality (N, Error), this indicates that at least operation speed did not slow down, which points towards participants switching from speech-based to button-based cancellation. So, this design aspect highlights an important facet regarding function cancellation: When errors occur in the speech-based interface and speech-based cancellation is slow, participants will switch over to the touch-based cancellation option. That makes sense, as when the speech-based interface fails, drivers likely switch to a more reliable interface that also provides faster operation speed. However, this introduces visual-manual distraction, which is detrimental to driving safety.

So, the results indicate that it is best when the speech-based interface is designed in a way that operation errors do not even occur in the first place. If errors do occur, however, cancellation via the speech-based interface must be easy, otherwise drivers will rely on other HMI modalities for cancellation. Cancellation via another HMI modality should also be as easy as possible.

Table 49: Speech-Based Interface - Correction of Operation Errors

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-SB-2		Correction of Operation Errors		One command, barge-in, immediately		Multiple commands, wait for assistant, no feedback		Direct cancellation via button/speech vs. cancellation with confirmation prompt or pressing button 3 times	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	-2	EoU	-2		
SDLP	0	Distr	-1	SDLP	-1	Distr	-2		
RT	0	Miss	0	RT	-2	N	0		
		Error	0			Error	0		
Sum	0	Sum	-1	Sum	-5	Sum	-4	1.65	0.34
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-SB-1 *Deactivation Option & Alternative Operation Option* is concerned with alternative access options to a function in case the speech-based interface is not available for operation (see Table 50). It is closely connected to the design principle M2-SB-1 *Fallback*. For the validation, correct fallback HMI modalities (see decision tree, section 5.2.3) were used. The simple task was operated via physical controls, the complex task via the touch-based interface. In both cases, the reference version of the interface was used. For the simple task, the driving performance remained unaffected, however, subjective operation quality improved (EoU, Distr). For the complex task, the switch in modality had a negative impact on driving performance (MDev, SDLP, RT) and subjective operation quality (Distr). For objective operation quality, more numbers were inserted (N) but also more errors made (Error). With the correct modality used, the simple task could be operated just as well as via the speech-based interface and participants clearly preferred this way of function operation. This extends the results of the global analysis where the interface types did not differ from one another. It also aligns with the decision tree where physical controls are intended as the primary access option. For the complex task, the touch-based interface led to results worse than the speech-based interface. This is in line with the global results and with the decision tree where primary function access shall be provided via the speech-based interface.

Taken together with the results of the validation of M2-SB-1, the correct fallback modality must be available for function operation in case the speech-based interface is not available.

Table 50: Speech-Based Interface - Deactivation Option & Alternative Operation Option

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-SB-1		Deactivation Option & Alternative Operation Option		Speech-based HMI deactivatable, fallback correct (decision tree)		Speech-based HMI not deactivatable, incorrect fallback (decision tree)		Correct fallback: indicator via physical controls, phone via touch	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	2	MDev	-2	EoU	0		
SDLP	0	Distr	2	SDLP	-1	Distr	-2		
RT	0	Miss	1	RT	-2	N	2		
		Error	-1			Error	-2		
Sum	0	Sum	4	Sum	-5	Sum	-2	1.57	0.21
Negative effect on driving performance									
Note: Correct fallback tested (simple task: physical controls, complex task: touch-based).									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

With the design aspect M1-SB-2 *Quality of the Speech Recognition*, the speech-based interface's ability to correctly recognize voice inputs is addressed (see Table 51). For the validation, the processing of the voice inputs was altered so that in 70% of cases the simple or complex function would be operated incorrectly. For the simple task, there was only a slight negative effect on driving performance (RT) and some slight negative effects on operation quality (subjective: EoU, Distr). The effects are more distinct for the complex task, both regarding driving performance (MDev, SDLP) and operation quality (subjective: EoU, Distr; objective: N). The stronger effects for the complex task can be explained in two ways. First, the operation errors only occurred within the function. For the simple task, there were only two operation options (indicating left or right) so that the error would be indicating in the opposite direction. For the complex task, however, there were more options for errors (calling another number, calling a shortcut contact, cancelling a nonexistent call). This might have led to greater distraction as the errors were more diverse and could not be as easily anticipated as for the simple task. Secondly, function operation was continuous whereas the simple task was only executed at each of the 18 lane changes in the LCT trial.

The results clearly demonstrate that operation errors in the form of not activating the target function lead to driver distraction. Furthermore, this effect is more pronounced if the errors are not clearly anticible. For speech-based interfaces, user inputs must therefore be reliably and correctly recognized.

Table 51: Speech-Based Interface - Quality of the Speech Recognition

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-SB-2		Quality of the Speech Recognition		Every voice input recognized		Prone to dialects and background noise		Reliable vs. 70% incorrect command processing	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	-1	MDev	-2	EoU	-2		
SDLP	0	Distr	-1	SDLP	-1	Distr	-2		
RT	-1	Miss	1	RT	0	N	-2		
		Error	0			Error	0		
Sum	-1	Sum	-1	Sum	-3	Sum	-6	1.53	0.33
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-SB-3 *Susceptibility to Operation Errors* refers to how likely the interface introduces errors in the operation process of a given function (see Table 52). It is closely linked to M1-SB-2 *Quality of the Speech Recognition*. While M1-SB-2 describes the general ability of the interface to process voice commands, M2-SB-3 is aimed more at the function itself. For validation, the same logic as M1-SB-2 was embedded into the speech-based interface: In 70% of cases, another function was activated. However, this time any function in the generic speech-based interface could be activated, not just a subfunction of the simple or complex task. For example, the music player or climate control functions could be erroneously activated. Like M1-SB-2, the complex task was more affected than the simple task. For the simple task, there were no effects on driving performance or on operation quality. For the complex task, both driving performance (MDev, SDLP) and operation quality (subjective: EoU, Distr; objective: N) were affected. As discussed with M1-SB-2, this is likely due to the higher complexity and continuous operation nature of the complex task.

The results further underline that errors in speech-based operation should be avoided, especially at higher task complexities.

Table 52: Speech-Based Interface - Susceptibility to Operation Errors

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-SB-3		Susceptibility to Operation Errors		Function reliably operable		Unreliable, other functions are activated		Reliable vs. 70% other subfunction activated	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	-2	EoU	-1		
SDLP	0	Distr	0	SDLP	-2	Distr	-2		
RT	0	Miss	0	RT	0	N	-2		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	-4	Sum	-5	1.59	0.29
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-SB-3 *Hands-free* specifies the initial activation of the speech-based interface. To mitigate visual-manual distraction, it should be activated without the use of buttons (see Table 53). For the validation, the design principle was violated by having participants push a steering wheel button to initiate a dialog with the speech-based interface. The complex task remained unaffected, whereas for the simple task, driving performance (MDev, SDLP, RT) and operation quality (subjective: Distr) got worse. The simple task being more affected was expected since the function operation takes place at the initiation of the lane change. Therefore, any visual-manual interference at the same time results in driving performance detriments. However, it is curious that there are no effects for the complex task. Compared to the simple task, the button could be pressed at any time and not just directly with the lane change initiation. Therefore, drivers could have timed the button press so that it would not occur during the lane change but during lane keeping only.

The results therefore indicate that hands-free speech-based HMI initiation is only relevant when driving maneuvers need to be performed at the same time. In uncritical driving situations, however, there is no difference between activation via a keyword or a steering wheel button.

Table 53: Speech-Based Interface - Hands-free

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-SB-3		Hands-free		Initial activation via key phrase		Push-to-talk button before each input on the steering wheel		“Hey Otto” activation vs. pressing button before each command	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-2	EoU	0	MDev	0	EoU	0		
SDLP	-1	Distr	-1	SDLP	0	Distr	0		
RT	-1	Miss	0	RT	0	N	0		
		Error	0			Error	0		
Sum	-4	Sum	-1	Sum	0	Sum	0	1.44	0.22
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-SB-4 *Shortcut* is a bonus item, i.e., it is not an essential design principle of the checklist. It refers to an option for shortening the operation of a complex function (see Table 54). In the validation, both a suboptimal and optimal approach were tested together. For the simple task, direct activation of the indicator was always possible (reference version). To test a suboptimal “non-shortcut” approach, participants now had to wake up the speech-based interface with a keyword. In contrast, this keyword always had to be used for the activation of the complex task (reference version). For the validation, the optimal approach was used by leaving out the keyword and allowing for directly voicing the phone number. The differences in the reference version (simple task: no keyword, complex task: keyword) stem from the decision tree (see section 5.2.3) and the maximum of allowed operation steps per respective level. The magnitude of effects is stronger for the simple task compared to the complex task. Here, negative effects on driving performance (MDev, RT) and operation performance (subjective: EoU, Distr; objective: Miss) can be seen. For the complex task, there were some positive effects on driving performance (RT) and operation quality (subjective: Distr, objective N). The opposite direction of effects between the tasks is fully expected: For the simple task, the non-shortcut led to more operation steps, thereby increasing operation demand. For the complex task, with the shortcut, the interaction got simpler and demand decreased. However, the magnitude of effects is larger for the simple task compared to the complex task. This is likely due to the simple task only having to be performed 18 times per trial and at the initiation of the lane change. The complex task was continuous and participants could incorporate pauses at critical driving moments like the lane change initiation. Furthermore, as the simple task had to be performed at a critical moment, the now preceding activation with the keyword extended the interaction process well into the lane change itself. Also, this appears to have been difficult as the number of misses increased.

The results show that shortcut operation is well suited for simple time-critical tasks, whereas benefits are not clearly visible for complex tasks. However, it must be discussed whether it is sensible to allow direct activation of critical functions in regard to operation errors: If the speech-based interface falsely detects a command while driving, this could lead to distracting erroneous function activations. Therefore, the activation via a button (see aspect M1-SB-3 *Hands-free*) and then using a direct command seems to be the more suitable option.

Table 54: Speech-Based Interface - Shortcut

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-SB-4		Shortcut		Own ideal command possible		Long predefined command, voice control gives hints		Without “Hey Otto” vs. with “Hey Otto” and more operation steps	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-2	EoU	-2	MDev	0	EoU	0		
SDLP	0	Distr	-2	SDLP	0	Distr	1		
RT	-2	Miss	-2	RT	1	N	2		
		Error	0			Error	0		
Sum	-4	Sum	-6	Sum	1	Sum	3	1.62	0.21
Negative effect on driving performance									
Note: Shortcut tested for complex task. Non-shortcut tested for simple task. Bonus item: Optimal implementation tested.									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-SB-5 *Risk of Operation Errors* refers to consequences of an erroneous function operation regarding driver distraction (see Table 55). It is connected to the previously discussed design aspect M2-SB-2 *Correction of Operation Errors*. The same simulated operation error was used (flashing screen with red and green colors and error sounds playback). Compared to M2-SB-2, however, participants could directly cancel the error by saying "Cancel!" or by pressing a button on the touchscreen. Compared to M2-SB-2, the effect on driving performance was less severe: Although the simple task remained yet again unaffected, only one measure for the complex task was affected (MDev) compared to all three with M2-SB-2. Operation performance was slightly less affected as well for both the simple task (subjective: Distr) and complex task (subjective: EoU, Distr; objective: Error). Compared to M2-SB-2, the results indicate that operation errors affect driving performance only to some extent, if the cancellation process is simple. As the cancellation was less cumbersome than with M2-SB-2, participants could reliably use the speech-based cancellation command and did not need to rely on the button press. However, button presses still occurred for the complex task as it required continuous operation. Furthermore, the interface was visually inspected to check the status of function operation, so that the red cancellation button likely appeared as the safer and quicker way to deactivate the malfunction.

As with M2-SB-2, the best solution regarding operation errors is to not let them occur in the first place. However, if they occur, their effect on driving performance is not as severe if they can be directly cancelled in a single step.

Table 55: Speech-Based Interface - Risk of Operation Errors

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-SB-5		Risk of Operation Errors		No safety risk		Strong distraction		Normal vs. flashscreen + error sound + cancellation via direct command or button	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	-2	EoU	-2		
SDLP	0	Distr	-1	SDLP	0	Distr	-2		
RT	0	Miss	0	RT	0	N	0		
		Error	-1			Error	-1		
Sum	0	Sum	-2	Sum	-2	Sum	-5	1.39	0.23
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-SB-6 *Additional Feedback* specifies the types of feedback the speech-based interface provides (see Table 56). For the validation, only speech-based feedback was provided without any chimes or visual feedback. Furthermore, the length of the speech-based feedback was increased to elongate the operation process. For the complex task, this had no effect on driving performance or on operation quality compared to the reference version. For the simple task, there was a slight effect on driving performance (MDev). However, every measure of operation quality was affected negatively (subjective: EoU, Distr; objective: Miss, Error). Regarding driver distraction, the length of the feedback has no substantial impact. However, operation quality suffers from elongated responses when the function operation is time-critical. Most often, this was because the interface was still responding to the first lane change when a second lane change appeared. Thereby, operating the simple task for this next lane change became impossible.

The results highlight that elongated feedback is not suitable for repeated quick function use. Therefore, feedback should ideally be kept brief and distinct.

Table 56: Speech-Based Interface - Additional Feedback

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-SB-6		Additional Feedback		Speech, chimes, visual if multi-step		Speech too long, visual too long		Additional feedback vs. no chimes, long filler phrase	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-1	EoU	-2	MDev		EoU			
SDLP	0	Distr	-2	SDLP		Distr			
RT	0	Miss	-2	RT		N			
		Error	-2			Error			
Sum	-1	Sum	-8	Sum	0	Sum	0	1.4	0.2
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-SB-7 *Diverse Commands* specifies that the speech-based interface must be able to recognize many expressions referring to the same target function (see Table 57). For validation, the speech-based interface's ability to understand commands was made stricter by requiring participants to always include the general function term (simple task "indicator,...", complex task "phone,...") before voicing the actual function command. For the complex task, this had no effect on driving performance or on operation quality. For the simple task, driving performance (MDev) and operation quality (subjective: EoU, Distr) were somewhat negatively affected. Perhaps violating this design principle by adding an additional preceding command could be easily learned by the participants so that the command did not really become more inflexible. Also, function operation was well learned as participants repeatedly used the function. Therefore, the absence and small magnitude of effects is expected.

In real life, functions are operated less frequently, and some functions are only ever rarely used. Therefore, we strongly advise that the interface can understand a multitude of commands addressing the same target function. By doing so, drivers do not have to think extensively about the right command and the likelihood of activating an unintended function is lowered.

Table 57: Speech-Based Interface - Diverse Commands

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-SB-7		Diverse Commands		Many common expressions recognized		One keyword must always be included		Various commands vs. “indicator” and “phone” must always be included	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	-1	EoU	-1	MDev	0	EoU	0		
SDLP	0	Distr	-1	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	0		
		Error	0			Error	0		
Sum	-1	Sum	-2	Sum	0	Sum	0	1.18	0.09
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design principle M1-SB-4 *Robustness of Activation* focuses on how reliably the speech-based interface can be activated (see Table 58). For the validation, activation errors were simulated by randomly activating the interface's listening mode. This had no effect on driving performance or on operation quality for the simple task. For the complex task, there were only marginal effects on driving performance (SDLP) and operation quality (objective: Error). As both functions were operated repeatedly, the random activation only ever rarely happened before the actual activation by the participant. Furthermore, with the interface doing the switch in a predefined time frame and by participants being familiar with the interface, they could anticipate this behavior. As the results demonstrate, this did not really interfere with driving or function operation.

However, in real life, drivers are not constantly engaging with a speech-based interface. So, a random activation of the interface would likely result in great distraction as drivers would now need to acquire an understanding of what malfunction just happened. As a result, they would have to think about possible solutions and visually search the vehicle's interface and take action to deactivate the speech-based interface. Therefore, we strongly recommend reliable activation, perhaps via a button (see aspect M1-SB-3 *Hands-free*) to ensure such random activation errors do not occur.

Table 58: Speech-Based Interface - Robustness of Activation

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-SB-4		Robustness of Activation		Key phrase and button not too sensitive		Too sensitive to noise and inputs		Reliable vs. randomized transition (5–10s) into listening mode	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	-1	Distr	0		
RT	0	Miss	0	RT	0	N	0		
		Error	0			Error	-1		
Sum	0	Sum	0	Sum	-1	Sum	-1	1.14	0.07
Negative effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

With the design aspect M1-SB-5 *Intelligibility / Quality of Voice Output*, the quality of the voice output of the speech-based interface is addressed (see Table 59). For the validation, a lower quality version of the text-to-speech voice was used and changed in the tablet's accessibility options. This had practically no effect on driving performance and operation quality for both tasks. Only some operation quality measures for the complex task were affected (subjective: EoU; objective: N). This result is within expectations since even though the voice was less well understandable it was still on an acceptable level. This is not surprising since the tablet would not ship with unacceptable voicing options. Perhaps effects would appear if voice quality was quite worse.

The results imply that function operation via the speech-based interface is not really affected by the quality of the voice output. Still, we advise using an easy-to-understand voice output.

Table 59: Speech-Based Interface - Intelligibility / Quality of Voice Output

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-SB-5		Intelligibility / Quality of Voice Output		Perfectly natural language		Rather artificial language		Voice “Viktor” (natural) vs. “Anna” (artificial)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	-1		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	-1		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	0	Sum	-2	1.07	0.04
No effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-SB-8 *In-process Abortion* specifies that a function operation with multiple steps can be aborted at any point (see Table 60). This design aspect was therefore only tested with the complex task. Here, the interface would not allow for cancellation in cases where participants started a call to the wrong number. However, as participants nearly never input the wrong number, it was expected that an alteration in the cancellation procedure would not have any effect.

Still, we argue that cancellation at any step is always important, especially regarding operation errors. If the operation cannot be aborted in a single step, this may result in long operation paths, inducing high levels of driver distraction.

Table 60: Speech-Based Interface - In-process Abortion

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-SB-8		In-process Abortion		At every point		At some points, only cumbersome		Cancellation possible (via speech or touch) vs. no cancellation	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	0		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	0	Sum	0	1	0
No effect on driving performance									
Note: Only complex task tested.									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-SB-6 *Feedback Dimensions* refers to the feedback channels used by the speech-based interface (see Table 61). Ideally, auditory (voice and chimes) and visual feedback are used for redundancy to ensure the driver can reliably monitor the system status. For the validation, the chimes and visual feedback were removed and only voice feedback was used. This had practically no effect on driving performance and function operation quality for both tasks. This makes for an interesting comparison with the aspect M2-SB-6 *Additional Feedback* where the feedback was limited in the same way but the length of the voice output was additionally increased. With increased length of the voice output, operation performance of the simple, time-critical task was negatively affected. Together with this result, there is no difference between richer and simpler feedback in regard to the feedback dimensions used.

Therefore, speech-based interfaces can use a variable amount of feedback channels. However, this likely depends on a per function basis as some functions benefit from visual feedback (showing a navigation route) while others do not (activating wipers, function activation itself can be seen).

Table 61: Speech-Based Interface - Feedback Dimensions

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-SB-6		Feedback Dimensions		Speech output with chimes and visual feedback		Speech output only, no chimes or visual feedback		Speech + chimes + visual vs. speech only	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	0		
SDLP	0	Distr	1	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	0		
		Error	0			Error	0		
Sum	0	Sum	1	Sum	0	Sum	0	1	-0.02
No effect on driving performance									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M2-SB-9 *Interaction Speed* describes that during a multiple-step function operation the interaction speed is defined by the user (see Table 62). Accordingly, the design aspect was tested with the complex task only. Here, the function would time out of the listening state within three seconds. For successful operation, participants therefore had to voice the number within three seconds. This was done to increase operation demand by incorporating time pressure. This had no effect on driving performance. Curiously, this had a positive impact on subjective (EoU, Distr) and objective operation quality (N). The increase in the number of inserted numbers was expected as participants hurried to enter them. The improved subjective operation quality may be due to the participants focusing more on the operation process and experiencing it more intensively, perhaps resulting in a feeling of success. From a driving performance perspective, the results would indicate that interaction speed is not a key component in reducing driver distraction. As mentioned previously, drivers were well acquainted with the operation process. In real life, function operation is much rarer.

Under these circumstances, real life function operation cannot be compared to the well learned operation process during the validation. As a result, a sudden reset would require the driver to understand the system status by looking at a screen. This would bind cognitive and visual resources, likely resulting in driver distraction. Therefore, we advise that the function waits for driver input instead of resetting or proceeding on its own.

Table 62: Speech-Based Interface - Interaction Speed

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M2-SB-9		Interaction Speed		Waits indefinitely		Waits briefly, gives instruction, cancels		Waits indefinitely vs. timeout after 3s in listening mode	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	0	EoU	1		
SDLP	0	Distr	0	SDLP	0	Distr	2		
RT	0	Miss	0	RT	0	N	2		
		Error	0			Error	0		
Sum	0	Sum	0	Sum	0	Sum	5	1	-0.09
No effect on driving performance									
Note: Only complex task tested.									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

The design aspect M1-SB-7 *Volume* refers to the output volume of the speech-based interface. Ideally, the output can be well heard by the driver under all circumstances (see Table 63). For validation, the volume was reduced to 10 % of the tablet's maximum loudness compared to the 50 % used in the reference version. This had basically no effect on driving performance and operation quality across both tasks. This result is within expectation as the validation took part in a quiet science lab with some noise from the driving simulation. However, this is not representative of the loudness and noises that occur during real life driving.

So, even though there was no effect during validation, we strongly advise that speech-based interfaces provide auditory feedback at appropriate loudness levels for all circumstances.

Table 63: Speech-Based Interface - Volume

Description									
Item		Description		CL Optimal		CL Suboptimal		Operationalization	
M1-SB-7		Volume		Default volume appropriate, adjustable, automatic		Default volume too loud/quiet, adjustable, no automatic		Appropriately loud (50% volume) vs. very quiet (10% vulme)	
Effects Compared to Reference Version									
Simple				Complex				Scores	
DP		OQ		DP		OQ		Penalty	Net
MDev	0	EoU	0	MDev	1	EoU	0		
SDLP	0	Distr	0	SDLP	0	Distr	0		
RT	0	Miss	0	RT	0	N	0		
		Error	0			Error	-1		
Sum	0	Sum	0	Sum	1	Sum	-1	1.04	-0.03
Possible positive effect on driving performance (likely no effect)									

Note. CL = checklist, DP = driving performance, OQ = operation quality. MDev = mean deviation from ideal trajectory, SDLP = standard deviation from ideal trajectory, RT = reaction time. EoU = ease of use, Distr = subjective distraction, Miss = number of misses, N = number of inserted numbers, Error = number of errors.

7.4 Summary

The validation was primarily aimed at identifying the most important design aspects for each HMI type in regard to driver distraction. By doing so, the HMI checklist will treat more pressing interface design issues in a stricter manner. This is not only valuable for the evaluation of an interface but also at the design stage for respecting fundamental design principles that ensure safe function operation while driving. Besides the evaluation of the design aspects within each HMI modality, the validation also allowed for a comparison between modalities with respect to task complexity.

The validation incorporated a dual task scenario where participants drove in the LCT and operated both a simple and complex task via generic interfaces. A generic interface was designed for each of the three HMI modalities (physical controls, touch-based, speech-based). In the reference version, each design aspect was implemented in the optimal way. For the comparison of the HMI modalities with respect to task complexity, this reference version was used. For the validation of the design aspects, one design aspect at a time was implemented in a suboptimal manner while the remaining aspects stayed optimal. The effect of this suboptimal implementation was then compared to the reference version (optimal implementation) regarding driver distraction and operation quality.

The global analysis compared the reference versions of each HMI type (physical controls, touch-based, speech-based) across both tasks (simple, complex) regarding driving performance. For lateral deviation (MDev, SDLP), there was no difference between the three HMI modalities for the simple task. However, as shown in the validation of the speech-based design aspects *Fallback* and *Deactivation & Alternative Operation Option*, physical controls and the speech-based interface performed on the same level and resulted in less driver distraction compared to touch-based operation. For the complex task, the speech-based interface resulted in the least driver distraction, followed by the touch-based interface, with physical controls being the most distracting operation option. Regarding temporal measures (RT), the speech-based interface resulted in the quickest steering responses across both tasks. For the simple task, physical controls were the second-best option whereas the touch-based option is the second-best option for the complex task. Overall, the speech-based interface resulted in the least driver distraction across all measures (MDev, SDLP, RT).

Table 64 provides an overview of the design aspects that were tested for physical controls. It summarizes the design aspects and their net score into overarching design concepts. The concepts should be seen as a suggestion since they are derived from our study's design and results. As often mentioned, the operationalization of the design aspects is always specific and will differ between studies and applications, i.e., vehicle cockpits. Besides the overarching concepts, we provide design recommendations drawn from the validation results. The design aspects are sorted in the same order as in Table 10. Therefore, the aspects whose violation had the greatest negative effect on driving performance are listed first. As described in their introduction, physical controls come in many shapes and forms and can be placed in various positions in the vehicle cockpit. The validation revealed the importance of reducing the motor effort required for operation. To achieve this, the control elements should be placed close to the driver, and the interaction must be made easy by incorporating, e.g., shallow travel distances. Less but still important is the reduction in operation steps as well as physical properties like labeling or antireflective properties. For the latter, it should be noted that during validation participants were very familiar with operation so that altering specific aspects (like labeling) during one trial did not have a substantial effect. Therefore, these design principles should not be underestimated regarding their effect on driver distraction.

Table 64: Physical Controls - Design Aspect Summary

Overarching Concept	Design Aspect	Design Recommendation
Mechanical Effort Ø Net Score: 0.28	Susceptibility to Operation Errors	→ Position close to the driver → Make interface easy to grab
	Blind Operation	→ Control element should require low motor effort (e.g., shallow travel distance)

Overarching Concept	Design Aspect	Design Recommendation
Proximity Ø Net Score: 0.21	Viewing Distance to the Right	→ Place control elements close to driver → Ideally steering wheel controls
	Expectation Conformity	→ Adhere to established control element placement and operation logic
	Reachability	→ Place control elements close to the driver → Place elements close to the driver's line of sight
	Viewing Distance Downwards	→ Center stack placement is acceptable
Interaction Effort Ø Net Score: 0.09	Correction of Operation Errors	→ Use as little operation steps as possible → Quick repeated inputs are acceptable
	Risk of Operation Errors	→ Reduce the likelihood of errors and additional operation steps
Physical Properties Ø Net Score: 0.05	Glare	→ Make control elements matte
	Comprehensibility	→ Labeling should be easy to read
	Legibility	→ Labeling should be easy to understand

Table 65 provides an overview of the design aspects that were tested for the touch-based interface. It summarizes the design aspects and their net score into overarching design concepts. Also, it shows the design recommendations drawn from the validation results. The design aspects are sorted in the same order as in Table 23. Therefore, the aspects whose violation had the greatest negative effect on driving performance are listed first. As the touch-based interface is visually rich and does not provide tactile feedback compared to physical controls, the validation revealed the importance of reducing visual-manual effort. To reduce the interaction effort, simple gestures (e.g., tapping instead of sliding) and appropriate button sizes should be used. Also, for the interaction process, extra inputs should be avoided when, for example, the interface is unresponsive or errors occur that need to be handled by the driver. Compared to physical controls, the proximity to the driver and physical properties are less relevant. Finally, the design of the visual interface did not really affect driver distraction. As noted for the validation of the according design principles, this is likely due to the participants being familiar with the interface so that they learned the function operation well enough to easily overcome an altered interface design. Therefore, these design principles should still be considered for less frequent real life function operation.

Table 65: Touch-based Interface - Design Aspect Summary

Overarching Concept	Design Aspect	Design Recommendation
Interaction Effort Ø Net Score: 0.43	Susceptibility to Operation Errors	→ Adhere to established operation logic → Align touch buttons → Make touch buttons of appropriate size
	Expectation Conformity	→ Use tapping as input instead of sliding or more complex touch gestures
Interaction Process Ø Net Score: 0.19	Input Precision	→ Reliably register touch inputs
	Risk of Operation Errors	→ Reduce likelihood of operation errors and additional operation steps → Keep button placements similar across menus
	Correction of Operation Errors	→ Reverting operation errors via a single or multiple quick taps is acceptable
	Interaction Speed	→ Interaction speed must be dictated by the driver

Overarching Concept	Design Aspect	Design Recommendation
	In-process Abortion	→ Complex function operation must be abortable at any point
	Shortcut	→ Enable skipping operation steps by including shortcut options
Proximity Ø Net Score: 0.10	Reachability	→ Place the screen close to the driver
	Driving-related Information Always Visible	→ Driving-related information must always be visible
	Viewing Distance Downwards	→ Place the touchscreen close to the driver's visual field (vertically)
	Viewing Distance to the Right	→ Place the touchscreen close to the driver's visual field (horizontally)
	View on the Road Not Blocked	→ Do not place the screen too much thin the driver's line of sight
Physical Properties Ø Net Score: 0.10	Reflection	→ Make the screen matte and antireflective
	Tilt	→ Tilt the screen towards the driver
Visual Interface Ø Net Score: 0.04	Legibility	→ Use a high contrast screen and interface → Make elements of appropriate size
	Brightness and Contrast	→ Use high contrast screen with automatic brightness regulation
	Attentional Guidance	→ Make the visual interface distinct (e.g., buttons are clearly identifiable)
	Minimalism	→ Avoid visual clutter → Display only appropriate an appropriate amount of content

Finally, Table 66 provides an overview of the design aspects that were tested for the speech-based interface. It summarizes the design aspects and their net score into overarching design concepts. Also, it shows the design recommendations drawn from the validation results. The design aspects are sorted in the same order as in Table 47. Therefore, the aspects whose violation had the greatest negative effect on driving performance are listed first. During validation, speech-based operation was compared to other interface modalities (physical controls, touch-based) that can be used to access the same functions. This was done since it is unlikely that primary function access can always be provided via the speech-based interface. This is due to the underlying technical complexity of the interface (e.g., speech-recognition software, internet access) and its susceptibility to external factors that make operation impossible (e.g., driver has no voice, noise). In general, the correct fallback modality must be used in regard to task complexity. For simple tasks, physical controls should be used that allow for low-distraction operation on the same level as the speech-based interface. For complex tasks, touch-based interfaces should be used that result in greater distraction compared to the speech-based interface. The speech-based interface should reliably and correctly identify commands to not introduce errors. For command recognition, short commands are best and the recognition of many common terms. The speech component of the auditory feedback should be kept short, however, the use of further kinds of feedback (e.g. chimes, visual) does not appear to be essential regarding distraction. Finally, the initiation of the speech-based interface both via keyword and button is acceptable. The timing and cancellation of a function operation process had no effect during validation, however, this was due to the fast timing and short interactions in our generic speech-based interfaces. Our procedure therefore did not sufficiently test longer, more complex interactions with the speech-based interface. As a result, we recommend still considering these principles when designing a speech-based interface.

Table 66: Speech-based Interface - Design Aspect Summary

Overarching Concept	Design Aspect	Design Recommendation
Fallback & Correct Modality Switch Options Ø Net Score: 0.39	Fallback	→ Ideally, make the interaction process reliable so that no fallback is needed → Use correct fallback modalities in accordance to task complexity
	Correction of Operation Errors	→ Reduce likelihood of operation errors and additional operation steps → Make the correction as easy as possible, ideally without falling back to other modalities
	Deactivation Option & Alternative Operation Option	→ Use correct alternative modalities in accordance to task complexity
	Hands-free	→ Both keyword and button activation of speech-based interface are acceptable → To reduce likelihood of errors, button activation method recommended
	Risk of Operation Errors	→ Reduce likelihood of operation errors and additional operation steps
Speech Recognition Ø Net Score: 0.31	Quality of the Speech Recognition	→ Correctly identify the target function within the driver's command
	Susceptibility to Operation Errors	
Command Length Ø Net Score: 0.15	Shortcut	→ Use shortcuts to shorten the driver-interface dialogue → Omit the use of a keyword for activation of the speech-based interface
	Diverse Commands	→ Make the interface able to understand a wide range of commands addressing the same function
Auditory Feedback Ø Net Score: 0.05	Additional Feedback	→ Keep the speech-based interface's dialogue short
	Feedback Dimensions	→ Both pure (auditory feedback only) and mixed (auditory & visual feedback) interfaces are acceptable
	Intelligibility / Quality of Voice Output	→ Use a natural, easy-to-understand voice for the interface
	Volume	→ Use appropriate loudness for every circumstance
Interaction Process Ø Net Score: -0.01	Robustness of Activation	→ Activation of the speech-based interface must be reliable → Use of buttons instead of keyword recommended
	In-process Abortion	→ The speech-based operation of a function must be cancelable at any point
	Interaction Speed	→ Interaction speed must be dictated by the driver

8 Comparison to Euro NCAP's General Vehicle Controls

With the 2026 revision of their protocols, Euro NCAP reworked its rating system for vehicle safety. The new methodology is centered around four stages of safety: safe driving, crash avoidance, crash protection, and post-crash safety. As part of the safe driving assessment, the driver engagement protocol is aimed at keeping the driver actively involved in driving. To this end, driver monitoring systems and the use of low-distraction HMIs used for function operation are evaluated.

8.1 General Vehicle Controls: Structure

The General Vehicle Controls (GVC) assessment, part of the driver engagement protocol (Euro NCAP, 2025), is an effort to combat the increased reliance on touch-based function operation. To get the best rating, essential functions (e.g., indicator, wipers, horn) must be accessible via physical controls. The basis for the GVC protocol was the HMI checklist and the respective research by Horn et al. (2023). The checklist has been adapted for the aim of testing final HMI designs in vehicles.

The GVC consists of two parts that are concerned with driving-related und non-driving-related functions. Within each part, there are subcategories that contain functions for which select actions and tasks must be performed (e.g., activate the direction indicator to the left and right). For each task, one or two requirements and the acceptable function implementation must be met to score points.

The acceptable implementation and requirements for each function are derived from the use context of the function. This is done via the previous version of the decision tree. Depending on the use context, four types of function implementations and requirements can be derived:

- Direct physical input
- Direct touch input
- Menu-based touch input
- Direct voice input

The requirements for each function implementation refer to the definition provided by the protocol, the description in the function implementation section, or additional information provided in the appendix of the according technical bulletin. For *direct physical input* (1), the definition requires the interface to provide mechanical movement and haptic feedback when operated. Also, the driver should be able to identify it by its physical properties alone and the interface should ideally require no off-road glances. Other solutions without mechanical movement are also allowed, as long as all other requirements are met. Besides the definition, the requirement *control identification – expected locations* demands that physical controls are placed in familiar places in the vehicle cockpit. Regarding *direct touch input* (2), the definition describes it as an interface accepting touch inputs and being always accessible. The first requirement *control identification – tactility* demands that the interactive area can be identified by its tactile properties and is bound from controls of other functions. The second requirement *control identification – interaction area* specifies that the interaction area should be sufficiently large and spaced from other interaction areas. *Menu-based touch input* (3) is only specified by a definition, which states that the interface is comprised of a menu whose operation requires a maximum of two steps counted from the homescreen. Finally, *direct voice input* (4) is defined as an interface that allows for operating devices, applications, and services via voice commands. An additional requirement is that the speech-based interface should be activated via a button and that an alternative control to the speech-based operation is implemented.

Besides the interface-specific requirements described previously, the requirement *essential driving information in direct driver's line of sight* refers to the visual feedback provided for monitoring the function's status. The requirement is assigned to three types of functions: light functions, vehicle assistance activation status, and vehicle assistance speed control. It mandates that the status of these functions is always visible in the driver's line of sight.

8.2 General Vehicle Controls and HMI Checklist: Assessment Process and Scope

Just as Euro NCAP's GVC, our HMI checklist is also aimed at improving driving safety by ensuring functions are implemented in a way that ensures low-distraction function operation. The GVC is a welcome development and already covers important functions and design principles, hereby ensuring that unacceptable implementations should not be considered by manufacturers. However, we think our checklist can be used to enhance the GVC, especially by providing more details regarding the specific implementation within an HMI modality. In the following sections, we want to draw attention to the similarities and differences between both instruments and where the GVC could potentially be expanded.

The general assessment process is rather similar between both instruments, however, there are differences in the assessment scope, process and overall length. This is largely due to their different roles in the HMI development lifecycle: the checklist is designed as a tool to inform and guide design decisions during the development process, whereas the GVC serves as an instrument to evaluate the final design outcome. The GVC uses predefined functions and tasks that need to be performed as part of the assessment process. If the requirements are met, points can be scored. On the other hand, the checklist is designed in an open way, i.e., the functions to be assessed are selected by the checklist users themselves. Therefore, the use context of the function and the acceptable implementation must always be derived first and cannot be provided beforehand like in the GVC. Based on the acceptable implementation of the function in the checklist, the according items need to be checked. Here, only general guidance is given on how to check the design principle as due to the open design, the principles cannot be function-specific. Unlike the GVC, the checklist's scoring on the item level is more granular: Instead of an acceptance criterium (met or not met), more options are available to evaluate the implementation quality of the function. For the weighting of scores, the GVC assigns higher scores (more weight) to functions that are closer to the driving task. No points are awarded if a single function of a subcategory fails the assessment. In contrast, the checklist weighs the scores not per task but per design aspect. Here, a weighting factor is used, acting as a multiplier, that was derived from the validation process. Overall, the checklist's assessment process takes longer and requires more expertise, we argue, compared to the GVC. Still, for the design stage, our instrument provides valuable insights for manufacturers, now that it is more in-depth and based on an empirical basis compared to the previous version by Horn et al. (2023). To transfer the results of the HMI checklist's validation to the GVC, we suggest partly expanding or revising the function implementation requirements of the protocol while keeping its original structure. The fixed structure is well suited for economical real-world lab testing compared to the checklist's approach. The GVC aims to penalize and to eliminate designs that are most detrimental to driver attention and driving performance.

As described previously, the GVC considers four acceptable implementation options that are specified by their definitions and requirements: direct physical input, direct touch input, menu-based touch input, and direct voice input. Each implementation is only outlined, and the requirements of this outline must be met. There is no further differentiation between "acceptable" and "unacceptable" in cases where not all requirements are met. Our checklist could enhance the GVC in two ways: First, it could lead to the inclusion of new requirements that are relevant for the acceptable implementation of a function in an HMI modality/type. These requirements would be based on the design aspects that we validated. Our idea would be to not include all design aspects, but rather those that had the greatest negative impact on driver distraction when implemented in a suboptimal way. Second, the evaluation could be made more nuanced by diverting from the acceptance-based approach, where the requirements are either met or not met. This could be achieved in several ways. For example, by making more important requirements mandatory while others could be optional. Another option could be to incorporate a weighting logic and numeric scoring of each requirement. Then, this score could be compared against a sum threshold. If the goal is to keep the general GVC structure, we would advise keeping the acceptance-based rating system and including more requirements as discussed in the first approach. To this end, we describe options for enhancing the requirements of each implementation (HMI type) in the next section.

8.3 GVC Function Implementation: Enhancement Options

The enhancement of the GVC's criteria will be done by linking the validation results of the checklist's design aspects with existing GVC requirements or by suggesting new requirements. As a foundation, the results summary tables (see section 7.4) will be used since the design aspects are already sorted according to their relevance regarding driver distraction. The most important design aspects are listed first in every table. By doing so, missing GVC requirements can be easily identified. Also, existing GVC requirements can be evaluated whether they are relevant or not regarding driver distraction.

Table 67 contains the design aspects that were tested during the checklist's validation for physical controls. The GVC only considers physical controls as a direct input method, therefore multiple-step operation via physical controls is not currently part of the GVC. Direct physical input can be linked the best with the checklist's validation results since it is specified to a greater degree compared to touch input (direct or menu based) or direct voice input. As a result, all suggestions are enhancements of the definition and the *control identification – expected locations* requirement. Although no completely new requirements are suggested, it can be discussed whether physical inputs should be considered regarding multiple-step operation. However, given the poor performance of physically controlled tasks that require multiple steps, we suggest not doing so.

Table 67: Suggestions for Euro NCAP's General Vehicle Controls (GVC) Assessment: Physical Controls

HMI Checklist		Suggestion for GVC Requirements
Design Aspect	Design Recommendation	Direct Physical Input
High relevance (net score > .10)		
Susceptibility to Operation Errors	→ Position close to the driver	→ Enhance control identification – <i>expected locations</i> by proximity to driver
	→ Make interface easy to grab	→ Enhance definition , add tangibility, mechanical integrity, sturdiness
Blind Operation	→ Control element should require low motor effort (e.g., shallow travel distance)	→ Enhance definition , further define mechanical movement aspect
Viewing Distance to the Right	→ Place control elements close to driver → Ideally steering wheel controls	→ Enhance expected locations by proximity to driver
Expectation Conformity	→ Adhere to established control element placement and operation logic	→ Enhance control identification – <i>expected locations</i> to include operation logic
Reachability	→ Place control elements close to the driver	→ Enhance control identification – <i>expected locations</i> by proximity to driver
	→ Place elements close to the driver's line of sight	
Viewing Distance Downwards	→ Center stack placement is acceptable	→ No action needed, already included with <i>expected locations</i>
Low relevance (net score < .10)		
Correction of Operation Errors	→ Use as little operation steps as possible	→ No action needed, already included with decision tree (use context of the function)
	→ Quick repeated inputs are acceptable	→ No action needed, direct physical input does not consider multiple-step operation

HMI Checklist		Suggestion for GVC Requirements
Design Aspect	Design Recommendation	Direct Physical Input
Risk of Operation Errors	→ Reduce the likelihood of errors and additional operation steps	→ No action needed, direct physical input does not consider multiple-step operation
Glare	→ Make control elements matte	→ Optional enhancement of <i>definition</i> , aspects were not integral in the validation
Comprehensibility	→ Labeling should be easy to read	
Legibility	→ Labeling should be easy to understand	

The design aspects that were tested for the touch-based interface during the checklist's validation are shown in Table 68. In contrast to the other implementations (HMI types), physical controls and voice input, the GVC considers touch-input not only directly but also menu-based. Therefore, single-step and multiple-step operation are covered. The suggestions are a mix of enhancements of the definitions and *control identification – interaction area* requirements as well as suggestions for new requirements that are currently not included in the GVC. The operation process could be further specified with an *expected principles* and *interaction principles* requirement. With a *multiple-step operation* requirement, longer operation paths in touch-based menus could be covered. Finally, an *interface design* requirement could push the touchscreen's interface to require less visual demand. As the list of design aspects covered is rather long, we suggest focusing on the suggested design requirements that are listed first in the table below.

Table 68: Suggestions for Euro NCAP's General Vehicle Controls (GVC) Assessment: Touch-Based HMIs

HMI Checklist		Suggestion for GVC Requirements
Design Aspect	Design Recommendation	Direct Touch Input & Menu Based
High relevance (net score > .10)		
Susceptibility to Operation Errors	→ Adhere to established operation logic	→ New <i>expected principles</i> requirement that demands logical operation paths and function placement in menus
	→ Align touch buttons → Make touch buttons of appropriate size	→ Enhance <i>control identification – interaction area size</i> (direct touch) to include button alignment → New <i>interface design</i> requirement that specifies button size and placement
Expectation Conformity	→ Use tapping as input instead of sliding or more complex touch gestures	→ New <i>interaction</i> requirement that demands simple input gestures
Input Precision	→ Reliably register touch inputs	→ New <i>interaction</i> requirement that demands reliable detection of touch input
Risk of Operation Errors	→ Reduce likelihood of operation errors and additional operation steps	→ New <i>multiple-step operation</i> requirement that demands a reduction of errors
	→ Keep button placements similar across menus	→ New <i>multiple-step operation</i> requirement that demands similar button placements across menus
Correction of Operation Errors	→ Reverting operation errors via a single or multiple quick taps is acceptable	→ New <i>multiple-step operation</i> requirement that demands quick error correction and abortion

HMI Checklist		Suggestion for GVC Requirements
Design Aspect	Design Recommendation	Direct Touch Input & Menu Based
Interaction Speed	→ Interaction speed must be dictated by the driver	→ New multiple-step operation requirement that defines that operation speed is managed by driver
In-process Abortion	→ Complex function operation must be abortable at any point	→ New multiple-step operation requirement that demands quick error correction and abortion
Shortcut	→ Enable skipping operation steps by including shortcut options	→ New multiple-step operation requirement that demands skipping options for longer operations
Reachability	→ Place the screen close to the driver	→ Enhance definition (both direct touch and menu-based) to include touchscreen placement
Driving-related Information Always Visible	→ Driving-related information must always be visible	→ Already included with requirement <i>driving-related information always visible</i>
Viewing Distance Downwards	→ Place the touchscreen close to the driver's visual field (vertically)	→ Enhance definition (both direct touch and menu-based) to include touchscreen placement
Viewing Distance to the Right	→ Place the touchscreen close to the driver's visual field (horizontally)	→ Enhance definition (both direct touch and menu-based) to include touchscreen placement
View on the Road Not Blocked	→ Do not place the screen too much thin the driver's line of sight	→ Enhance definition (both direct touch and menu-based) to include touchscreen placement
Reflection	→ Make the screen matte and antireflective	→ Enhance definition (both direct touch and menu-based) to include touchscreen's antireflective properties
Tilt	→ Tilt the screen towards the driver	→ Enhance definition (both direct touch and menu-based) to include touchscreen placement
Low relevance (net score < .10)		
Legibility	→ Use a high contrast screen and interface	→ New interface design requirement that demands appropriate contrast
	→ Make elements of appropriate size	→ New interface design requirement demanding appropriate sizing
Brightness and Contrast	→ Use high contrast screen with automatic brightness regulation	→ Enhance definition to require touchscreen to be of appropriate brightness
Attentional Guidance	→ Make the visual interface distinct (e.g., buttons are clearly identifiable)	→ New interface design requirement specifying distinct button design
Minimalism	→ Avoid visual clutter	→ New interface design requirement demanding a tidy design
	→ Display only appropriate an appropriate amount of content	→ New interface design requirement demanding a tidy design

Finally, Table 69 contains the design aspects that were tested for the speech-based interface during the checklist's validation. As with direct physical input, only single-step operation is considered by the GVC but not an operation spanning across multiple steps. As direct voice input is only defined by a definition, the suggestions are largely proposals for new requirements that define the basic properties of the speech-based interface. The GVC already covers the button-based activation of the speech-based interface and the use of fallback options. However, based on the HMI checklist's validation, new requirements could be included that specify the *voice processing* and *output* as well as *dialogue management*. The latter could be relevant if functions were part of the assessment where dialogues with multiple turns between driver and interface take place. As with the touch-based suggestions, we suggest focusing on the suggestions listed first. Also, some requirements are marked as optional since they were not essential regarding driver distraction during validation.

Table 69: Suggestions for Euro NCAP's General Vehicle Controls (GVC) Assessment: Speech-Based HMIs

HMI Checklist		Suggestion for GVC Requirements
Design Aspect	Design Recommendation	Direct Voice Input
High relevance (net score > .10)		
Fallback	→ Ideally, make the interaction process reliable so that no fallback is needed	→ Discuss whether functions should be included that are primarily accessible via voice input instead of physical controls
	→ Use correct fallback modalities in accordance to task complexity	→ Enhance <i>alternative control</i> requirement by relying on correct HMI modality (for simple task: physical controls)
Correction of Operation Errors	→ Reduce likelihood of operation errors and additional operation steps	→ New <i>voice processing</i> requirement that demands reliable command detection
	→ Make the correction as easy as possible, ideally without falling back to other modalities	→ New <i>dialogue management</i> requirement that demands interruptibility of dialogue-based operation
Deactivation Option & Alternative Operation Option	→ Use correct alternative modalities in accordance to task complexity	→ Enhance <i>alternative control</i> requirement by relying on correct HMI modality (for simple task: physical controls)
Hands-free	→ Both keyword and button activation of speech-based interface are acceptable → To reduce likelihood of errors, button activation method recommended	→ Already included via <i>dedicated button</i> requirement
Risk of Operation Errors	→ Reduce likelihood of operation errors and additional operation steps	→ New <i>voice processing</i> requirement that demands reliable command detection
Quality of the Speech Recognition	→ Correctly identify the target function within the driver's command	→ New <i>voice processing</i> requirement that demands reliable command detection
Susceptibility to Operation Errors		
Shortcut	→ Use shortcuts to shorten the driver-interface dialogue → Omit the use of a keyword for activation of the speech-based interface	→ New <i>dialogue management</i> requirement that demands flexibility regarding the directness of the addressed function

HMI Checklist		Suggestion for GVC Requirements
Design Aspect	Design Recommendation	Direct Voice Input
Diverse Commands	→ Make the interface able to understand a wide range of commands addressing the same function	→ New <i>dialogue input</i> requirement that demands flexible command detection
Low relevance (net score < .10)		
Additional Feedback	→ Keep the speech-based interface's dialogue short	→ New <i>dialogue output</i> requirement that demands short voice output
Feedback Dimensions	→ Both pure (auditory feedback only) and mixed (auditory & visual feedback) interfaces are acceptable	→ Optional: New <i>dialogue output</i> requirement that specifies the use of mixed modality feedback
Intelligibility / Quality of Voice Output	→ Use a natural, easy-to-understand voice for the interface	→ Optional: New <i>dialogue output</i> requirement that demands natural voice output
Volume	→ Use appropriate loudness for every circumstance	→ New <i>dialogue output</i> requirement that demands suitable loudness level
Robustness of Activation	→ Activation of the speech-based interface must be reliable → Use of buttons instead of keyword recommended	→ Already included via <i>dedicated button</i> requirement
In-process Abortion	→ The speech-based operation of a function must be cancelable at any point	→ New <i>dialogue management</i> requirement that demands interruptibility of dialogue-based operation
Interaction Speed	→ Interaction speed must be dictated by the driver	→ New <i>dialogue management</i> requirement that demands that the interaction is paced by the driver

9 Conclusion

The main motivation of this project was the further development of the HMI checklist to include the implementation quality of in-vehicle functions regarding driver distraction and safe function operation. Modern vehicles are equipped with a wide range of functions, to which any HMI type (physical controls, touch-based, speech-based) could be used to provide access to. In recent times, manufacturers increasingly relied on touch-based interfaces to control many of the vehicle's functions. As touchscreens have a visually rich interface and require precise motor actions for input, they lead to high levels of visual-manual distraction compared to physical controls. Beyond visual-manual interfaces, some vehicles are equipped with speech-based interfaces in the hope of reducing driver distraction to a minimum. The influx of functions and interfaces as well as the increased reliance on the touchscreen has resulted in a landscape where function operation might not be in an optimal state. The development of the HMI checklist is an effort to push function operation towards a safe, low-distraction state.

The vehicle function overview (see chapter 3) has revealed that vehicles are equipped with at least 150 functions accessible to the driver. Usually, the more driving-related a function is, the more likely it is operated in a single step via physical controls. However, most of the functions available are not needed for the essential driving task, often require more than one step, and are accessible via touch- or speech-based interfaces. Although the assignment of a function to an interface type, i.e., physical controls, touch-, and speech-based interfaces, often seems suitable, the specific design of each interface type differs quite a bit between manufacturers. So, beyond mismatched function-interface type combinations, the research gap lies in specific interface design guidelines for each respective interface type that ensure safe function operation.

9.1 HMI Checklist: Towards Implementation Quality

A first step in narrowing down low-distraction interface design is the predecessor of the HMI checklist that was developed by Horn et al. (2023). Based on this, the Euro NCAP General Vehicle Controls (Euro NCAP, 2025) was developed that is aimed at putting essential driving-related functions on dedicated, physical controls. However, both instruments lack details regarding specific implementation principles. Horn et al.'s (2023) instrument was largely derived from the NHTSA visual-manual distraction guideline. As a result, only visual-manual interfaces are considered, and speech-based interfaces are missing. Also, the items do not suitably differentiate between both interface types and provide sufficient information on interface design dimensions. Although Euro NCAP's General Vehicle Controls also includes speech-based operation, the instrument does not include criteria for the implementation quality for each interface type. Another concern with both instruments is that they were not empirically validated yet. The current version of the HMI checklist was developed to overcome both concerns. First, by collecting more aspects on good interface design for every HMI type. Second, by using an empirical validation to ensure the aspects do indeed affect driver distraction and to identify the most relevant principles.

To this end, we started by carrying over the base structure of Horn et al.'s (2023) instrument. As a result, our checklist also employs two modules. The first module is concerned with HMI features that are shared between functions (e.g., the hardware of the touchscreen). The second module focuses on function-specific design aspects. This was done to ensure the assessment process stays efficient. Next, all three HMI modalities, physical controls, touch-based, and speech-based, are included in both modules and the decision tree. The latter was also revised to rank the modalities and acceptable operation steps regarding the use context of a function. The core of our checklist are the design aspects that are checked in both modules. These were derived by summing up design guidelines, standards, and research studies. These resources differ in their level of detail and are largely not function specific. This is a major strength of our instrument that tries to sum up the information and intends to link it to task complexity.

For each design aspect, an item was derived that provides a description of the principle, suggests a test procedure, and allows for a precise rating on a five-point rating scale. Furthermore, the items are weighted according to their distraction potential derived from the validation process. In summary, the checklist pushes the function implementation idea beyond simply matching the function task to a suitable interface type and then checking whether

select, unspecific design aspects are met or not. Now, the actual implementation quality is extensively checked across a range of validated design aspects via a quantitative rating scheme.

9.2 Empirical Validation: Ranking of the Design Aspects

With the empirical validation, the relative effects of the design aspects on driver distraction and operation performance were determined. As a byproduct, this also allowed for an analysis of how task complexity and HMI type interact with each other. The latter follows the results of studies where tasks of varying complexity were operated via different interface types (see section 4.4). Accordingly, speech-based interaction resulted in the least distraction. For simple tasks, physical controls performed essentially just as well as the speech-based interface and better than touch-based operation. For the complex task, speech-based operation was clearly the best, followed by touch-based operation and then physical controls. As the results follow established results, this indicates that the basic validation procedure sufficiently depicts in-vehicle function operation while driving. The validation itself used the Lane Change Task (LCT; International Organization for Standardization, 2010) for the driving simulation and generic interfaces based on self-programmed iOS applications. The LCT is very sensitive for capturing driver distraction in a continuous manner and has been implemented as an ISO standard. Furthermore, the generic interfaces allowed us to precisely control the interface's behavior and the implementation of the design principles. Taken together, the validation approach itself was well suited for measuring the effects of the function operation design aspects across the three HMI modalities.

The evaluation of the individual design aspects per HMI type is the second major strength of the HMI checklist beyond summing up the low-distraction interface design literature. The design aspects were tested for each of the three HMI modalities, namely physical controls, the touch-based, and speech-based interface. According to Wicken's Multiple Resource Theory (2002) the first two are visual-manual interfaces, the speech-based interface relies on the auditory channel. The design aspect's validation clearly demonstrated how the principles affect the resource usage per HMI type. For physical controls, the manual resource has had a greater impact compared to the visual aspect, although only the complex task really relied on it during validation. The most relevant design aspects refer to the manual effort required, either by the mechanical integrity of the control element (e.g., travel distance, stable mounting) or by the proximity to the driver, i.e., placement in the cockpit. Physical properties (e.g., glare, labeling) had nearly no impact.

This shifts when looking at the touch-based interface's validation results: Besides the manual effort, visual effort clearly affects driver distraction during function operation. When buttons are small or touch inputs are not registered, the visual interface needs to be looked at, and the target of the next tap must be visually identified. Therefore, interface design and operation paths are the most important design characteristics.

Finally, the speech-based interface is the newest interface type compared to both visual-manual interface types. As a result, there is nearly no comprehensive literature on spoken human-machine interaction, especially when driving. Our validation contributes to filling this gap: As speech-based interaction is often used as an "add-on" on top of the other interfaces, mixed modality interactions are expected. Our results clearly demonstrate that this can lead to high levels of driver distraction. This raises questions concerning the fundamental implementation of speech-based accessible functions: Following the results, function access should ideally remain pure, i.e., only via the speech-based interface. Looking at the use context of the functions (see decision tree in section 5.2.3) affected, this largely applies to multiple-step functions. Accordingly, access to complex functions should only be provided via the speech-based interface and not the touchscreen. However, this might lead to problems when the speech-based interface is not available and might conflict with driver's expected operation principles of functions (e.g., via the touch screen). Beyond the mixed modality discussion, speech-based interfaces must reliably detect commands, otherwise this leads to errors, resulting in high levels of driver distraction. This is in line with results on unexpected speech-based HMI behavior (Reichel et al., 2014). Interestingly, the quality of the auditory feedback was not as important.

The validation of the design principles for all three HMI modalities across the same tasks and comparable generic interfaces is a first in function implementation research. The results contribute towards specifying the concept of implementation quality.

9.3 Future Directions

The HMI checklist aims to push function operation towards low-distraction solutions, thereby increasing road safety. Our approach is the first to extensively examine interface design aspects and compare them in a validation approach. We hope that our research helps manufacturers, policy makers, safety organizations, and research to make future vehicles safer for every road user, and not just the driver. This can only be achieved through joint effort, and when the idea of function implementation quality is further researched, discussed, and carried over into practice.

To this end, we would like to point out some key points to overcome and to consolidate the concept of function implementation quality. Our research tested over fifty design aspects with a sample of 130 people, therefore, the sample size per aspect is rather small. Larger sample sizes per aspect (or comparing more aspects per participant) would be helpful in identifying the effects of each design aspect more precisely.

For the design aspects, the different levels on the five-point rating scale could be explored further by testing each level individually. The current validation approach only compares the optimal (best level) and suboptimal (worst level) implementation. Future research could vary implementation quality more gradually by implementing the different steps of the HMI checklist's rating scale. A more general point to consider is that most design aspects were present in the interface concurrently and acted together – both in the reference version (where every aspect was optimally implemented) but also in the design aspect's test interface (the single aspect was implemented suboptimally, the rest stayed at the optimal level). Therefore, technically speaking, a constellation of the aspects was tested that had a combined effect on driving and operation performance (Gestalt psychology: the whole is different from the sum of its parts). As a result, testing the same interface design aspects in a different configuration, i.e., in a different generic interface, may lead to different results. Also, with the gestalt idea in mind, it is hard to say whether and how the effect of a suboptimal implementation of a single design aspect is mediated or moderated by the remaining design aspects. Such interactions should also be considered when implementing functions in production vehicles.

For the validation approach itself, future research could incorporate a medium complex task and not just a simple and complex task. To increase external validity, other distraction measurement tasks could be used (e.g., box task; Trommler et al., 2021) or the transfer into a more complex driving environment could be done. The latter would be useful in understanding how drivers time their operation of functions, e.g., by stopping operation during critical moments. Also, driver errors could be explored this way and connected to the design aspects. The current approach is somewhat unrealistic since drivers were constantly performing lane changes and repeatedly operating the function. In reality, they would use timing strategies to mitigate the effects of driver distraction. Also, driving in real life is more than simply staying on an ideal trajectory. Still, the LCT approach is good in capturing the distraction effects of the design aspects, especially with regard to visual-manual interfaces that are the more pressing issue compared to speech-based interfaces. Also, the focused approach of the LCT allowed us to test the large number of design aspects in the first place.

There are two more elaborate challenges to overcome that go beyond the current validation. First, when implementing a given function in an interface, the operationalization of the interface and design aspects is always rather specific. It would be interesting to explore how each design principle could be applied across different interface variants and to compare these operationalizations against each other. This way, the concept, i.e., definitions, dimensions, of the interface design aspect could be specified. Second, participants had learned the function operation well, which is unsurprising given the repeated use of the interfaces over the course of the experiment. As discussed for some design aspects, especially those that had seemingly no real effect on driver distraction, effects would likely become visible when the driver is not familiar with the function. Perhaps, the function operation process

should be further explored (e.g., interactions, strategies) and researched regarding this “first experience” with a function. The initial contact and operation of the function could point towards pain points and problems in the operation process. When done in the driving simulator, this could capture the distraction dynamics of discovering an unfamiliar function. Also, when the function is then used repeatedly (or in intervals), the learning process could be depicted. Another area of function operation behavior that should be explored are compensation effects, i.e., function operation is stopped and driving is prioritized. It would be interesting to investigate when drivers start prioritizing driving and what limits such a behavior has, especially regarding accidents.

References

- ADAC. (2020, October 19). *Touchscreen im Auto: Wann die Bedienung verboten ist*. <https://www.adac.de/verkehr/recht/verkehrsvorschriften-deutschland/tesla-touchscreen-handyverbot/>
- Ahmed, H. U., Huang, Y., Lu, P., & Bridgelall, R. (2022). Technology Developments and Impacts of Connected and Autonomous Vehicles: An Overview. *Smart Cities*, 5(1), 382–404. <https://doi.org/10.3390/smartcities5010022>
- Alex Sibila (Director). (2023, March 7). *Tesla Touchscreen FULL WALKTHROUGH 2023! (IN DEPTH)*. <https://www.youtube.com/watch?v=rAkXY2uP09o>
- Alliance of Automobile Manufacturers. (2003). *Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems*.
- Apple Inc. (2024). *XCode* (16.2). <https://developer.apple.com/xcode/>
- Braun, M., Mainz, A., Chadowitz, R., Pfleging, B., & Alt, F. (2019). At Your Service: Designing Voice Assistant Personalities to Improve Automotive User Interfaces. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–11. <https://doi.org/10.1145/3290605.3300270>
- Breuer, J. I., Bengler, K., Heinrich, C., & Reichelt, W. (2003). Development of advanced driver attention metrics (ADAM). In H. Strasser, Karsten. Kluth, H. Rausch, & H. Bubb (Eds.), *Quality of Work and Products in Enterprises of the Future* (pp. 37–40). Ergonomia Verlag. <https://mediatum.ub.tum.de/1318660>
- Bubb, H., Bengler, K., Grünen, R. E., & Vollrath, M. (2015). *Automobilergonomie*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-8348-2297-0>
- Burns, P. C., Trbovich, P. L., McCurdie, T., & Harbluk, J. L. (2005). Measuring Distraction: Task Duration and the Lane-Change Test (LCT). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(22), 1980–1983. <https://doi.org/10.1177/154193120504902220>
- Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, 113(10), 2636–2641. <https://doi.org/10.1073/pnas.1513271113>
- Ebel, P., Lingenfelder, C., & Vogelsang, A. (2023). On the forces of driver distraction: Explainable predictions for the visual demand of in-vehicle touchscreen interactions. *Accident Analysis & Prevention*, 183, 106956. <https://doi.org/10.1016/j.aap.2023.106956>
- Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37(1), 32–64. <https://doi.org/10.1518/001872095779049543>
- Eriksson, A., Lindström, A., Seward, A., Seward, A., & Kircher, K. (2014). Can User-Paced, Menu-free Spoken Language Interfaces Improve Dual Task Handling While Driving? In M. Kurosu (Ed.), *Human-Computer Interaction. Advanced Interaction Modalities and Techniques* (Vol. 8511, pp. 394–405). Springer International Publishing. https://doi.org/10.1007/978-3-319-07230-2_38
- Euro NCAP. (2025). *Euro NCAP Protocol—Safe Driving—Driver Engagement*. <https://www.euroncap.com/en/for-engineers/protocols/2026-protocols/>
- European Union. (2008). *Commission Recommendation of 26 May 2008 on safe and efficient in-vehicle information and communication systems: Update of the European Statement of Principles on human-machine interface (notified under document number C(2008) 1742)*. <http://data.europa.eu/eli/reco/2008/653/oj/eng>
- Fahrzeugfunktionen Einfach Erklärt (Director). (2020, August 7). *VW Golf 8, Neuer Caddy: Bedienung und Konfiguration Homescreen und Controlcenter*. <https://www.youtube.com/watch?v=VI-khnZQ7PM>
- Grahn, H., & Kujala, T. (2020). Impacts of Touch Screen Size, User Interface Design, and Subtask Boundaries on In-Car Task's Visual Demand and Driver Distraction. *International Journal of Human-Computer Studies*, 142, 102467. <https://doi.org/10.1016/j.ijhcs.2020.102467>
- Harbluk, J. L., Burns, J. C., Lochner, M., & Trbovich, P. L. (2007, July 10). Using the Lane-Change Test (LCT) to Assess Distraction: Tests of Visual-Manual and Speech-Based Operation of Navigation System Interfaces. *Driving Assessment Conference 2007*. Driving Assessment Conference 2007. <https://doi.org/10.17077/drivingassessment.1208>

- He, J., Chaparro, A., Nguyen, B., Burge, R. J., Crandall, J., Chaparro, B., Ni, R., & Cao, S. (2014). Texting while driving: Is speech-based text entry less risky than handheld text entry? *Accident Analysis & Prevention*, 72, 287–295. <https://doi.org/10.1016/j.aap.2014.07.014>
- Horn, H.-P., Oehme, A., Kröling, S., & Gehlert, T. (2023). *Ablenkung im Fahrzeug*. Gesamtverband der Deutschen Versicherungswirtschaft e.V., Unfallforschung der Versicherer.
- Hua, Z., & Ng, W. L. (2010). Speech recognition interface design for in-vehicle system. *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 29–33. <https://doi.org/10.1145/1969773.1969780>
- Huemer, A. K., & Vollrath, M. (2010). Alcohol-related impairment in the Lane Change Task. *Accident Analysis & Prevention*, 42(6), 1983–1988. <https://doi.org/10.1016/j.aap.2010.06.005>
- Huemer, A. K., & Vollrath, M. (2011). *Driver secondary tasks in Germany: Detailed analysis of face-to-face interviews*. 2nd International conference on driver distraction and inattention. <https://trid.trb.org/View/1345978>
- International Organization for Standardization. (1996). *Passenger cars—Driver hand-control reach* (ISO 3958:1996). <https://www.iso.org/standard/9613.html>
- International Organization for Standardization. (2009). *Road vehicles Location of hand controls, indicators and tell-tales in motor vehicles* (ISO 4040:2009; Version 5). <https://www.iso.org/standard/44856.html>
- International Organization for Standardization. (2010). *Road vehicles—Ergonomic aspects of transport information and control systems—Simulated lane change test to assess in-vehicle secondary task demand* (ISO 26022:2010). <https://www.iso.org/standard/43361.html>
- International Organization for Standardization. (2011). *Road vehicles—Ergonomic aspects of transport information and control systems—Specifications for in-vehicle auditory presentation* (ISO 15006:2011). <https://www.iso.org/standard/55322.html>
- International Organization for Standardization. (2016). *Road vehicles—Transport information and control systems—Detection-response task (DRT) for assessing attentional effects of cognitive load in driving* (ISO 17488:2016). <https://www.iso.org/standard/59887.html>
- International Organization for Standardization. (2017a). *Road vehicles—Ergonomic aspects of transport information and control systems—Specifications and test procedures for in-vehicle visual presentation* (ISO 15008:2017). <https://www.iso.org/standard/62784.html>
- International Organization for Standardization. (2017b). *Road vehicles—Ergonomic aspects of transportation and control systems—Dialogue management principles and compliance procedures* (ISO 15005:2017). <https://www.iso.org/standard/69238.html>
- Japan Automobile Manufacturers Association. (2004). *Guidelines for In-vehicle Display Systems—Version 3.0*. https://www.jama.or.jp/operation/safety/display_system/pdf/jama_guidelines_v30_en.pdf
- Karray, F., Alemzadeh, M., Saleh, J. A., & Arab, M. N. (2017). Human-Computer Interaction: Overview on State of the Art. *International Journal on Smart Sensing and Intelligent Systems*, 1(1), 137–159. <https://doi.org/10.21307/ijssis-2017-283>
- Khdier, H. Y., Jasim, W. M., & Aliesawi, S. A. (2021). Deep Learning Algorithms based Voiceprint Recognition System in Noisy Environment. *Journal of Physics: Conference Series*, 1804(1), 012042. <https://doi.org/10.1088/1742-6596/1804/1/012042>
- Kraftfahrt-Bundesamt. (n.d.). *Monatliche Neuzulassungen Dezember 2024*. Retrieved January 30, 2025, from https://www.kba.de/DE/Statistik/Fahrzeuge/Neuzulassungen/MonatlicheNeuzulassungen/monatl_neuzulassungen_node.html
- Kurtz, M., Oehl, M., & Sutter, C. (2021). “Just One Short Voice Message”—Comparing the Effects of Text- vs. Voice-Based Answering to Text Messages via Smartphone on Young Drivers’ Driving Performances. *Safety*, 7(3), 57. <https://doi.org/10.3390/safety7030057>
- Liu, X., Sun, H., Gao, Y., Zhang, W., Ge, Y., & Qu, W. (2022). Exploring the performance of click and slide gestures on large in-vehicle touch screens. *Applied Ergonomics*, 99, 103613. <https://doi.org/10.1016/j.apergo.2021.103613>

- Loew, A., Koniakowsky, I., Forster, Y., Naujoks, F., & Keinath, A. (2023). The impact of speech-based assistants on the driver's cognitive distraction. *Accident Analysis & Prevention*, 179, 106898. <https://doi.org/10.1016/j.aap.2022.106898>
- Lyashenko, V., Laariedh, F., Sotnik, S., & Ahmad, M. A. (2021). Recognition of Voice Commands Based on Neural Network. *TEM Journal*, 583–591. <https://doi.org/10.18421/TEM102-13>
- Ma, J., & Gong, Z. (2024). *Automotive Human-Machine Interaction (HMI) Evaluation Method*. Springer Nature Singapore. <https://doi.org/10.1007/978-981-99-9951-4>
- Ma, J., Li, J., & Gong, Z. (2022). Evaluation of driver distraction from in-vehicle information systems: A simulator study of interaction modes and secondary tasks classes on eight production cars. *International Journal of Industrial Ergonomics*, 92, 103380. <https://doi.org/10.1016/j.ergon.2022.103380>
- Maciej, J., & Vollrath, M. (2009). Comparison of manual vs. Speech-based interaction with in-vehicle information systems. *Accident Analysis & Prevention*, 41(5), 924–930. <https://doi.org/10.1016/j.aap.2009.05.007>
- Mattes, S. (2003). The Lane-Change-Task as a Tool for Driver Distraction Evaluation. In H. Strasser, Karsten. Kluth, H. Rausch, & H. Bubb, (Eds.), *Quality of Work and Products in Enterprises of the Future* (pp. 57–60). Ergonomia Verlag.
- Meixner, G., & Müller, C. (Eds.). (2017). *Automotive User Interfaces: Creating Interactive Experiences in the Car*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-49448-7>
- Morgenstern, T., Trommler, D., Naujoks, F., Karl, I., Krems, J. F., & Keinath, A. (2023). Comparing the sensitivity of the box task combined with the detection response task to the lane change test. *Transportation Research Part F: Traffic Psychology and Behaviour*, 93, 159–171. <https://doi.org/10.1016/j.trf.2023.01.004>
- Murad, C., Munteanu, C., Clark, L., & Cowan, B. R. (2018). Design guidelines for hands-free speech interaction. *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, 269–276. <https://doi.org/10.1145/3236112.3236149>
- National Highway Traffic Safety Administration. (2013). *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices* (No. 81). <https://www.federalregister.gov/documents/2013/04/26/2013-09883/visual-manual-nhtsa-driver-distraction-guidelines-for-in-vehicle-electronic-devices>
- National Highway Traffic Safety Administration. (2016). *Visual-Manual NHTSA Driver Distraction Guidelines for Portable and Aftermarket Devices*. <https://www.federalregister.gov/documents/2016/12/05/2016-29051/visual-manual-nhtsa-driver-distraction-guidelines-for-portable-and-aftermarket-devices>
- Naujoks, F., Wiedemann, K., Schömig, N., Hergeth, S., & Keinath, A. (2019). Towards guidelines and verification methods for automated vehicle HMIs. *Transportation Research Part F: Traffic Psychology and Behaviour*, 60, 121–136. <https://doi.org/10.1016/j.trf.2018.10.012>
- Nicolas, L., Rod, M., & Thomas, E. (2024). *Designing Touchscreen Menu Interfaces for In-Vehicle Infotainment Systems: The Effect of Depth and Breadth Trade-off and Task Types on Visual-Manual Distraction* (arXiv:2404.11469). arXiv. <https://doi.org/10.48550/arXiv.2404.11469>
- Patel, S., Liu, Y., Zhao, R., Liu, X., & Li, Y. (2022). Inspection of In-Vehicle Touchscreen Infotainment Display for Different Screen Locations, Menu Types, and Positions. In H. Krömker (Ed.), *HCI in Mobility, Transport, and Automotive Systems* (Vol. 13335, pp. 258–279). Springer International Publishing. https://doi.org/10.1007/978-3-031-04987-3_18
- Pitts, M. J., Skrypchuk, L., Wellings, T., Attridge, A., & Williams, M. A. (2012). Evaluating User Response to In-Car Haptic Feedback Touchscreens Using the Lane Change Test. *Advances in Human-Computer Interaction*, 2012(1), 598739. <https://doi.org/10.1155/2012/598739>
- Reichel, S., Sohn, J., Ehrlich, U., Berton, A., & Weber, M. (2014). Out-of-Domain Spoken Dialogs in the Car: A WoZ Study. *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, 12–21. <https://doi.org/10.3115/v1/W14-4303>
- Robertson, D. A., & Cox, E. L. (2024). Steering Away from Automotive HMI De Facto Standards—The Effect on Control Discoverability. *Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 175–183. <https://doi.org/10.1145/3640792.3675726>
- Schwaben - Checker (Director). (2020, January 21). *VW Golf 8 Sprachbedienung Review, Funktion, Verständnis, Innovision Cockpit, Sprachmodus Test, Info*. https://www.youtube.com/watch?v=lqXn__UQdho

- Society of Automotive Engineers. (2015). *Guidelines for Speech Input and Audible Output in a Driver Vehicle Interface* (SAE J2988). https://doi.org/10.4271/J2988_201506
- Society of Automotive Engineers. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles—SAE International*. https://www.sae.org/standards/content/j3016_202104/
- Starkey, N. J., & Charlton, S. G. (2020). Drivers Use of In-Vehicle Information Systems and Perceptions of Their Effects on Driving. *Frontiers in Sustainable Cities*, 2, 39. <https://doi.org/10.3389/frsc.2020.00039>
- Straßenverkehrs-Ordnung. (2013). § 23 StVO. https://www.gesetze-im-internet.de/stvo_2013/_23.html
- Strayer, D. L. (2015). Is the Technology in Your Car Driving You to Distraction? *Policy Insights from the Behavioral and Brain Sciences*, 2(1), 157–165. <https://doi.org/10.1177/2372732215600885>
- Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., & Hopman, R. J. (2017). The smartphone and the driver's cognitive workload: A comparison of Apple, Google, and Microsoft's intelligent personal assistants. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, 71(2), 93–110. <https://doi.org/10.1037/cep0000104>
- Suhm, B. (2003). Towards best practices for speech user interface design. *8th European Conference on Speech Communication and Technology (Eurospeech 2003)*, 2217–2220. <https://doi.org/10.21437/Eurospeech.2003-621>
- Technikfieber (Director). (2023, June 25). *60 Sprachbefehle die JEDER TESLA 2023 versteht im Vergleich mit Mercedes MBUX!* <https://www.youtube.com/watch?v=slyiMxBzfyQ>
- Trommler, D., Morgenstern, T., Wögerbauer, E. M., Naujoks, F., Krems, J. F., & Keinath, A. (2021). The box task—A method for assessing in-vehicle system demand. *MethodsX*, 8, 101261. <https://doi.org/10.1016/j.mex.2021.101261>
- Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2), 167–190. <https://doi.org/10.1016/j.trf.2005.04.014>
- Wang, M., Lee, S. C., Montavon, G., Qin, J., & Jeon, M. (2022). Conversational Voice Agents are Preferred and Lead to Better Driving Performance in Conditionally Automated Vehicles. *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 86–95. <https://doi.org/10.1145/3543174.3546830>
- Weir, D., Kwok, A., & Peak, J. (2007). Exploratory lane change test experiments in support of ISO WG8 LCT task force. *Presentation given at the Meeting of the Task Force Calibration Task*.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159–177. <https://doi.org/10.1080/14639220210123806>
- Wong, P. N. Y., Brumby, D. P., Babu, H. V. R., & Kobayashi, K. (2019). Voices in Self-Driving Cars Should be Assertive to More Quickly Grab a Distracted Driver's Attention. *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 165–176. <https://doi.org/10.1145/3342197.3344535>
- Young, K. L., Lenné, M. G., & Williamson, A. R. (2011). Sensitivity of the lane change test as a measure of in-vehicle system demand. *Applied Ergonomics*, 42(4), 611–618. <https://doi.org/10.1016/j.apergo.2010.06.020>
- Young, R. A., & Zhang, J. (2017). Driven to Distraction? A Review of Speech Technologies in the Automobile. *Voice Interaction Design*.
- Zhang, F., & Roberts, S. C. (2023). Factors affecting drivers' off-road glance behavior while interacting with in-vehicle voice interfaces. *Accident Analysis & Prevention*, 179, 106883. <https://doi.org/10.1016/j.aap.2022.106883>
- Zhang, T., Liu, X., Zeng, W., Tao, D., Li, G., & Qu, X. (2023). Input modality matters: A comparison of touch, speech, and gesture based in-vehicle interaction. *Applied Ergonomics*, 108, 103958. <https://doi.org/10.1016/j.apergo.2022.103958>
- Zhong, Q., Guo, G., & Zhi, J. (2022). Address inputting while driving: A comparison of four alternative text input methods on in-vehicle navigation displays usability and driver distraction. *Traffic Injury Prevention*, 23(4), 163–168. <https://doi.org/10.1080/15389588.2022.2047958>

List of Figures

Figure 1: Function Cluster for Optimal Operability	15
Figure 2: Checklist Structure	24
Figure 3: Item Structure	25
Figure 4: Example of a bonus item.	26
Figure 5: Decision tree for determining the optimal implementation	28
Figure 6: Basic driving simulator setup used for all three generic HMIs (physical controls, touch-based, speech-based).....	33
Figure 7: Physical controls: Physical steering wheel controls (left) and software interface (right).....	34
Figure 8: Touch-based interface: Touch interface used for the simple task (left) and complex task (right).	34
Figure 9: Speech-based interface: Software interface.	35
Figure 10: Lateral driving performance for every combination of task (simple, complex) and HMI type (physical controls, touch-based, speech-based). Baseline performance indicated by dotted line. Sorted in ascending order by MDev. RV = Reference version.	38
Figure 11: Reaction time for every combination of task (simple, complex) and HMI type (physical controls, touch-based, speech-based). Baseline performance indicated by dotted line. Sorted in ascending order by MDev. RV = Reference version.....	39
Figure 12: Interaction diagram for task (simple, complex) and HMI type (physical controls, touch-based, speech-based): Mean deviation (MDev) change score.	40
Figure 13: Interaction diagram for task (simple, complex) and HMI type (physical controls, touch-based, speech-based): Standard deviation of lane position (SDLP) change score.	41
Figure 14: Interaction diagram for task (simple, complex) and HMI type (physical controls, touch-based, speech-based): Reaction time (RT) change score.....	42

List of Tables

Table 1: Volkswagen Golf 8 – functions accessible via physical controls and touch-based HMIs	10
Table 2: Volkswagen Golf 8 – functions accessible via the speech-based interface.....	11
Table 3: Tesla Model 3 – functions accessible via physical controls and touch-based HMIs	12
Table 4: Tesla Model 3 – functions accessible via the speech-based interface	13
Table 5: Visual-Manual Guideline Criteria Summary – Adapted from Horn et al. (2023).....	17
Table 6: Speech-Based Standard Criteria Summary	20
Table 7: Example Functions on Each Level of the Decision Tree	28
Table 8: Experimental Design	31
Table 9: Score Overview	45
Table 10: Validation Results for the Design Aspects of Physical Controls	48
Table 11: Physical Controls - Susceptibility to Operation Errors	50
Table 12: Physical Controls - Viewing Distance to the Right	51
Table 13: Physical Controls - Blind Operation	52
Table 14: Physical Controls - Glare.....	53
Table 15: Physical Controls - Expectation Conformity	54
Table 16: Physical Controls - Reachability	55
Table 17: Physical Controls - Correction of Operation Errors	56

Table 18: Physical Controls - Risk of Operation Errors	57
Table 19: Physical Controls - Viewing Distance Downwards	58
Table 20: Physical Controls - Additional Feedback.....	59
Table 21: Physical Controls - Comprehensibility	60
Table 22: Physical Controls - Legibility	61
Table 23: Validation Results for the Design Aspects of the Touch-Based Interface	64
Table 24: Touch-Based Interface - Input Precision	67
Table 25: Touch-Based Interface - Susceptibility to Operation Errors	68
Table 26: Touch-Based Interface - Expectation Conformity	69
Table 27: Touch-Based Interface - Reachability	70
Table 28: Touch-Based Interface - Gesture Control	71
Table 29: Touch-Based Interface - Risk of Operation Errors.....	72
Table 30: Touch-Based Interface - Correction of Operation Errors	73
Table 31: Touch-Based Interface - Driving-related Information Always Visible	74
Table 32: Touch-Based Interface - Interaction Speed	75
Table 33: Touch-Based Interface - Additional Feedback	76
Table 34: Touch-Based Interface - Reflection	77
Table 35: Touch-Based Interface - Viewing Distance Downwards.....	78
Table 36: Touch-Based Interface - Tilt	79
Table 37: Touch-Based Interface - Legibility.....	80
Table 38: Touch-Based Interface - Brightness and Contrast.....	81
Table 39: Touch-Based Interface - Attentional Guidance	82
Table 40: Touch-Based Interface - Minimalism	83
Table 41: Touch-Based Interface - In-process Abortion	84
Table 42: Touch-Based Interface - Viewing Distance to the Right.....	85
Table 43: Touch-Based Interface - View on the Road Not Blocked.....	86
Table 44: Touch-Based Interface - Shortcut.....	87
Table 45: Touch-Based Interface - Deactivation	88
Table 46: Touch-Based Interface - Blocking of Distracting Functions	89
Table 47: Validation Results for the Design Aspects of the Speech-Based Interface	92
Table 48: Speech-Based Interface - Fallback	95
Table 49: Speech-Based Interface - Correction of Operation Errors	96
Table 50: Speech-Based Interface - Deactivation Option & Alternative Operation Option.....	97
Table 51: Speech-Based Interface - Quality of the Speech Recognition	98
Table 52: Speech-Based Interface - Susceptibility to Operation Errors	99
Table 53: Speech-Based Interface - Hands-free.....	100
Table 54: Speech-Based Interface - Shortcut.....	101
Table 55: Speech-Based Interface - Risk of Operation Errors.....	102
Table 56: Speech-Based Interface - Additional Feedback	103
Table 57: Speech-Based Interface - Diverse Commands	104
Table 58: Speech-Based Interface - Robustness of Activation	105
Table 59: Speech-Based Interface - Intelligibility / Quality of Voice Output	106
Table 60: Speech-Based Interface - In-process Abortion	107
Table 61: Speech-Based Interface - Feedback Dimensions	108

Table 62: Speech-Based Interface - Interaction Speed	109
Table 63: Speech-Based Interface - Volume	110
Table 64: Physical Controls - Design Aspect Summary	111
Table 65: Touch-based Interface - Design Aspect Summary	112
Table 66: Speech-based Interface - Design Aspect Summary	114
Table 67: Suggestions for Euro NCAP's General Vehicle Controls (GVC) Assessment: Physical Controls	117
Table 68: Suggestions for Euro NCAP's General Vehicle Controls (GVC) Assessment: Touch-Based HMIs	118
Table 69: Suggestions for Euro NCAP's General Vehicle Controls (GVC) Assessment: Speech-Based HMIs	120

HMI Checklist: Distraction and Operation Safety

Assessment Scope

1) Define the function(s) to be assessed:

This assessment tool only evaluates functions provided by the vehicle manufacturer. It does not cover third-party integrations such as Apple CarPlay or Google Android Auto. Additionally, functions directly affecting vehicle dynamics (acceleration, braking, steering, and gear shifting) are outside the scope of this assessment.

Please tick the function group(s) relevant to your assessment. If your target functions don't fit into the existing categories, you may add a custom function group.

	Function Group	Functions
☐	Light and visibility	Exterior Lights, windshield wipers, defogging windows, dimming mirrors
☐	Extended longitudinal control	Gear shifter/gear selector
☐	(Warning) signals	Horn, headlight flashing, indicators
☐	Assistance	ACC, LDW
☐	Windows and interior light	Windows, interior lighting, ambience
☐	Trip data and service	Trip distance, duration, fuel consumption
☐	Media Control	Volume, track, playback control
☐	Climate Control	Heated seats, adjust temperature, airflow
☐	Social	Calls, messages
☐	Setup	Align headlights, reset tire pressure control
☐	Custom:	

2) Determine the HMI Types(s) that are currently employed for operating the function(s):

A function can be controlled via different types of human-machine interfaces (HMIs). They can be distinguished as follows:

Physical Controls: Physical controls are primarily physical-haptic in nature. Often, an element controls only a single function. They are usually mechanical (levers, switches, buttons, rotary switches) and are operated by manual interaction. Physical controls provide tactile feedback that is generated by the movement of the control element. Touch buttons that are not part of the central touchscreen are also included in this group since they too represent a physical interface that only controls a single function, although the tactile feedback might be missing. The actions for interacting with physical controls include pressing, turning or pulling.

Touch-Based HMI: Touch-based interfaces are usually implemented via a central touch display located in the middle of the vehicle cockpit. They are flexible in the structure of their visual interface and can therefore be used to control a wide range of functions. Interaction takes place by touching the screen, tapping, or swiping on it. Feedback is primarily provided visually in the display, but acoustic or tactile feedback via actuators is also possible.

Speech-Based HMI: Speech-based interfaces allow functions to be controlled via verbal commands and usually provide auditory feedback. The range of functions controllable by voice varies depending on the manufacturer. The voice-based interface can be activated via a push-to-talk button or via the use of a keyword ("Hey XY!"). Voice recognition is handled by algorithms that can understand the intention of the driver's commands. Feedback is mostly provided via the use of chimes and/or speech.

Please list all the functions that you would like to assess in the table below. If possible, also assign a function group. Finally, indicate which HMI type(s) is/are currently employed for the operation of the function(s).

Function Group: _____	HMI Type		
	Physical Controls	Touch-Based HMI	Speech-Based HMI
Functions:			
1.	☐	☐	☐
2.	☐	☐	☐
3.	☐	☐	☐
4.	☐	☐	☐
5.	☐	☐	☐
6.	☐	☐	☐
Implementations to be checked in module I and II b:	☐	☐	☐

Now begin the assessment by starting with module I. Please evaluate all the items of the HMI type(s) that are used by the function(s) which you would like to evaluate.

Module I – Shared HMI Features

For each HMI type(s) used by the function(s) (physical controls, touch-based HMI, and/or speech-based HMI), please check whether and how the criteria listed below apply. If a rating of “unacceptable” is given for any item, please do not continue with the assessment. Here, the HMI type should be revised to ensure an acceptable operation level. You may, however, proceed with the check of a/the remaining HMI type(s).

The first module must be only completed once – a repetition for every function that shall be assessed is not necessary. The criteria in module I apply to the interface in general and thus are shared by all functions that are implemented in it.

Module I – Physical Controls

The following criteria must be checked for the HMI type. With each item, the deviation from an ideal design is determined. The sum (score on the rating scale multiplied with the weighting factor) of all items results in an HMI operation score. The items most relevant regarding driver distraction are listed first.

M1-PC-1 – Viewing Distance to the Right				
The control element is not too far away from the driver's line of sight (horizontally - "to the right").				
0° - 15°	15° - 30°	30° - 45°	45° - 60°	> 60°
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 3)	+ (0.05 × 3.8)	= 1.49
Explanation	The control element should be as close as possible to the driver's line of sight when looking straight ahead (0°) at the road. This is regarding the view "to the right", which should be as close as possible to the driver's line of sight.			
Testing	Look straight ahead at the road. Then look to the right at the control element. Can you see the control element with just your eye movement (up to 30°), or do you need to move your head (more than 30°) as well?			
References	Visual-Manual Guideline Criteria Summary – Installation Naujoks et al. (2019)			

M1-PC-2 – Blind Operation				
The control element has haptic features that allow for blind operation.				
Haptics when reaching the control element, haptic feedback during operation, setting can be felt.	Haptics when reaching the control element, haptic feedback during operation, setting can be felt.	Haptics when reaching the device, haptic feedback during operation.	Haptics when reaching the control element.	No haptic properties that allow blind operation.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 2)	+ (0.05 × 2.2)	= 1.31
Explanation	It should be possible to operate the control element blindly in all steps of the operation process: when reaching it, when choosing a setting, and the setting should also be able to be felt.			
Testing	Carry out a complete operation (reaching, setting, feeling the setting).			
References	Visual-Manual Guideline Criteria Summary – Interaction Ma et al. (2022)			

M1-PC-3 – Glare				
The control element does not cause glare when exposed to sunlight during the daytime or when light from streetlamps falls on it at night. The surface and position of the element prevents that light can reach it.				
No glare, matte surface, concealed position.	Some glare, matte surface, concealed position.	Some glare, glossy surface, concealed position.	Glare, reflective surface, concealed position.	Lots of glare, very reflective surface, exposed position.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 2)	+ (0.05 × 1.5)	= 1.27
Explanation	The control element should be designed and positioned in such a way that light is not reflected towards the driver.			
Testing	First check the surface and positioning of the instrument. Consider whether the instrument could cause reflections that would lead to distraction. If so, ask another person to shine a light source from outside the vehicle onto the control element at different angles. Look straight ahead at the road and check whether and how much you notice the reflections.			
References	Visual-Manual Guideline Criteria Summary – Installation			

M1-PC-4 – Reachability				
The control element has been installed in a position where it can be easily reached by the driver.				
Both hands on the steering wheel.	One hand on the steering wheel, arm not extended.	One hand on the steering wheel, arm slightly extended.	One hand on the steering wheel, arm widely extended.	Both hands necessary.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 1)	+ (0.05 × 4.5)	= 1.33
Explanation	The control element should be close to the driver so that minimal motor effort is needed.			
Testing	Put both hands on the steering wheel. Now operate the control element and pay attention to your hands and whether one hand needs to be removed from the steering wheel. If so, pay attention to how much you need to reach for the control element.			
References	Visual-Manual Guideline Criteria Summary – Installation, Interaction ISO 3958:1996 & ISO 15005:2017			

M1-PC-5 – Viewing Distance Downwards				
The control element is not too far away from the driver's line of sight (vertical - "downwards").				
0° - 10°	10° - 20°	20° - 30°	30° - 45°	> 45°
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects			Operation Quality ↓	
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 3)	= 1.15
Explanation	The control element should be as close as possible to the driver's line of sight when looking straight ahead (0°) at the road. This is regarding the view "downwards", which should be as close as possible to the driver's line of sight.			
Testing	Look straight ahead at the road. Then look down at the control element. Can you see the control element with just your eye movement (up to 30°), or do you need to move your head as well (from approx. 30°)?			
References	Visual-Manual Guideline Criteria Summary – Installation Naujoks et al. (2019)			

M1-PC-6 – Legibility				
The labeling of the control element is high-contrast, large, and distinct, so it only takes a short time to look at it.				
Labeling: high-contrast, distinct, appropriate size.	Labeling - one criterion violated: high-contrast, clear, appropriate size.	Labeling - two criteria violated: high-contrast, clear, appropriate size.	Labeling - three criteria violated: high-contrast, clear, appropriate size.	No labeling.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects				
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0)	= 1
Explanation	The labeling of the control element should be easily visible. This is only about legibility, not about the meaning of the labeling (that is checked in module II b). The labeling should also be distinctive (unambiguous) so that attention is not drawn to irrelevant functions with similar symbols.			
Testing	Look at the control element and evaluate if the contrast, distinctiveness (symbols/texts do not overlap with neighboring control elements), and size are appropriate.			
References	Visual-Manual Guideline Criteria Summary – Information Presentation ISO 15005:2017			

Module I – Touch-Based HMI

The following criteria must be checked for the HMI type. With each item, the deviation from an ideal design is determined. The sum (score on the rating scale multiplied with the weighting factor) of all items results in an HMI operation score. The items most relevant regarding driver distraction are listed first.

M1-TB-1 – Input Precision				
The display registers touch inputs with high precision.				
Every touch input is registered.	Almost every touch input is registered.	Every now and then something is not registered.	Often something is not registered.	Touch inputs unreliable
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 6)	+ (0.05 × 4.5)	= 1.83
Explanation	The touchscreen should reliably register all finger inputs (taps or swipes).			
Testing	Use the touchscreen and make sure that all inputs (taps and swipes) are reliably registered. We recommend wearing gloves or using people with "dry" hands for the test. If possible, operation should also be tested at different temperatures to determine the functionality of the touchscreen in very cold and warm environments.			
References	Visual-Manual Guideline Criteria Summary – Interaction ISO 15005:2017			

M1-TB-2 – Reachability				
The touchscreen has been installed in a position where it can be easily reached by the driver.				
One hand on the steering wheel, arm not extended, seating position is maintained.	One hand on the steering wheel, arm slightly extended, seating position maintained.	One hand on the steering wheel, arm widely extended, seating position maintained.	One hand on the steering wheel, arm widely extended, leaving seating position.	Both hands necessary.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 4)	+ (0.05 × 3)	= 1.55
Explanation	The touchscreen should be close to the driver so that the hand and arm movements needed for the operation are minimized.			
Testing	First take both hands on the steering wheel. Now operate the touchscreen and pay attention to how far you need to reach with your right hand and whether you need to move your upper body in the direction of the screen as well. Try to operate the controls on the right-hand edge of the screen.			
References	Visual-Manual Guideline Criteria Summary – Installation & Interaction ISO 15008:2017 & ISO 3958:1996 Ma & Gong (2024)			

M1-TB-3 – Driving-related Information Always Visible				
Driving-related information is always visible. When presented in a separate display, the information is not physically covered by the touchscreen. When displayed in the touchscreen interface, the information is not covered by other elements.				
Information always visible, driving-related displays separated from touchscreen.	Information always visible, touchscreen contains information, fixed position.	Information always visible, touchscreen contains information, variable position.	Information sometimes covered, touchscreen contains information, fixed or variable position.	Information always covered.
0 □	1 □	2 □	3 □	unacceptable □
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 1)	+ (0.05 × 0.8)	= 1.14
Explanation	Information relevant to driving (e.g., speed) should always be available to the driver. Ideally, the information should be shown in separate displays and not just on the touchscreen. If driving-relevant information is only displayed on the touchscreen, it should always be visible and not be occluded by other interface elements.			
Testing	Check where driving-relevant information is displayed. If the information is only displayed on the touchscreen, systematically test some touchscreen functions and test whether the driving-relevant information can be shifted or obscured by the other function.			
References	Visual-Manual Guideline Criteria Summary - Installation			

M1-TB-4 – Reflection				
The surface of the touchscreen is anti-reflective (anti-glare coating or matte) so that there are no reflections.				
Matte, strong light absorption, not disturbing when switched off.	Matte, weak light absorption, slightly disturbing when switched off.	Glossy, anti-reflective coating, slightly disturbing when switched off.	Glossy, anti-reflective, very distracting when switched off.	No anti-reflective surface treatment.
0 □	1 □	2 □	3 □	unacceptable □
Weighting		Driving Performance ↓	Operation Quality ↓	
Affects	1.0	+ (0.1 × 1)	+ (0.05 × 2.2)	= 1.21
Explanation	The surface of the touchscreen should be designed in such a way that reflections are minimized, especially in bright driving environments.			
Testing	If possible, switch the touchscreen off. Look for reflections or shine a light source onto the display. Assess the light absorption for a matte display and the anti-reflective coating for a glossy display.			
References	Visual-Manual Guideline Criteria Summary - Installation ISO 15008:2017			

M1-TB-5 – Viewing Distance Downwards				
The touchscreen is not too far away from the driver's line of sight (vertical - "downwards").				
0° - 10°	10° - 20°	20° - 30°	30° - 45°	> 45°
0 □	1 □	2 □	3 □	unacceptable □
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 1)	+ (0.05 × 2.2)	= 1.21
Explanation	The touchscreen should be as close as possible to the driver's line of sight when the looking straight ahead (0°) at the road. This is regarding the view "downwards", which should be as close as possible to the driver's line of sight.			
Testing	Look straight ahead at the road. Then look down at the touchscreen. Try to look at the bottom edge of the display. Can you see it with just your eye movement (up to 30°), or do you also have to move your head (from approx. 30°)?			
References	Visual-Manual Guideline Criteria Summary – Installation ISO 15008:2017 Ma & Gong (2024)			

M1-TB-6 – Tilt				
The touchscreen is tilted towards the driver.				
Strong horizontal and vertical tilt towards the driver.	Slight horizontal and vertical tilt towards the driver.	Strong tilt only vertical or horizontal.	Weak tilt only vertical or horizontal.	No tilt.
0 □	1 □	2 □	3 □	4 □
Affects			Operation Quality ↓	
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 1.5)	= 1.07
Explanation	The touchscreen should be tilted towards the driver to increase visibility and reduce vertical and horizontal operating distances.			
Testing	Look at the touchscreen and assess whether it is tilted and if so, how much and in which direction.			
References	Visual-Manual Guideline Criteria Summary) – Installation & Information Presentation ISO 15008:2017 Ma & Gong (2024)			

M1-TB-7 – Brightness and Contrast				
The brightness and contrast of the display allow for optimal readability in both the brightest (sun shining directly on the display) and the darkest (night) lighting conditions. The brightness can be adjusted and is also regulated automatically.				
Contrast and brightness (maximum and minimum) optimal, adjustable and automatic regulation.	Contrast or brightness (maximum and minimum) suboptimal, adjustable and automatic regulation.	Contrast and brightness (maximum and minimum) suboptimal, adjustable and automatic regulation.	Contrast and brightness (maximum and minimum) suboptimal, adjustable, no automatic regulation.	Contrast and brightness (maximum and minimum) insufficient, not adjustable, no automatic regulation.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects			Operation Quality ↓	
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 1.5)	= 1.07
Explanation	The touch display is easy to read and easy on the eyes (not blinding) in both very bright and very dark environments without disturbing the driver.			
Testing	Try adjusting the brightness of the display. Test the maximum brightness in a bright environment and the minimum brightness in a dark environment. Change the environment while driving and ensure that the brightness is automatically adjusted.			
References	Visual-Manual Guideline Criteria Summary – Information Presentation ISO 15008:2017 Naujoks et al. (2019)			

M1-TB-8 – Attentional Guidance				
The interface is designed in such a way that it conveys interaction principles and guides attention towards relevant elements without being distracting.				
Interaction principles are optimally communicated, adequate attentional guidance.	Interaction principles are only partially communicated, adequate attentional guidance.	Interaction principles are not communicated, no attentional guidance.	Principles of interaction are not communicated, distracting attentional guidance.	Principles of interaction confusing, strongly distracting attentional guidance.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects			Operation Quality ↓	
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 1.5)	= 1.07
Explanation	The surface of the touchscreen is designed in a way that supports the interaction and prevents operation errors.			
Testing	Operate different functions by switching them on, off and, if possible, adjusting them. Pay attention to the design of the touch interface. Are interaction principles conveyed, e.g., by indicating that buttons are pressable surfaces or showing that sliders can be moved in certain way? Also, is attention also guided towards the relevant control elements, either through their arrangement, color scheme or contrast? Finally, make sure that attention is not drawn to areas that are irrelevant for the function operation.			
References	Visual-Manual Guideline Criteria Summary – Information Presentation ISO 15005:2017			

M1-TB-9 – Minimalism				
The touchscreen interface has a minimalist design and information is presented compactly. This minimizes visual search times and the associated distraction.				
All content is short and concise.	Almost all content is short and concise.	Content is sometimes too long.	Contents rather long.	Contents too long.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects			Operation Quality ↓	
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0.8)	= 1.04
Explanation	The touchscreen's interface and its contents are not overloaded. The structure of the system can be quickly and easily understood.			
Testing	Look at the interface of various functions and evaluate the information presented in terms of brevity and conciseness.			
References	Visual-Manual Guideline Criteria Summary – Information Presentation ISO 15005:2017 Naujoks et al. (2019)			

M1-TB-10 – Viewing Distance to the Right				
The touchscreen is not too far away from the driver's line of sight (horizontal - "to the right").				
0° - 15°	15° - 30°	30° - 45°	45° - 60°	> 60°
0 □	1 □	2 □	3 □	unacceptable □
Affects			Operation Quality ↓	
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0.8)	= 1.04
Explanation	The touchscreen should be as close as possible to the driver's line of sight when looking straight ahead (0°) at the road. This is regarding the view "to the right", which should be as close as possible to the driver's line of sight.			
Testing	Look straight ahead at the road. Then look to the right at the touchscreen. Try to look at the right edge of the display. Can you see it with just your eye movement (up to 30°), or do you also have to move your head (from approx. 30°)?			
References	Visual-Manual Guideline Criteria Summary – Installation ISO 15008:2017 Ma & Gong (2024)			

M1-TB-11 – View on the Road Not Blocked				
The touchscreen has been installed in such a way that it does not block the view on the road.				
Clear field of view.	Peripheral field of view blocked when looking to the right.	Central field of view blocked when looking to the right.	Peripheral field of view blocked when looking straight ahead.	Central field of view blocked when looking straight ahead.
0 □	1 □	2 □	3 □	unacceptable □
Affects		Driving Performance ↓		
Weighting	1.0	+ (0.1 × 1)	+ (0.05 × 0)	= 1.1
Explanation	The touchscreen should not cover any areas of the driving environment.			
Testing	First look straight ahead onto the road. Then keep looking to the right until you reach the right-hand A-pillar. Pay attention to whether, when and which area of your field of view is covered by the touchscreen.			
References	Visual-Manual Guideline Criteria Summary – Installation			

M1-TB-12 – Deactivation				
The touchscreen can be switched off unless it contains driving-related information.				
Touchscreen can be switched off.	Touchscreen can be dimmed considerably.	Display of a static interface (e.g., home screen).	Display of a standby interface (e.g., clock).	Switching off not possible OR Switching off possible despite displaying driving-relevant information.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects				
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0)	= 1
Explanation	The central display should be able to be switched off or at least be able to be set to a state that reduces distraction.			
Testing	Check whether you can switch off the display. If this is not the case, test whether the display can be dimmed considerably or switched to a static state.			
References	Visual-Manual Guideline Criteria Summary - System Response ISO 15005:2017			

M1-TB-13 – Blocking of Distracting Functions				
The system blocks functions that are highly distracting.				
The following content is hidden while driving: Text input, video, scrolling text, images, irrelevant text.	No blocking of images & text.	No blocking of the keyboard.	No blocking of dynamic content.	No blocking of highly distracting content.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects				
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0)	= 1
Explanation	The touch display blocks content that introduces high levels of distraction.			
Testing	Systematically search through the entertainment functions and check whether they are blocked while driving. Videos containing color and brightness changes are particularly distracting. This is followed by dynamic content such as scrolling text. With keyboards, the distraction is not due to their display per se, but rather, that drivers could try to interact with them while driving. Finally, static images and text that are not related to the driving task should be hidden.			
References	Visual-Manual Guideline Criteria Summary – System Response ISO 15005:2017			

Module I – Speech-Based HMI

The following criteria must be checked for the HMI type. With each item, the deviation from an ideal design is determined. The sum (score on the rating scale multiplied with the weighting factor) of all items results in an HMI operation score. The items most relevant regarding driver distraction are listed first.

M1-SB-1 – Deactivation Option & Alternative Operation Option				
The key phrase activation can be turned off. The entire speech-based HMI can also be deactivated. All voice assistant functions are also available via the touch-based HMI and physical controls.				
Key phrase can be deactivated, assistant can be deactivated, correct fallback of functions (physical controls & touch correct).	Key phrase can be deactivated, assistant can be deactivated, correct fallback of functions (physical controls & touch correct).	Key phrase can be deactivated, assistant cannot be deactivated, incorrect fallback of functions (physical controls & touch correct).	Key phrase cannot be deactivated, assistant cannot be deactivated, incorrect fallback of functions (physical controls & touch correct).	Key phrase cannot be deactivated, assistant cannot be deactivated, no fallback of functions.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 5)	+ (0.05 × 1.5)	= 1.57
Explanation	The speech-based HMI offers the option to turn off the activation by key phrase. In addition, the entire assistant can be deactivated to eliminate any risk of cognitive distraction. All functions are therefore also available via other, appropriate interfaces (physical controls or touch-based, see decision tree in module II b).			
Testing	Look through the speech-based HMI's menu (e.g., in the touch display) for options to deactivate the activation via key phrase. Also check whether the speech-based HMI can be deactivated completely. Finally, test whether functions can also be accessed via physical controls and the touchscreen.			
References	Hua & Ng (2010) Suhm (2003)			

M1-SB-2 – Quality of the Speech Recognition				
The speech recognition is of high-quality and works well with dialects and background noise.				
Every voice input is recognized, even with a dialect or background noise.	Almost every voice input is recognized even with a dialect or background noise.	Sometimes voice input is not recognized due to dialects or background noise.	Voice input is often not recognized due to dialects or background noise.	Voice input completely unreliable with dialects or background noise.
0 □	1 □	2 □	3 □	unacceptable □
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 3)	+ (0.05 × 4.5)	= 1.53
Explanation	The speech recognition is able to identify commands of varying speech qualities. It can understand words despite dialects or background noise.			
Testing	Start a dialog with the system and try to operate various functions. To test the robustness of the speech recognition, the operation should be carried out by at least two different people in order to identify any major problems with the speech recognition. It would be ideal to use a person who can speak in a dialect. Also check to which extent background noise (e.g., hissing, clicking, rustling, etc.) interferes with speech recognition. You could try making noises with objects or playing them back on an external device.			
References	Speech-Based Standard Criteria Summary – Interaction			

M1-SB-3 – Hands-free				
Hands-free operation of the speech-based HMI is possible.				
Initial activation via a key phrase.	Activation via key phrase before each entry.	Initial activation via push-to-talk button on the steering wheel.	Push-to-talk button before each input on the steering wheel.	Push-to-talk button not on steering wheel.
0 □	1 □	2 □	3 □	unacceptable □
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 4)	+ (0.05 × 0.8)	= 1.44
Explanation	The speech-based HMI can be operated without using your hands.			
Testing	Try to initiate a dialog using a key phrase ("Hey, XY"). If this is not possible and if the system uses push-to-talk, observe if you only must press the button once at the start of the whole conversation or before each voice input.			
References	Speech-Based Standard Criteria Summary – Interaction			

M1-SB-4 – Robustness of Activation				
The speech-based HMI is not too easy to activate accidentally.				
Accidental activation unlikely. Push-to-talk button not easily accessible. No reaction to similar phrases to key phrase. No response to passengers.	Accidental activation rather unlikely. Push-to-talk button not easily accessible. Reaction to similar phrases to key phrase. No response to passengers.	Accidental activation likely. Push-to-talk button easily accessible. Reaction to similar phrases to key phrase. Reaction to passengers.	Accidental activation very likely. Push-to-talk button easily accessible. Reaction to similar phrases to key phrase. Reaction to passengers.	Permanent accidental activation. Push-to-talk button easily accessible. Reaction to any noise. Reaction to passengers.
0 □	1 □	2 □	3 □	unacceptable □
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 1)	+ (0.05 × 0.8)	= 1.14
Explanation	The speech-based HMI cannot be easily activated by mistake - neither via a push-to-talk button nor a key phrase.			
Testing	Try to trigger the speech-based HMI unintentionally. If a push-to-talk button is installed, think about how easily it can be triggered accidentally. If the system uses a key phrase, try to incorporate similar sounding phrases into a sentence. Also check whether the system can distinguish between driver and passengers.			
References	Speech-Based Standard Criteria Summary – Interaction Naujoks et al. (2019)			

M1-SB-5 – Intelligibility / Quality of Voice Output				
The output speech is characterized by a high level of comprehensibility. The pronunciation is free from any dialect. The words can be easily distinguished from one another.				
Perfectly natural language.	Very natural language.	Almost natural language.	Rather artificial language.	No voice output.
0 □	1 □	2 □	3 □	unacceptable □
Affects			Operation Quality ↓	
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 1.5)	= 1.07
Explanation	If the speech-based HMI provides speech output, the language should be easily understandable (intelligibility) as to not distract the driver.			
Testing	Start a dialog with the system and activate various functions. Evaluate the language used by the system.			
References	Speech-Based Standard Criteria Summary – Audio Signal Suhm (2003)			

M1-SB-6 – Feedback Dimensions				
The speech-based HMI primarily uses acoustic feedback in the form of speech output. Optionally, chimes are used. Visual feedback is shown to support the operation.				
Speech output with chimes and visual feedback.	Speech output with chimes. No visual feedback.	Speech output with visual feedback. No chimes.	Speech output. No chimes or visual feedback.	Visual feedback only.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects				
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0)	= 1
Explanation	The speech-based HMI uses speech output as the primary feedback to avoid that the driver looks away from the road. In addition, chimes are used to make the dialog and its segments more distinctive and to indicate whether the system is listening or processing a command. Visual feedback is only used as a fallback option in case the driver has not adequately understood the speech output.			
Testing	Start a dialog with the system and activate and adjust various functions. Pay attention to which feedback types are used by the system.			
References	Speech-Based Standard Criteria Summary – Information Presentation Hua & Ng (2010)			

M1-SB-7 – Volume				
The speech output uses an appropriate volume. The volume is adjustable and is automatically adjusted at high noise levels. Other audio sources are reduced during speech output.				
Volume appropriate in default setting. The volume is adjustable. Automatic control and reduction of other audio sources.	Volume slightly too loud/quiet in default setting. The volume is adjustable. Automatic control and reduction of other audio sources.	Volume too loud/quiet in default setting. Volume adjustable, but no automatic control or reduction of other audio sources.	Volume too loud/quiet in default setting. Volume adjustable, but no automatic control and reduction of other audio sources.	Volume not adjustable. Or No speech output.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects			Operation Quality ↓	
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0.8)	= 1.04
Explanation	By default, the speech-based HMI is loud enough to be easily understood. The volume can be adjusted if necessary. To increase intelligibility, the system automatically adjusts the volume and reduces the noise level of potentially disturbing audio content (e.g., music).			
Testing	First start a dialog with the system and evaluate the default volume. Now try to adjust the volume. Then drive the vehicle and check whether the system increases the volume. Start audio content (e.g., radio) and initiate a dialog. Check whether the system reduces the volume of the other audio content.			
References	Speech-Based Standard Criteria Summary – Audio Signal			

Analysis

Module I – Shared HMI Features

After answering all items of the HMI types used by the function(s) you want to assess, please sum up the ratings to calculate the HMI operation score. If your function(s) do/es not use an HMI type, check the “not used” box. If an item scored an “unacceptable” check the “unacceptable” box.

HMI Type	Not used?	Unacceptable?	HMI Operation Score	Point Range
Physical Controls	☐	☐		0 (best) → 23 (worst)
Touch-Based HMI	☐	☐		0 (best) → 46 (worst)
Speech-Based HMI	☐	☐		0 (best) → 27 (worst)

Transfer the results to the evaluation sheet. If "unacceptable" occurred in module I, continue with module II on the next page.

Module II – Function-Specific HMI Features

In the second module, the implementation (HMI type & operation steps) of each individual function is checked. Repeat the module until all the functions which you would like to evaluate have been assessed in both parts (a & b) of module II.

Module II a – Optimal HMI implementation

Please answer the following questions to determine the use context of the function:

Question 1: What is the cause for the use of the function?

- external factors (in relation to the driving task): e.g., using indicators due to the infrastructure, turning off the cruise control because of an obstacle, enabling windshield wipers in the rain
- internal (voluntary by the driver): e.g., changing the song in the media control, setting the temperature in the climate control, answering a call

If "external" in question 1 → **Question 2:** How much time is available to operate the function? By when must the operation be completed?

- instantly: e.g., braking, swerving
- a few seconds: e.g., setting cruise control, switching on windshield wipers, switching on lights
- longer-term: e.g., adding a petrol station to the navigation, avoid traffic jams in the navigation system

If "longer term" in question 2 → **Question 4:** How great is the risk of and how severe are the effects of an erroneous operation?

- high risk: interference with lateral and longitudinal control (e.g., assistant switched off), strong distraction (e.g., radio very loud)
- low risk: slight distraction (e.g., ventilation at full strength)

If "internal" in question 1 → **Question 3:** How often is function used?

- often: e.g., adjusting the volume of the radio, adjusting climate control
- seldomly: e.g., changing settings in the configuration menu, pairing a new phone

If "often" in question 3 → **Question 4:** How great is the risk of and how severe are the effects of an erroneous operation?

- high risk: interference with lateral and longitudinal control (e.g., assistant switched off), strong distraction (e.g., radio very loud)
- low risk: slight distraction (e.g., ventilation at full strength)

The questions are part of the decision tree for the determining the optimal implementation of the function (see page 22). The implementation is made up of the *Directness* and *HMI Type*:

1) Directness: This describes the number of operation steps required. The fewer operation steps required, the more directly the function can be accessed. The following operation actions can be distinguished depending on the modality:

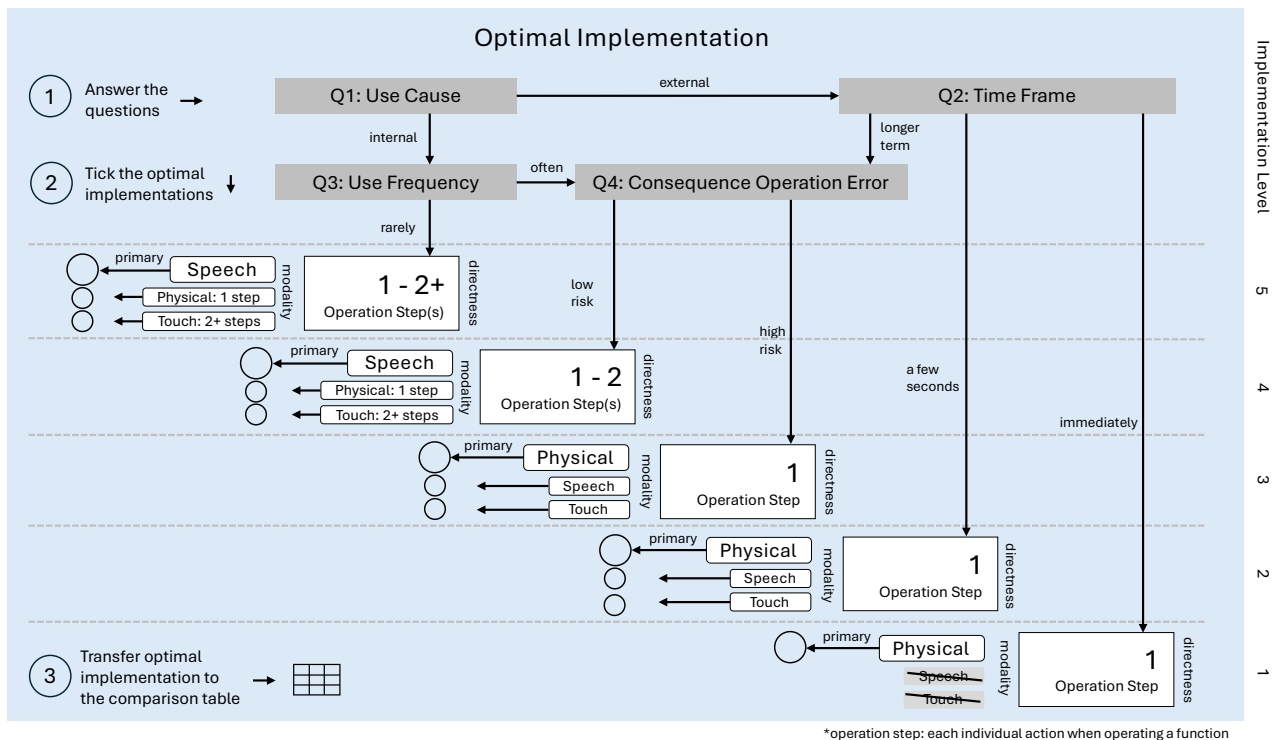
Physical Controls: pressing, turning, sliding, tilting; **Touch-based HMI:** tapping, swiping, zooming; **Speech-based HMI:** commands, parts of a dialog

Operation step: Each individual action to control a function. The actions must be counted individually and not be grouped together. For example, the first click on a button in the touch display, whereupon a pop-up menu appears, and the subsequent selection of an entry using a button count as two operating actions. It should also be noted whether the operation is "step-by-step" or continuous. If, for example, rotary switches with individual grid points are used to select between different points of a list, each grid step is a single operating step. If, on the other hand, the volume is adjusted with a continuous action using rotary switches with grid steps, this is seen as a singular operating action.

2) HMI types: The HMI types differ in their main modality used for the interaction with the driver (see definitions in the assessment scope). The bold and larger implementation placed at the very top within each level of the decision tree is the *primary modality* and is always required on implementation levels 1 – 3. The implementations listed below the primary modality are called *secondary modalities*. On level 1, no secondary modality is allowed; on levels 2 and 3, the bold secondary modality is the recommended secondary modality. On levels 4 and 5, the secondary modality must be suitable for the function's complexity, and the primary modality is not necessarily needed.

Now, use the decision tree (see next page) to determine the optimal implementation. Then, use the comparison table to compare the actual implementation against the actual implementation.

1. **Follow the decision path:** Determine the optimal implementation by running through the decision tree's paths. Any path will end at one of the five implementation levels. Each implementation level is made up of the number of operating steps ("directness") and the primary modality (written in a bold and larger font, at the very top) as well as other possible, secondary modalities (written in smaller letters, below) or not permitted modalities (crossed out). Tick the primary modality and, if possible, the secondary modalities.
2. **Transfer the optimal implementation into the comparison table:** For the ideal implementation, transfer the number of operation steps and primary and secondary modalities into the comparison table. Also transfer modalities which are not permitted into the table and cross them out, if applicable (level 1).
3. **Insert the actual implementation into the comparison table:** For the actual implementation, add the number of operation steps of the primary modality if used by the actual function implementation. Next, add secondary modalities, if applicable.
4. **Compare:** The comparison is made in three steps, and the outcome differs between the implementation levels. The colored columns of the table correspond to the ratings of optimal (green), suboptimal (blue) and unacceptable (red).
 - **Step 1:** Compare the operation steps.
 - Levels 1 - 4: more operation steps are unacceptable.
 - Levels 4 & 5: The same number of operation steps is optimal, less is deemed suboptimal.
 - **Step 2:** Compare the primary modality.
 - Levels 1 - 3: The primary modality must be used, otherwise this is rated as unacceptable.
 - Levels 4 & 5: The primary modality should be used. If the primary modality is not used but a secondary modality is used this is rated as suboptimal (see next step).
 - **Step 3:** Compare the secondary modalities.
 - Level 1: If an unpermitted secondary modality is used, this is rated as unacceptable.
 - Level 4 & 5: If the primary modality was not used but a suitable (for the task's complexity) secondary modality is used, this is rated as suboptimal. If an unsuitable secondary modality is used, this is rated as unacceptable.



Decision tree for determining the optimal implementation (modality/directness of access) regarding use context.

Comparison Table: Optimal and Actual Implementation					
Implementation	Insert: Optimal	Insert: Actual	Compare: Actual vs. optimal		
Step 1: Directness			↓ optimal	↓ suboptimal	↓ unacceptable
Number of Operation steps	_____	_____	□ Same number of steps	□ Less steps (L4 & 5)	□ more than 1 step (L1 - 3), more than 2 steps (L4)
Step 2: Primary Modality					
Primary modality	_____	_____	□ used	□ not used but... (L 4 & 5)	□ not used (L 1 - 3)
Step 3: Secondary Modality					
Secondary modality	_____	_____	□ ...suitable secondary modality used (L 4 & 5)	□ not allowed (L1)	□ ...unsuitable secondary modality used (L4 & 5)

Analysis

Module II a – Optimal HMI implementation

After determining the optimal implementation and checking whether the actual implementation complies with it, please fill out the following table to sum up the results. If the evaluation resulted in a “optimal” or “suboptimal”, you may continue with the assessment in part b of module II. If the actual implementation is deemed “unacceptable”, the assessment of the function has to be stopped and its implementation needs to be revised. You may, however, still continue with the assessment of other functions.

Function: _____	optimal	suboptimal	unacceptable
Comparison Result: Optimal vs. Actual Implementation			
Rating	☐	☐	☐
Reason for an “unacceptable” or “suboptimal”: _____ _____			

Transfer the results to the corresponding section of the evaluation sheet.

Module II b – HMI Function Interaction

Please indicate which HMI type(s) are used by the function that you are currently assessing:

Function: _____	
HMI Type	
Physical Controls	☐
Touch-Based HMI	☐
Speech-Based HMI	☐

For each HMI type used for controlling the function, please check whether and how the criteria (items) listed below apply. If a rating of “unacceptable” is given for any item, please do not continue with the assessment. Here, the HMI type should be revised to ensure an acceptable operation level. You may, however, continue with the assessment of another HMI type if is used for the operation of the function.

Repeat module II b for every function you would like to assess.

Module II b – Physical Controls

The following criteria must be checked for the HMI type. With each item, the deviation from an ideal design is determined. The sum (score on the rating scale multiplied with the weighting factor) of all items results in an HMI operation score. The items most relevant regarding driver distraction are listed first.

M2-PC-1 – Susceptibility to Operation Errors				
The operation of the function is not prone to errors.				
Not error-prone. The control element offers an optimal operating resistance and is placed at an optimal position.	Somewhat error-prone. The control element offers a good operating resistance and is easy to reach.	Moderately error-prone. The control element offers a suboptimal operating resistance (too light or heavy) or is difficult to reach.	Rather error-prone. The control element offers a suboptimal operating resistance (too light or heavy) and is difficult to reach.	Extremely error-prone. The control element offers a poor operating resistance (much too light or heavy) and is difficult to reach.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 3)	+ (0.05 × 3)	= 1.45
Explanation	The interaction with the control element should be carried out in a way that allows for the successful operation of the function.			
Testing	Operate the function that is controlled via the control element. Pay attention to the operating resistance and the position where the element is placed in the cockpit. Consider whether you would have problems operating the function on an uneven road.			
References	Naujoks et al. (2019)			

M2-PC-2 – Expectation Conformity				
The control element is positioned where drivers would usually expect to find it.				
Fully in line with expectations.	Slight deviation, but still conform to expectations.	Moderate deviation, but still conform to expectations.	Other position where it only occurs sometimes.	Not conform to expectations.
0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 1)	+ (0.05 × 5.2)	= 1.36
Explanation	The control element should be placed in an area where a driver would normally expect to find it.			
Testing	Check the position of the control panel and assess the extent to which this is in line with expectations.			
References	ISO 15005:2017 & ISO 3958:1996			

M2-PC-3 – Correction of Operation Errors				
In the event of an operation error, the operation can be undone or canceled.				
Yes, with one operation step, immediately.	Yes, with one operation step, very quickly.	Yes, with multiple operation steps, quite quickly.	Yes, with multiple operating steps, moderately fast.	No and/or only over a longer period of time.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 1)	+ (0.05 × 1.5)	= 1.18
Explanation	If the function is erroneously adjusted, switched on or off, it must be possible to reverse the operating action quickly.			
Testing	Switch the function on and, if possible, check for each setting how quickly you can return the function to the previous or a safe state. Pay attention to the time and the number of operation steps required.			
References	Ma & Gong (2024) Naujoks et al. (2019)			

M2-PC-4 – Risk of Operation Errors				
There is no safety risk if the function is operated incorrectly.				
No safety risk.	Safety risk: weak distraction (visual or acoustic).	Safety risk: strong distraction (visual or acoustic).	Safety risk for others.	Extreme safety risk: Intervention in longitudinal and lateral vehicle movement.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓		
Weighting	1.0	+ (0.1 × 1)	+ (0.05 × 0)	= 1.1
Explanation	In the event of an operation error, the function is either unintentionally adjusted, switched on, or switched off. This should not result in a safety risk.			
Testing	Switch the function on and off. If possible, switch the function to its maximum and minimum setting and check to what extent safety risks arise. A weaker group of safety risks is the distraction of the driver (e.g., radio very loud, ventilation fully turned up). The next more serious risk is the distraction of others (e.g., blinding others with high beams). Interventions in the lateral and longitudinal movement of the vehicle are unacceptable.			
References	Visual-Manual Guideline Criteria Summary – System Response			

M2-PC-5 – Additional Feedback				
The operation of the function is accompanied by clear feedback. This supplements the feedback from the function itself, if available.				
Yes, haptic, acoustic and visual feedback.	Yes, haptic & acoustic feedback.	Yes, haptic & visual feedback.	Not haptic, only acoustic and/or visual feedback.	No additional feedback.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects			Operation Quality ↓	
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 1.5)	= 1.07
Explanation	The operation of the function is accompanied by haptic feedback. Ideally, this feedback is made even more distinct by using other feedback channels (separately via loudspeakers and displays). If the function itself provides feedback (e.g., ventilation), it is enhanced by additional feedback.			
Testing	Use the control panel to switch the function on, off, or adjust it. Pay attention to which extent feedback occurs. Haptic feedback from the controls element itself (e.g., different position, detent steps). Acoustic or visual feedback from the function itself. Acoustic feedback via loudspeakers. Visual feedback via displays.			
References	Visual-Manual Guideline Criteria Summary – Interaction ISO 15005:2017			

M2-PC-6 – Comprehensibility				
The control element is labeled with distinct letters and symbols so that it does not take much time to look at it.				
Yes, distinct symbols and text.	Yes, only distinct symbols, no text.	Yes, only distinct text, no symbols.	Text and/or symbols not distinct.	No labeling.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0)	= 1
Explanation	The control element is labeled with symbols or short text (e.g., "AC") to minimize glance times.			
Testing	Look at the control panel and assess whether the symbols and/or text used are clear and distinct. This is about the meaning of the symbols and text, but not about the legibility of the labeling, which has already been assessed in module I.			
References	Visual-Manual Guideline Criteria Summary – Information presentation ISO 15005:2017 Naujoks et al. (2019)			

Module II b – Touch-Based HMI

The following criteria must be checked for the HMI type. With each item, the deviation from an ideal design is determined. The sum (score on the rating scale multiplied with the weighting factor) of all items results in an HMI operation score. The items most relevant regarding driver distraction are listed first.

M2-TB-1 – Susceptibility to Operation Errors				
The operation of the function is not prone to errors.				
Not error-prone. The operation logic is clear. The buttons are appropriate for the size and sensitivity of the touchscreen.	Somewhat error-prone. The operation logic is not clear. The buttons are appropriate for the size and sensitivity of the touchscreen.	Moderately error-prone. The operation logic is clear. The buttons are inappropriate for the size and sensitivity of the touchscreen.	Rather error-prone. The operation logic is not clear. The buttons are inappropriate for the size and sensitivity of the touchscreen.	Extremely error-prone. The operation logic is confusing. The buttons are completely inappropriate for the size and sensitivity of the touchscreen.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 4)	+ (0.05 × 4.5)	= 1.62
Explanation	The interaction with the touch interface should work in a way that allows for the successful operation of the function without any errors.			
Testing	Operate the function via the touch interface. Pay attention to the operation logic (path from the home screen, grouping in submenus, etc.) and evaluate if the operation elements (buttons, sliders) are appropriate regarding the size and sensitivity of the touchscreen.			
References	Grahn (2020) Liu et al. (2022) Naujoks et al. (2019)			

M2-TB-2 – Expectation Conformity				
The texts and symbols of the function follow familiar conventions. The operation logic is similar to physical controls.				
Texts and symbols follow familiar conventions completely, the operation logic is analogous to physically controlled systems.	Texts and symbols follow familiar conventions largely, the operation logic is almost analogous to physically controlled systems.	Texts and symbols follow familiar conventions only partly, the operation logic deviates somewhat from physically controlled systems.	Texts and symbols do not follow familiar conventions, the operation logic differs greatly from physically controlled systems.	Texts and symbols contradict familiar conventions, the operation logic differs greatly from physically controlled systems.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 4)	+ (0.05 × 4.5)	= 1.62
Explanation	The function's touch interface uses symbols and texts that correspond to conventions found in other function implementations (physical controls). Furthermore, the operation logic corresponds to already established implementations of the function.			
Testing	Look at the function when it is switched on and off and, if possible, in every setting. Assess whether the texts and symbols used are understandable for a user with an average "automotive operating experience". Also check whether the operation logic is compatible with previous physically controlled systems (e.g., on/off via a button that then appears pressed; higher setting further to the right, etc.).			
References	Visual-Manual Guideline Criteria Summary – Information Presentation ISO 15005:2017 Naujoks et al. (2019)			

M2-TB-3 – Risk of Operation Errors				
There is no safety risk if the function is operated incorrectly.				
No safety risk.	Safety risk: weak distraction (visual or acoustic).	Safety risk: strong distraction (visual or acoustic).	Safety risk for others.	Extreme safety risk: Intervention in longitudinal and lateral vehicle movement.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 3)	+ (0.05 × 1.5)	= 1.38
Explanation	In the event of an operation error, the function is either unintentionally adjusted, switched on, or switched off. This should not result in a safety risk.			
Testing	Switch the function on and off. If possible, switch the function to its maximum and minimum setting and check to what extent safety risks arise. A weaker group of safety risks is the distraction of the driver (e.g., radio very loud, ventilation fully turned up). The next more serious risk is the distraction of others (e.g., blinding others with high beams). Interventions in the lateral and longitudinal movement of the vehicle are unacceptable.			
References	Visual-Manual Guideline Criteria Summary – System Response			

M2-TB-4 – Correction of Operation Errors				
In the event of an operation error, the operation can be undone or canceled.				
Yes, with one operation step, immediately.	Yes, with one operation step, very quickly.	Yes, with multiple operation steps, quite quickly.	Yes, with multiple operating actions, moderately fast.	No and/or only over a longer period of time.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 3)	+ (0.05 × 0.8)	= 1.34
Explanation	If the function is mistakenly adjusted, switched on or off, it must be possible to reverse the operation action quickly.			
Testing	Switch the function on and, if possible, check for each setting how quickly you can return the function to the previous or a safe state. Pay attention to the operation time and the number of operation steps required.			
References	Ma & Gong (2024) Naujoks et al. (2019)			

M2-TB-5 – Interaction Speed (only applicable to multiple-step operation)				
The driver determines the interaction speed at which the function is operated. The function does not continue automatically and does not generate any pressure for the continuation of the operation.				
Function waits indefinitely for driver input.	Function waits a long time for driver input, then returns to the initial state or gives further instructions.	Function waits a little for driver input, then returns to the initial state or gives additional instructions.	Function waits briefly for driver input, then returns to the initial state or gives additional instructions.	Function almost does not wait for driver input, quickly returns to the initial state and does not provide any further information.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 2)	+ (0.05 × 1.5)	= 1.27
Explanation	In the case of multiple-step operation, the function waits for the driver's input at each operation step. It does not continue automatically.			
Testing	Start a multiple-step operation process within the function. After each interaction step, test whether and for how long the function waits for driver input.			
References	Visual-Manual Guideline Criteria Summary – Interaction ISO 15005:2017			

M2-TB-6 – Additional Feedback				
The operation of the function is accompanied by clear feedback at every operation step (activation, deactivation, adjustment) and in the event of errors. This supplements the feedback from the function itself, if available.				
Acoustic and visual feedback for every operation step and in the event of errors.	Acoustic or visual feedback for every operation step and in the event of errors.	Acoustic and visual feedback only for operation steps or errors.	Only visual or acoustic feedback for operation steps or errors.	No additional feedback for operation steps or errors.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 1)	+ (0.05 × 3)	= 1.25
Explanation	Each operating step and error of the function is accompanied by clear feedback. If the function itself provides feedback (e.g., ventilation), this is supplemented by additional feedback.			
Testing	Switch the function on, off and, if possible, adjust it. Check whether visual and/or acoustic feedback is/are used for each operation step. Check to what extent error feedback is given, e.g., when a blocked option is selected or when the end of a list is reached.			
References	Visual-Manual Guideline Criteria Summary – Interaction ISO 15005:2017			

M2-TB-7 – Legibility				
The text, symbols, and images used in the function menu are easy to read and high-contrast.				
Text, images, or symbols are optimally sized and high-contrast.	Text, images, or symbols are slightly too small or low-contrast.	Text, images, or symbols are too small or low-contrast.	Text, images, or symbols are very small or low-contrast.	Text, images, or symbols are much too small or without contrast.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects			Operation Quality ↓	
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0.8)	= 1.04
Explanation	The texts, pictures, and/or symbols displayed in the function menu are easy to read in terms of their size and contrast.			
Testing	Look at the function when it is switched on and off and, if possible, in every setting. Assess whether the elements used (text, images, symbols) are easy to read in every setting.			
References	Visual-Manual Guideline Criteria Summary – Information Presentation Naujoks et al. (2019) ISO 15008:2017			

M2-TB-8 – In-process Abortion (only applicable to multiple-step operation)				
A multiple-step operation that has been started can be canceled during the process.				
The function can be canceled at any point.	The function can be canceled at almost every point.	The function can be canceled at a few points.	The function can be canceled at several points or can only be canceled in a cumbersome manner.	The function cannot be canceled.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects				
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0)	= 1
Explanation	In the case of multiple-step operation, the operation process can be stopped at any point.			
Testing	Start a multiple-step operating process within the function (e.g., a navigation or phone function). At any operation point, test whether you can cancel the function operation. Aborting can be done via a separate button, by calling up the home menu, or by calling up another function menu. A cancellation is considered cumbersome if it requires several operating steps (e.g., if "back" has to be pressed several times).			
References	Visual-Manual Guideline Criteria Summary – Interaction ISO 15005:2017			

Bonus part: The following criteria are an optional part of the HMI type assessment. These are deemed desirable but not necessary. If they do not apply, skip the items. If they apply, bonus points are collected which are used to lower the HMI operation score.

M2-TB-9 – Shortcut				
The operation of the function can be shortened by skipping individual operation steps.				
Function can be operated via a shortcut with a single operation step.	Function can be operated via a shortcut with fewer operation steps.	Function accessible via physical controls.	Function accessible via speech-based HMI.	Function does not offer a shortcut.
-4 ☐	-3 ☐	-2 ☐	-1 ☐	0 ☐
Affects				
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0)	= 1
Explanation	The operation of the function can be shortened by using a shortcut (touch-based HMI, physical controls, and/or voice-based HMI) which skips operation steps.			
Testing	First try to create a shortcut for the function on the home screen. Then check whether you can operate the function so that the final operation step can be accessed directly via this shortcut. Also check to what extent the function can also be accessed via physical controls and the voice-based HMI.			
References	Ebel et al. (2023)			

Module II b – Speech-Based HMI

The following criteria must be checked for the HMI type. With each item, the deviation from an ideal design is determined. The sum (score on the rating scale multiplied with the weighting factor) of all items results in an HMI operation score. The items most relevant regarding driver distraction are listed first.

M2-SB-1 – Fallback				
The function can also be operated using a modality (HMI type) other than speech. The fallback uses the optimal implementation according to module II a.				
Yes, via another optimal modality (primary or upper secondary modality).	Yes, via another optimal modality and suboptimal modality.	Yes, via suboptimal modalities.	Yes, via suboptimal and unacceptable modalities.	No, no fallback.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 6)	+ (0.05 × 3.8)	= 1.79
Explanation	The function can be operated not only via voice, but also via another modality (physical controls and/or touch-based HMI). This ensures operability even if there are errors in the speech-based HMI.			
Testing	Check to what extent the function accessible via the speech-based HMI can also be accessed via other interfaces (physical controls and/or touch-based HMI). Also check whether the implementation is optimal in accordance with module II a.			
References	Hua & Ng (2010)			

M2-SB-2 – Correction of Operation Errors				
In the event of incorrect operation, it can be undone or canceled.				
Yes, with a single command and via barge-in, immediately. Cancellation is confirmed verbally.	Yes, with several commands and via barge-in. Cancellation is confirmed verbally.	Yes, with a single command, but only after the speech-based HMI has finished its dialog. Cancellation is confirmed verbally.	Yes, with several commands but only after the speech-based HMI has finished its dialog. Cancellation is not confirmed verbally.	No, it is not possible to cancel via voice.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 5)	+ (0.05 × 3)	= 1.65
Explanation	If the function is mistakenly adjusted, switched on, or off via the speech-based HMI, it must be possible to reverse the operation action quickly.			
Testing	Activate or adjust the function using the speech-based HMI. Try to interrupt the interaction by barging-in with a command such as "Cancel!". If this is not possible, try canceling the interaction after the assistant has finished speaking. If necessary, test other commands to cancel the function.			
References	Speech-Based Standard Criteria Summary – System Response ISO 15005:2017			

M2-SB-3 – Susceptibility to Operation Errors				
The speech-based operation of the function is not prone to errors.				
Not error-prone. The speech-based HMI works reliably. The target function can always be operated.	Somewhat error-prone. The speech-based HMI almost always works reliably. Instead of activating something else, it first asks how to proceed.	Moderately error-prone. The speech-based HMI works unreliably from time to time. Other functions are activated without asking.	Rather error-prone. The speech-based HMI is unreliable. Other functions are activated without asking.	Extremely error-prone. The speech-based HMI is very unreliable. Other functions are constantly activated without asking.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 4)	+ (0.05 × 3.8)	= 1.59
Explanation	The interaction with the speech-based HMI should work in a way that allows for the successful operation of the function without any errors.			
Testing	Operate the function using the speech-based HMI. Make sure that you can reliably switch the target function on, off and, if possible, adjust it. Try varying your expressions to provoke a misunderstanding (e.g., speaking quietly, mumbling, alternative terms) and check whether the speech-based HMI asks if it understood you correctly or if it directly activates another function.			
References	Naujoks et al. (2019) Reichel et al. (2014)			

M2-SB-5 – Risk of Operation Errors				
There is no safety risk if the function is operated incorrectly.				
No safety risk.	Safety risk: weak distraction (visual or acoustic).	Safety risk: strong distraction (visual or acoustic).	Safety risk for others.	Extreme safety risk: Intervention in longitudinal and lateral vehicle movement.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 2)	+ (0.05 × 3.8)	= 1.39
Explanation	In the event of an operation error, the function is either unintentionally adjusted, switched on or switched off. This should not result in a safety risk.			
Testing	Switch the function on and off. If possible, switch the function to its maximum and minimum setting and check to what extent safety risks arise. A weaker group of safety risks is the distraction of the driver (e.g., radio very loud, ventilation fully turned up). The next more serious risk is the distraction of others (e.g., blinding others with high beams). Interventions in the lateral and longitudinal movement of the vehicle are unacceptable.			
References	Speech-Based Standard Criteria Summary – System Response Visual-Manual Guideline Criteria Summary – System Response			

M2-SB-6 – Additional Feedback				
Feedback as part of the function operation is provided verbally (speech) and acoustically (chimes). Visual feedback is compact and not distracting. This complements the feedback from the function itself, if available.				
Short and concise speech output and chimes. Visual feedback short and only used for multi-step interaction.	Rather long speech output or no chimes. Visual feedback not as short as possible.	Rather long speech output and no chimes. Visual feedback too long.	Rather long or incomprehensible speech output. Visual feedback too long.	Only visual feedback, which is very long.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 1)	+ (0.05 × 6)	= 1.4
Explanation	The function operation using the speech-based HMI is accompanied by acoustic feedback. Ideally, chimes are used, and the operation is confirmed via speech output. Visual feedback should be brief and only used for multiple-step functions.			
Testing	Activate or adjust the function using the speech-based HMI. When doing so, pay attention to the acoustic (speech and/or chimes) and visual feedback (in the instrument cluster, on the touch display or on physical switches) that is used.			
References	Speech-Based Standard Criteria Summary – Information Presentation Hua & Ng (2010)			

M2-SB-7 – Diverse Commands				
The system uses a naturalistic language model. The function can be activated using various common expressions.				
Many common expressions for the function are recognized.	Several common expressions for the function are recognized.	Some common expressions for the function are recognized.	Multiple expressions possible, but they must always contain a keyword.	Only a specific expression with a keyword is recognized for the function.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 1)	+ (0.05 × 1.5)	= 1.18
Explanation	Various common expressions can be used to activate and/or control the function.			
Testing	Try using various commands to switch the function on, off and, if possible, adjust it. Vary the expressions you use for this. Expressions for controlling the seat heating, for example, could be "Switch on the seat heating", "I'm cold", etc. Pay attention to whether the system insists on the use of certain key terms (e.g., "seat heating" must always be mentioned to control the seat heating).			
References	Speech-Based Standard Criteria Summary – Interaction ISO 15005:2017 Eriksson et al. (2014)			

M2-SB-8 – In-process Abortion (only applicable to multiple-step operation)				
A multiple-step operation can be canceled mid-process by informing the speech-based HMI of the cancellation request.				
The function can be canceled at any point.	The function can be canceled at almost every point.	The function can be canceled at a few points.	The function can be canceled at several points or can only be canceled in a cumbersome manner.	The function cannot be canceled.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects				
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0)	= 1
Explanation	In the case of multiple-step operation, the operation process can be stopped at any point			
Testing	Start a multiple-step operation process within the function. Within each interaction step, test whether you can cancel the operation of the function. A cancellation may happen via a separate command, e.g., "Cancel!" or by cancelling the speech-based HMI altogether. The function may also be canceled by starting the operation of another function. Finally, the cancelling can also be carried out via another modality (touch-based HMI or physical controls). Canceling is inconvenient if several dialogs with the speech-based HMI are necessary.			
References	Speech-Based Standard Criteria Summary – Interaction			

M2-SB-9 – Interaction Speed (only applicable to multiple-step operation)				
The driver determines the interaction speed at which the function is operated. The speech-based HMI does not continue automatically and does not generate any pressure for the continuation of the operation.				
The speech-based HMI waits indefinitely for driver input.	The speech-based HMI waits a long time for driver input, first gives instructions and then cancels the operation.	The speech-based HMI waits briefly for driver input, first gives instructions and then cancels the operation.	The speech-based HMI waits briefly for driver input, gives no instructions and cancels the operation.	The speech-based HMI does not wait for driver input, gives no instructions and cancels the operation.
0 ☐	1 ☐	2 ☐	3 ☐	unacceptable ☐
Affects				
Weighting	1.0	+ (0.1 × 0)	+ (0.05 × 0)	= 1
Explanation	In the case of multiple-step operation, the speech-based HMI waits for the driver's input at each operation step. It does not continue automatically.			
Testing	Start a multiple-step operation process within the function. After each voice output, test whether and for how long the speech-based HMI waits for driver input.			
References	Speech-Based Standard Criteria Summary – Interaction			

Bonus part: The following criteria are an optional part of the HMI type assessment. These are deemed desirable but not necessary. If they do not apply, skip the items. If they apply, bonus points are collected which are used to lower the HMI operation score.

M2-SB-4 – Shortcut				
The function can be activated with a single short command. The speech-based HMI provides information on the ideal command during the interaction.				
Yes, short own ideal command possible.	Yes, short predefined ideal command, speech-based HMI gives hints.	Yes, medium-long predefined ideal command, speech-based HMI gives hints.	Yes, long predefined ideal command, no instructions from the speech-based HMI.	No, no shortened operation possible.
-4 ☐	-3 ☐	-2 ☐	-1 ☐	0 ☐
Affects		Driving Performance ↓	Operation Quality ↓	
Weighting	1.0	+ (0.1 × 4)	+ (0.05 × 4.5)	= 1.62
Explanation	The operation of the function via speech-based HMI can be shortened by using one's own or a predefined ideal voice command. The speech-based HMI should provide information about the ideal command.			
Testing	Start a voice-controlled operation of the function in which you try to operate it in a complicated way (e.g., many individual steps). Pay attention to whether the speech-based HMI gives you tips for an ideal command. You can also ask the speech-based HMI about the function and check whether it provides operation commands. Finally, you should check whether you can assign your own commands - either directly via the speech-based HMI or via the settings in the central touch display.			
References	Hua & Ng (2010) Response Ebel et al. (2023)			

Analysis

Module II b – HMI Function Interaction

After answering all items of the HMI types used by the function you want to assess, please sum up the ratings to calculate the HMI operation score. If your function does not use an HMI type, check the “not used” box. If an item scored an “unacceptable” check the “unacceptable” box.

Function: _____				
HMI Type	Not used?	Unacceptable?	HMI Oeration Score	Point Range
Physical Controls	☐	☐		0 (best) → 22 (worst)
Touch-Based HMI	☐	☐		-4 (best) → 32 (worst)
Speech-Based HMI	☐	☐		-7 (best) → 33 (worst)

Transfer the results to the evaluation sheet.

Evaluation Sheet

Assessment Scope

Function	
Function Group	

Module I – Shared HMI Features

HMI Type	Not used?	Unacceptable?	HMI Op. Score	Point Range
Physical Controls	☐	☐		0 (best) → 23 (worst)
Touch-Based HMI	☐	☐		0 (best) → 46 (worst)
Speech-Based HMI	☐	☐		0 (best) → 27 (worst)

STOP

Recommendations:

- “unacceptable”: Revise the associated interface characteristic
- High HMI Operation Score: Revise those items whose rating is higher than 2

check only once

Module II a – Optimal HMI implementation

Function: _____	optimal	suboptimal	unacceptable
Comparison Result			
Rating	☐	☐	☐
STOP Reason for an “unacceptable” or “suboptimal”: _____ _____			

Recommendations:

- “unacceptable”: Provide a suitable, optimal implementation for the function
- “suboptimal”: Improve the function implementation

check for every single function

Module II b – HMI Function Interaction

HMI Type	Not used?	Unacceptable?	HMI Op. Score	Point Range
Physical Controls	☐	☐		0 (best) → 22 (worst)
Touch-Based HMI	☐	☐		-4 (best) → 32 (worst)
Speech-Based HMI	☐	☐		-7 (best) → 33 (worst)

STOP

Recommendations:

- “unacceptable”: Revise the associated interface characteristic
- High HMI Operation Score: Revise those items whose rating is higher than 2

Appendix B: Operationalization of the HMI Checklist's Interface Design Aspects

For the validation of the checklist, a driving simulator experiment was conducted. The impact of the individual design aspects (items) per HMI type on the distraction dynamics while driving was examined. To capture the effect on distraction, driving performance and operation quality were measured.

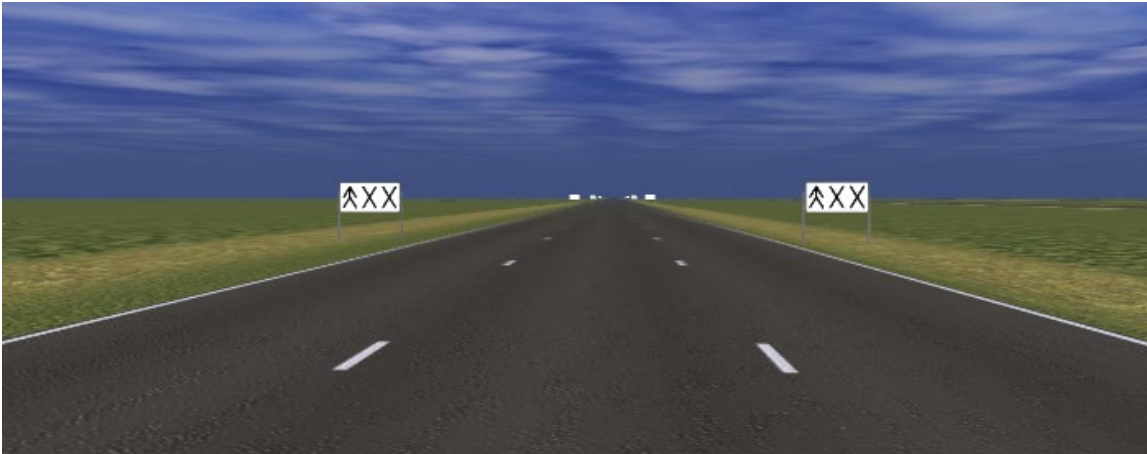
To adequately understand the results, the operationalization of the checklist parameters as well as the driving task and operation tasks must be considered. The operationalization describes the specific implementation of the driving task, function operation tasks, and checklist's design aspects in the study. Therefore, it refers to how the distraction, influenced by various aspects of an interface, was made measurable.

The operationalization establishes a very specific context of driving and function operation. In real life, driving maneuvers are diverse, as are the operation tasks and the manufacturer-specific implementation. Therefore, the study's results provide only an indication toward relevant aspects for real vehicle function operation; in any case, they do not claim absolute validity. The transfer of the checklist parameters to further contexts must be discussed carefully.

In this document, the driving task and operation tasks (simple and complex) are first described, before the operationalization of the checklist parameters is explained per HMI type.

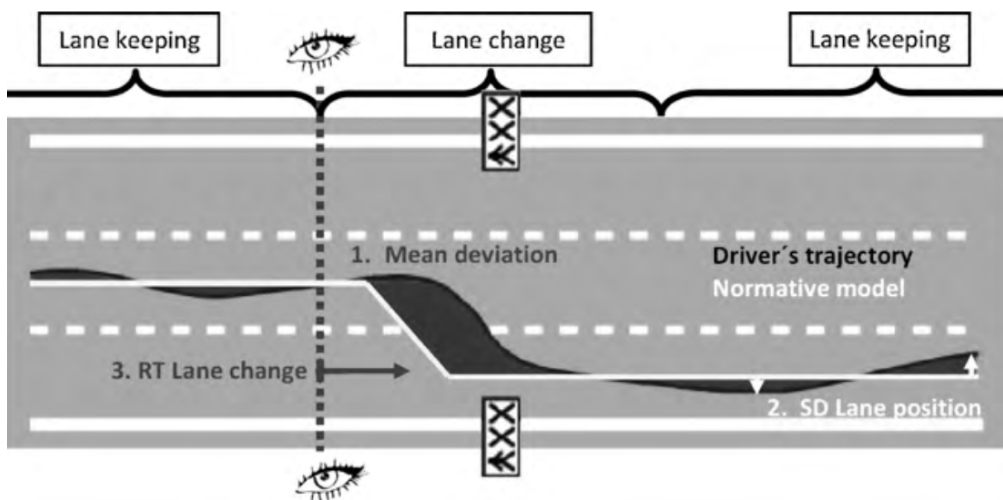
Beyond the operationalization, examples are provided for each design principle to make the transfer to real-life function operation easier.

Driving Task: Lane-Change Task



Driving environment in the Lane-Change Task.

In the Lane-Change Task (LCT), participants drive at a constant speed of 60 km/h along a three-lane road. Signs on the right and left side of the road, placed at intervals of 150 meters, indicate which lane should be changed to. The driving task consists of two subtasks: lane keeping and lane changing.



Driving tasks in the Lane-Change Task.

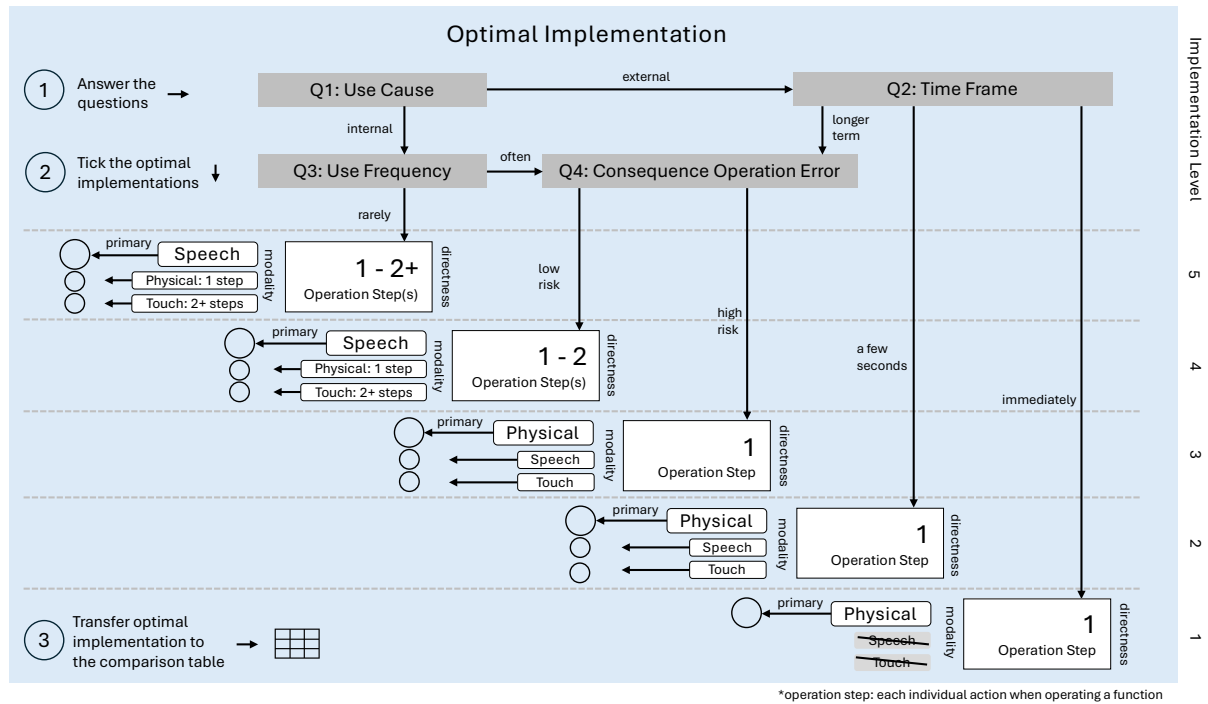
To measure driving performance, two aspects are considered: the deviation from the ideal trajectory and the reaction time between the lane change sign becoming visible and the first steering input. The deviation from the ideal trajectory represents the visual-manual distraction, while the reaction time serves as an indicator of cognitive distraction.

Although the LCT is well-suited for capturing distraction, it can be discussed to what extent the results can be transferred to real-life driving. For instance, there are no other traffic participants present on the road, and the infrastructure only consists of a three-lane, always- straight road. Complex driving maneuvers are therefore not part of the driving simulation.

Function Operation Tasks

In addition to validating the checklist's interface design aspects (items), a further objective of the study was to examine the effect of task complexity on driving performance and its interaction with HMI type (physical controls, touch-based, speech-based) as well as the checklist parameters.

For this purpose, two operation tasks were selected that correspond to the highest (5) and lowest (1) implementation levels of the decision tree (Module IIa): entering a phone numbers (complex task, level 5) and indicating (simple task, level 1) The levels of the decision tree indicate the acceptable task complexity (operation steps) and HMI type in relation to the use context.



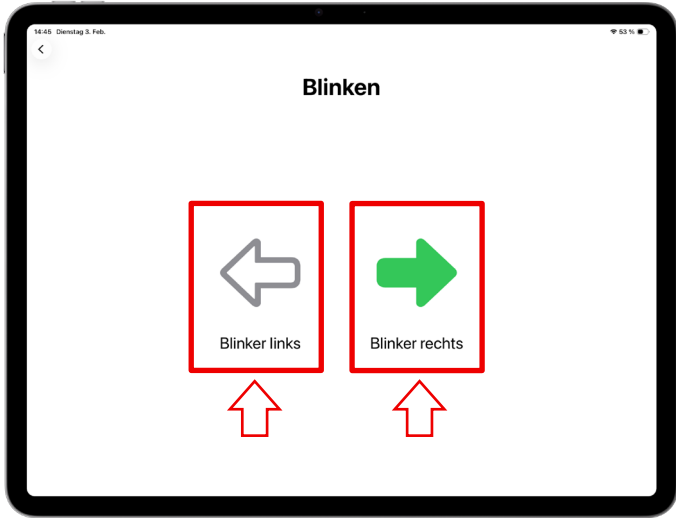
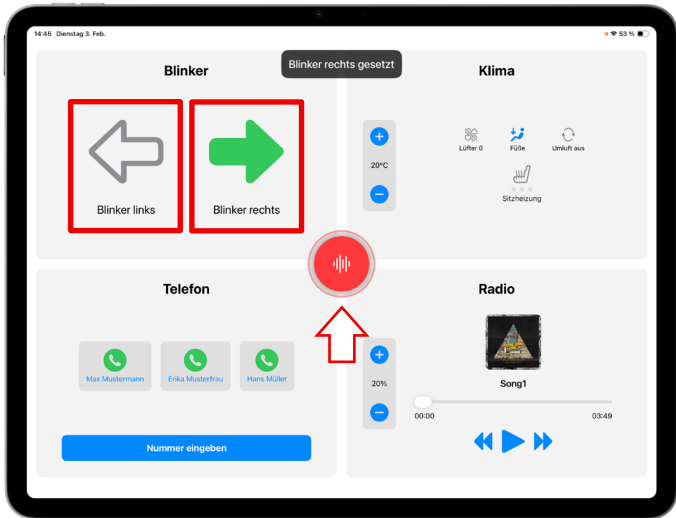
Decision tree (module IIa) from the HMI checklist.

Simple Task: Indicating

Indicating is an operation task at the first level of the decision tree. The uses context is therefore external, and the operation must be possible in a single step. In the experiment, the external use context arises from the lane change required in the LCT.

The operation was tested for each of the three HMI types:

HMI Type	Image	Operation Sequence
Physical Controls		Press steering wheel button, visual-acoustic feedback (arrow blinking & click sound)

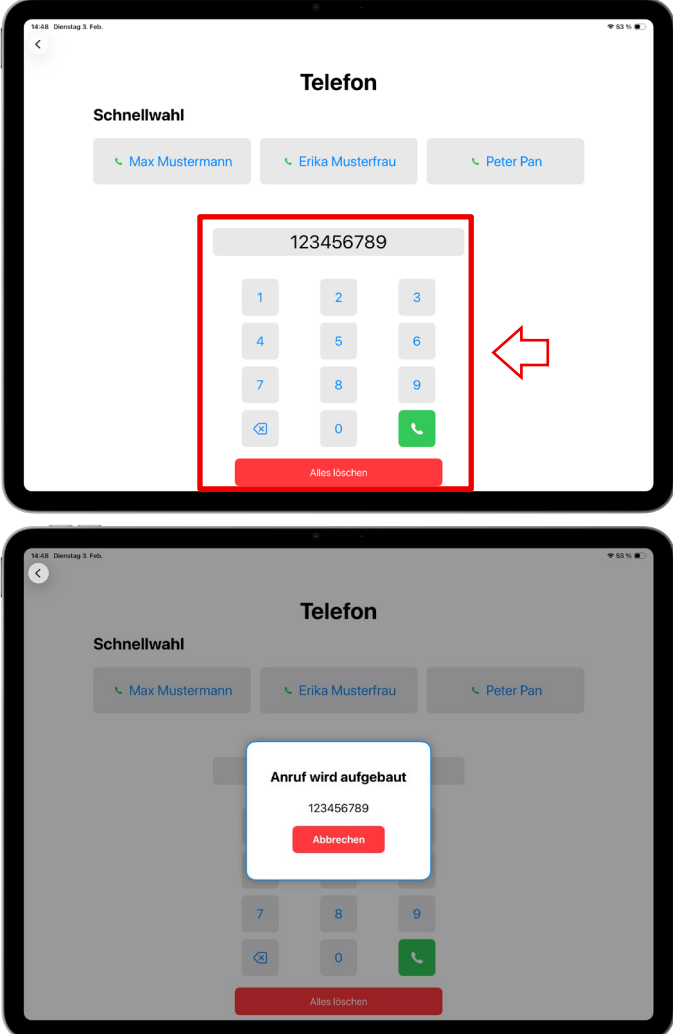
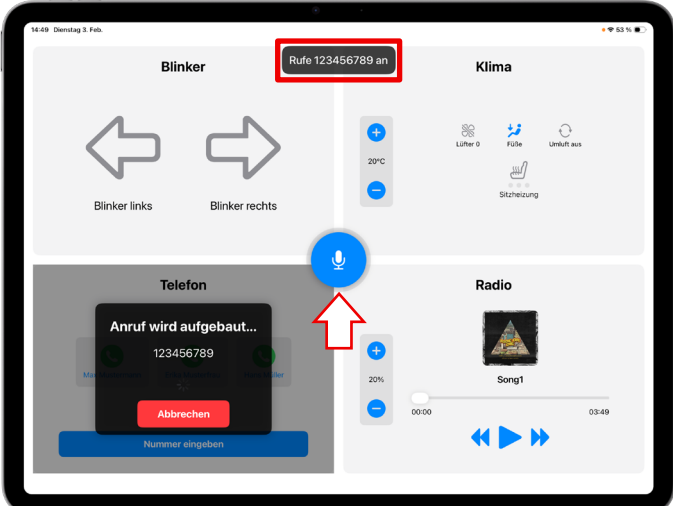
HMI Type	Image	Operation Sequence
Touch-Based		Press touch buttons, visual-acoustic feedback (arrow blinking & click sound)
Speech-Based		Voice command (without keyword), visual-acoustic feedback (arrow blinking, click sound & speech output)

Complex Task: Phone Number Dialing and Calling

Phone operation (entering and calling a phone number) is an operation task at the fifth level of the decision tree. The use context is therefore internal, and operation generally occurs only rarely and can require multiple steps. In the study, the phone numbers were provided in segments of three digits by the experimenter.

The operation was tested for each of the three HMI types:

HMI Type	Image	Operation Sequence
Physical Controls	 	Press steering wheel buttons, visual-acoustic feedback (call screen & dial tones)

HMI Type	Image	Operation Sequence
Touch-Based		Press touch buttons, visual-acoustic feedback (call screen & dial tones)
Speech-Based		Voice commands (without keyword), visual-acoustic feedback (call screen, dial tones & speech output)

Interface Design Aspects

The interface design aspects refer to properties of the HMI type affecting function implementation quality. The design aspects are assessed with the items of the checklist. In module I, these aspects describe characteristics that impact the operation quality of multiple functions. These include, for example, the size and positioning of a touchscreen. Every function operated via the touchscreen is equally affected by these aspects. On the other hand, module II b assesses aspects for each individual function, such as whether a function can be interrupted during the operation process or follows established operation sequences.

For the experiment, a generic interface was created for each HMI type (physical controls, touch-based HMI, speech-based HMI) where all parameters were implemented optimally (0). This version was used to set the reference distraction level; therefore, it is called the “reference version”. To test the impact of a design aspect’s violation, it was implemented in a suboptimal way (3). The remaining interface design aspect remained optimally implemented.

13 – Minimalism				
The touchscreen interface has a minimalist design and information is presented compactly. This minimizes visual search times and the associated distraction.				
All content is short and concise	Almost all content is short and concise	Content is sometimes too long.	Contents rather long	Contents too long
0	1	2	3	unacceptable
☐	☐	☐	☐	☐
optimal		suboptimal		

The design aspects and their implementation for all three HMI types in the study are described below. The first group of aspects refer to cross-functional aspects (module I). The second group of aspects is function-specific (module II b). For each design aspect, examples from automotive settings are provided to facilitate the transfer of the design principle.

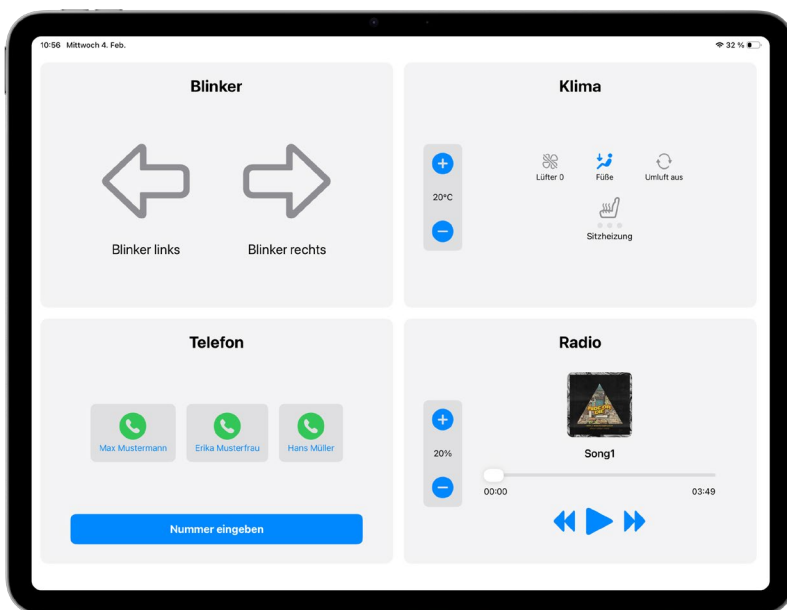
Note: Some design aspects were “bonus” items. Here, instead of measuring the impact of a suboptimal implementation, the implementation of the bonus design principle was tested. Also, some design aspects were measured only for the simple or complex task. Affected design aspects (items) are marked in case they are bonus items and/or were only measured for one of the tasks.

PHYSICAL CONTROLS

Reference Version

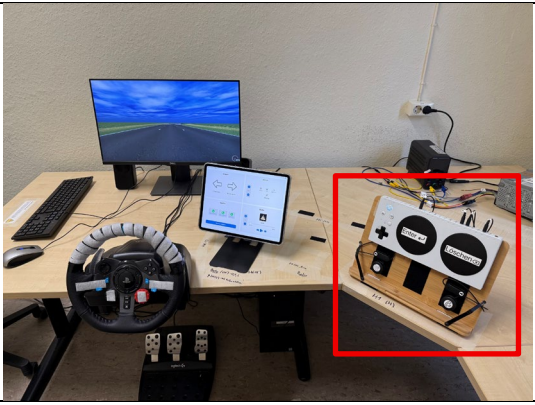
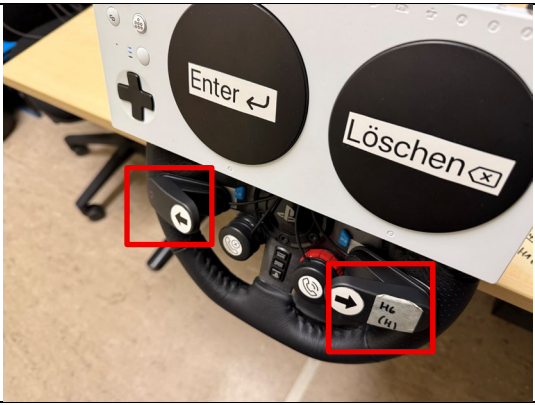


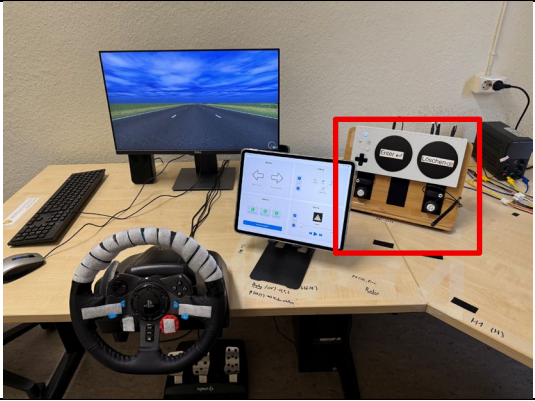
Experimental setup of the physically controlled HMI.

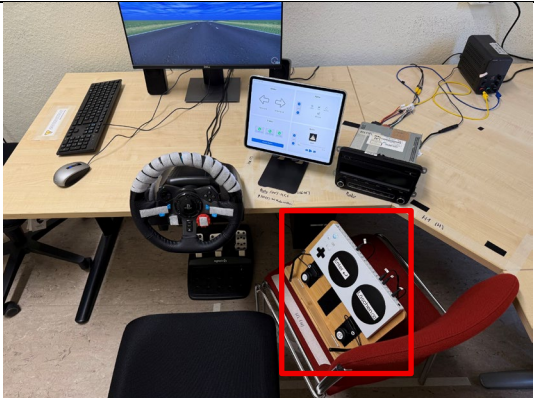
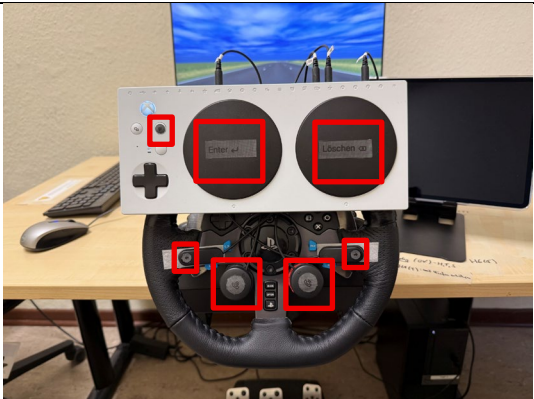
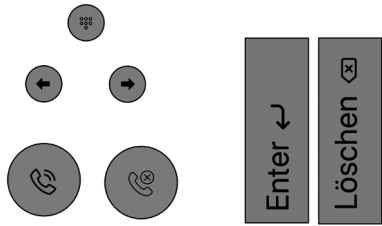


Software of the physically controlled HMI: Function tiles for indicating function activation status.

Module I – Shared HMI Features

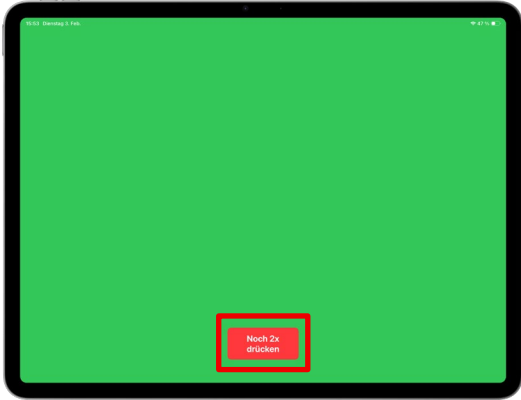
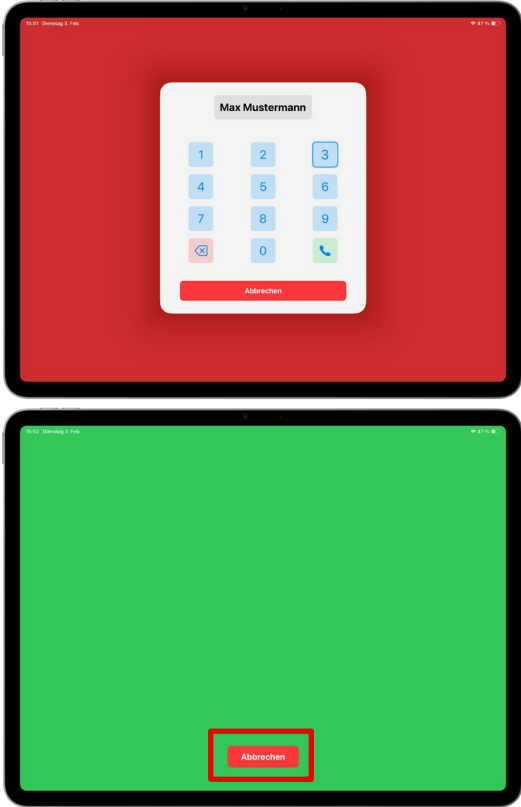
Item	Description	Optimal	Suboptimal
M1-PC-1	Viewing Distance to the Right	0° – 15°	45° – 60°
	Validation	All controls on steering wheel, tablet next to steering wheel	Controls attached to a holder on the right
			
	Example	• Cruise control, volume, media buttons on steering wheel • Turn signal, wiper levers on steering column • Horn in steering wheel center	• Climate control panel at far right of center stack • Window switches on driver door panel • Hazard light button placed far back on console
M1-PC-2	Blind Operation	Haptics when reaching, during operation, setting can be felt	Haptics only when reaching
	Validation	Controls on steering wheel with click and resistance	Trigger (no click sound and no resistance) used as control element on steering wheel
	Note: Indicator only		
	Example	• Rotary volume knob with detents (feeleable steps) • Turn signal lever with audible/tactile click • Heated seat button with raised texture and distinct position	• Capacitive buttons with no click feedback • Flat piano-black control strips • Smooth rocker switches with no tactile differentiation

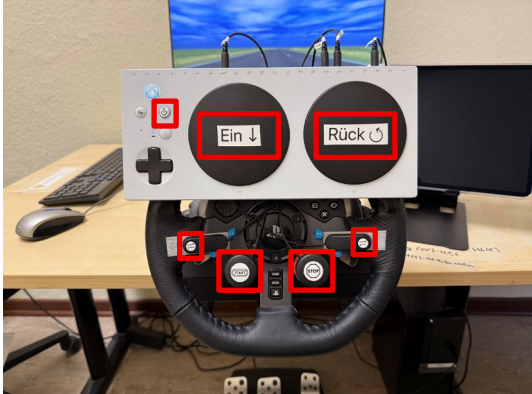
Item	Description	Optimal	Suboptimal
M1-PC-3	Glare	No glare, matte surface	Reflective surface
	Validation	No illumination of driver	Flashlight on the radio aimed toward the driver, so the driver is illuminated (brightness: level 1)
			
	Example	- Matte plastic buttons housed under dashboard overhang - Controls in recessed areas shielded from direct sunlight	- Chrome-ringed buttons on exposed center console - Glossy piano-black control surfaces - Polished metallic trim directly surrounding buttons
M1-PC-4	Reachability	Both hands on the steering wheel	One hand on the steering wheel, arm widely extended
	Validation	All controls on steering wheel, tablet next to steering wheel	Holder with controls placed far to the right so that the participant must stretch
			
	Example	- Cruise control, volume, media buttons on steering wheel - Turn signal, wiper levers on steering column - Horn in steering wheel center	- Rear window defroster button low on center stack - Hazard light button recessed far back on console - Seat adjustment buttons on side of seat

Item	Description	Optimal	Suboptimal
M1-PC-5	Viewing Distance Downwards	0° – 15°	30° – 45°
	Validation	All controls on steering wheel, tablet next to steering wheel	Control holder positioned downward, on chair
			
	Example	- Steering wheel controls - Buttons on upper center stack	- Controls at bottom of center stack - Controls on center console
M1-PC-6	Legibility	High-contrast, distinct, appropriate size	Low-contrast, ambiguous, small
	Validation	Good labeling, large, white background, black text	Light grey labeling, far too small
			 
	Example	- High-contrast backlit labels visible day and night - Large, distinct icons for AC, fan speed, temperature, etc.	- Small grey symbols on light grey background (no backlight) - Unlabeled multi-function buttons (symbol only, very small) - Labels invisible in daylight without backlight

Module II b – Function-Specific HMI Features

Item	Description	Optimal	Suboptimal
M2-PC-1	Susceptibility to Operation Errors	Resistance and position optimal	Resistance unfavorable and hard to reach
	Validation	Controls on the steering wheel	Controller and button placed loosely on the table
			
	Example	- Volume knob with firm detents (no accidental activation) - Turn signal lever with spring-return mechanism - E-call button with protective cover	- Feather-light touch buttons (triggered by brushing) - Rocker switch with no tactile differentiation between settings - Loose toggle switch with no defined end position
M2-PC-2	Expectation Conformity	Fully in line with expectations	Other position
	Validation	All control elements on steering wheel	Indicator: buttons on steering wheel swapped; Phone: Controller placed flat on the left side next to steering wheel
			
	Example	- Turn signal on left stalk, wipers on right stalk (left-hand drive) - Volume on left side of steering wheel - Gear shift in center console	- Turn signal as touchscreen button - Gear selector as buttons on dashboard - Horn relocated to spoke button instead of center pad

Item	Description	Optimal	Suboptimal
M2-PC-3	Correction of Operation Errors	One action, immediately	Multiple actions, slowly
	Validation	Direct cancellation possible (M2-PC-4)	Cancellation requires multiple steps (press physical cancel button three times)
			
	Example	- Turn signal: simple push back or tap to cancel - Volume: single rotation to previous level - AC: single button press to toggle off	- Function requiring multiple back-button presses to undo - Setting change that triggers a confirmation dialog before reverting - No undo, must scroll back to previous state manually
M2-PC-4	Risk of Operation Errors	No safety risk	Strong distraction
	Validation	Function works normally	Feedback becomes very salient – flashscreen (green: max. brightness, red: normal brightness) & error sound; direct cancellation possible by pressing physical cancel button
			

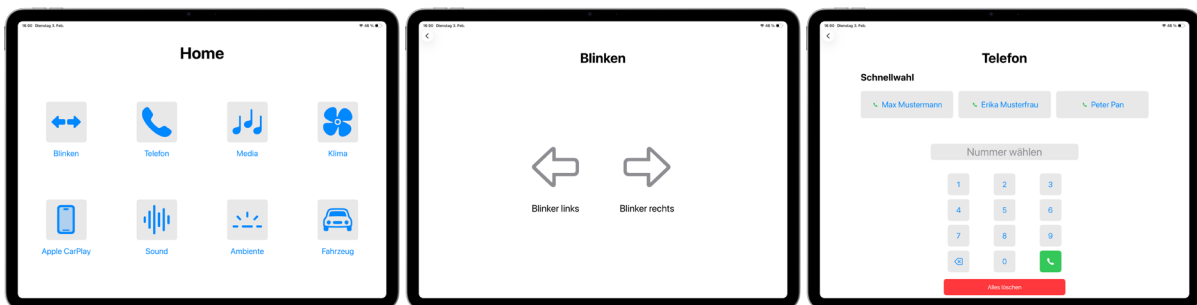
Item	Description	Optimal	Suboptimal
	Example	- Accidental seat heating activation: mild thermal effect only - Accidentally pressing rear wiper button: minor visual distraction	- Accidentally triggering full fan speed (sudden loud noise) - Accidentally activating entertainment function (strong visual distraction) - Accidentally disabling lane assist (safety risk)
M2-PC-5	Additional Feedback	Haptic, acoustic, visual	Only acoustic or visual
	Validation	App provides additional feedback	App is muted, no acoustic feedback
	Example	- Turn signal: mechanical click + dashboard indicator + stalk position - AC button: click sound + LED backlight change + display update - Volume knob: detent feedback + bar visualization on cluster + volume	- Silent capacitive button with only LED color change - Function activated with no acoustic or visual confirmation - Only the function itself provides feedback (no supplemental signal)
M2-PC-6	Comprehensibility	Symbols and text distinct	Text and/or symbols unclear
	Validation	Symbols on buttons matching those in the instructions	Different symbols (ambiguous text, geometric shapes) on the buttons
			
	Example	- Standard symbols for AC, fan, temperature, heated seat - Clear text labels next to buttons (e.g., "AC", "REAR") - Distinct icon shapes for neighboring controls	- Brand-specific abstract symbols without text (unrecognizable) - Multiple similar-looking icons in close proximity - Symbol meaning only revealed by long-press tooltip

TOUCH-BASED HMI

Reference Version:

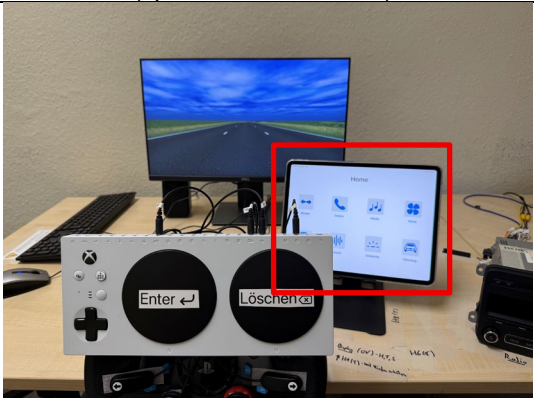


Experimental setup of the touch-based HMI.



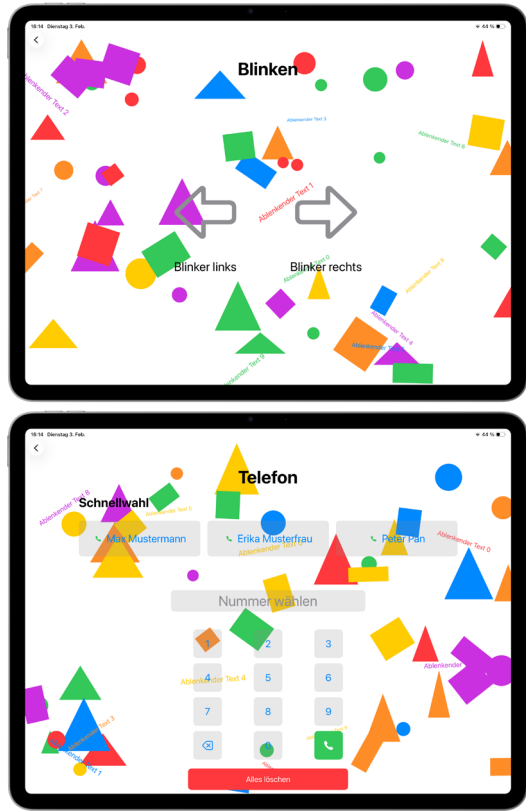
Software of the touch-based HMI: Home screen, indicator, and telephone menu.

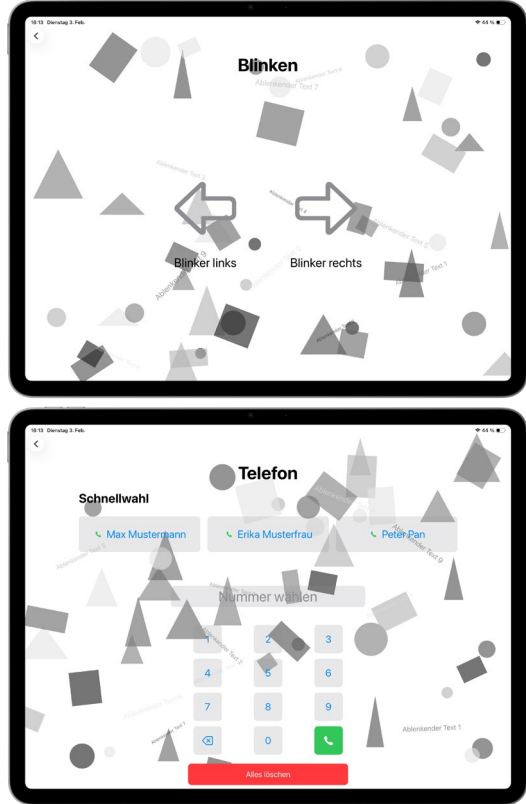
Module I – Shared HMI Features

Item	Description	Optimal	Suboptimal
M1-TB-1	Input Precision	Every touch input registered	Often not registered
		Every input is processed	Inputs processed with 33% probability only
	Example	- Display responds reliably with bare fingers, light touch - Touch registered across full display area including edges	- Screen unresponsive with cold or dry hands - Display requires firm press (resistive type) - Gloves not recognized
M1-TB-2	Reachability	One hand on the steering wheel, arm not extended	One hand on the steering wheel, arm widely extended
	Validation	Tablet close to the right of the steering wheel	Tablet placed farther away
			
	Example	- Display close to steering wheel, within arm reach without leaning - Compact vehicle with short reach to display	- Display positioned more toward passenger side - Wide landscape screen requiring driver to lean forward and right - Large SUV with wide console: display far to the right
M1-TB-3	Driving-related Information Always Visible	Always visible; display separate from touchscreen	Information sometimes covered by touchscreen
	Validation	Tablet close to the right of the steering wheel	Tablet placed in front of the monitor so that the lower right part of the LCT is obstructed (speedometer location)
			

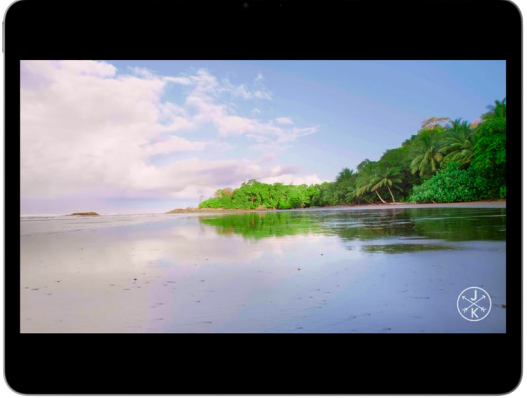
Item	Description	Optimal	Suboptimal
	Example	- Speed shown in dedicated instrument cluster, independent of touchscreen - HUD always displaying speed regardless of touchscreen state - Fixed status bar with speed, range, etc. always at top of display	- Navigation overlay hiding speed - Full-screen media player obscuring vehicle status information - Pop-up notification covering speedometer on combined screen
M1-TB-4	Reflection	Matte, strong light absorption	Glossy, anti-reflective coating
	Validation	No illumination of driver	Flashlight on the radio so that the lateral light cone hits the driver (same as M1-PC-4 condition for physical controls)
			
	Example	- Matte display coating with strong light absorption - Display looks clean when switched off, no mirror effect - Anti-glare surface visibly reducing reflections	- High-gloss display acting as a mirror in sunlight - Piano-black surrounding adding additional reflection zone - Display mounted so that it reflects outside environment
M1-TB-5	Viewing Distance Downwards	0° – 15°	30° – 45°
	Validation	Tablet close to the right of the steering wheel	Tablet on stand placed far downward where center console would be
			

Item	Description	Optimal	Suboptimal
	Example	- Display mounted high, near top of dashboard - Display at roughly instrument cluster height	- Display mounted low in center console - Display at armrest level requiring strong head tilt
M1-TB-6	Tilt	Strong horizontal and vertical tilt	No tilt
	Validation	Tablet tilted toward driver	Tablet in line with steering wheel, no tilt toward driver
			
	Example	- Display angled toward driver both horizontally and vertically - Adjustable display tilt	Display mounted vertically flush with dashboard (no tilt)
M1-TB-7	Brightness and Contrast	Optimal, adjustable and automatic	Insufficient, adjustable, no automatic regulation
	Validation	Tablet at normal brightness (80% brightness)	Tablet set quite dark (20% brightness)
			
	Example	- Automatic ambient-light brightness adjustment - High maximum brightness (sunlight-readable) - Very low minimum brightness for night driving	- Display that is too bright at night and cannot be dimmed sufficiently - Display washed out in direct sunlight with no way to boost brightness - No auto-brightness: must adjust manually while driving

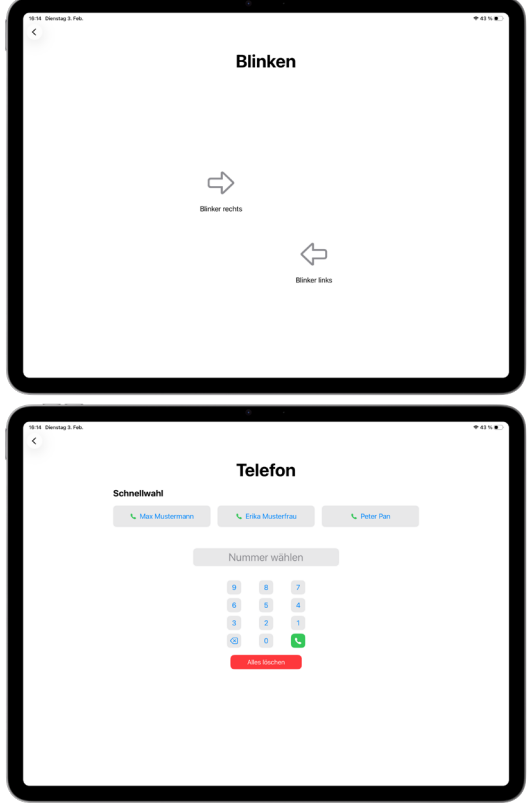
Item	Description	Optimal	Suboptimal
M1-TB-8	Attentional Guidance	Interaction principles and guidance optimal	No communication, guidance distracting
	Validation	Concise interface	Home screen & function menus: salient distractors (geometric shapes and text; colorful, low opacity)
			
	Example	- Clearly visible button borders indicating pressable surfaces - Highlighted active states showing current setting - Breadcrumb navigation showing current menu position	- Animated backgrounds drawing eyes to irrelevant areas - Pop-up notifications appearing during driving - Cluttered home screen with many competing elements of equal salience

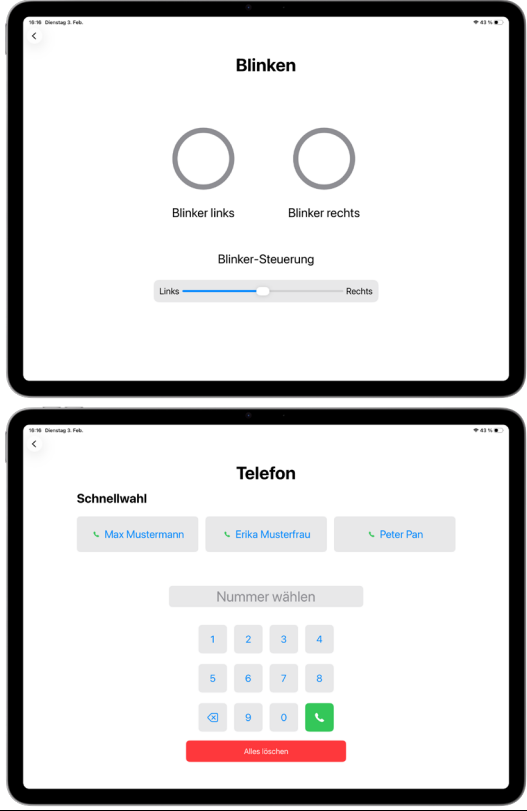
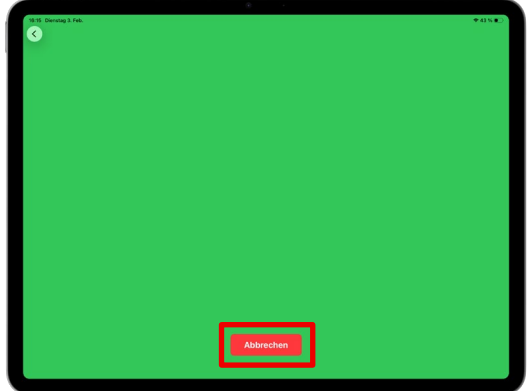
Item	Description	Optimal	Suboptimal
M1-TB-9	Minimalism	All content short and concise	Contents rather long
	Validation	Concise interface	Home screen & function menus: non-salient distractors (geometric shapes and text; grey, high opacity)
			
	Example	- Simple icon-based home menu with one function per button - Short labels (max. 1-2 words) for all menu items - Clean layout with generous spacing	- Information-dense screens requiring several seconds to read - Long scrollable text menus without icons - Multiple simultaneous functions shown with small overlapping text
M1-TB-10	Viewing Distance to the Right	0° – 15°	45° – 60°
	Validation	Tablet close to the right of the steering wheel	Tablet on stand placed to the right of the participant, same vertical height
			

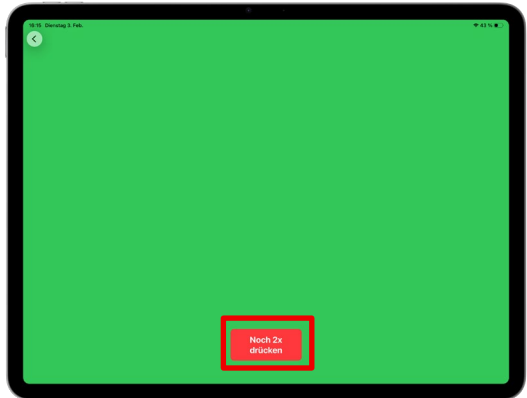
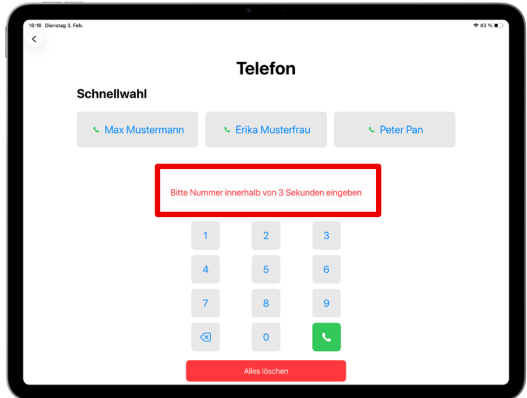
Item	Description	Optimal	Suboptimal
	Example	- Display directly to the right of steering wheel - Display close to vehicle centerline - Display within normal right-eye-movement range	- Display positioned at far right of dashboard - Passenger-side mounted display
M1-TB-11	View on the Road Not Blocked	Clear field of view	Peripheral field of view blocked
	Validation	Tablet close to the right of the steering wheel	Tablet placed in front of the monitor so that the lower right part of the LCT is obstructed – even more than M1-TB-5
			
	Example	Display positioned below window line	- Display mounted on top of dashboard crossing the sightline - Wide landscape display partially blocking lower windshield view
M1-TB-12	Deactivation	Yes, can be switched off	Standby interface
	Validation Note: No function operation	No operation: corresponds to a trial without tablet → baseline measurement used	Tablet placed next to participant showing home screen (no operation)
	Example	- Display with a dedicated off button “Do Not Disturb / Drive Mode” turns off screen completely - Display can be dimmed to near-black	- Always-on display that cannot be switched off or dimmed fully - Display locked to home screen as minimum state

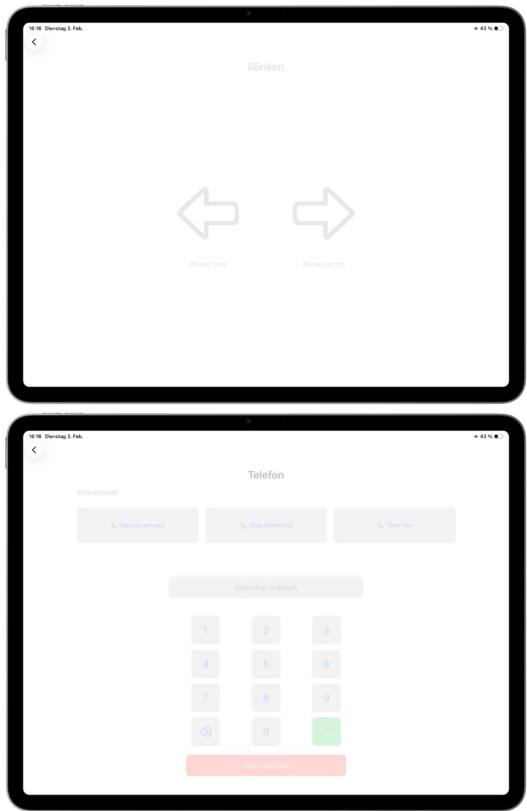
Item	Description	Optimal	Suboptimal
M1-TB-13	Blocking of Distracting Functions	Dynamic/moving content	No blocking
	Validation	Corresponds to a drive without tablet → baseline measurement used	Video playing on the tablet (without sound)
	Note: No function operation		
	Example	- Keyboard automatically hidden when vehicle is moving - Video playback blocked for driver while driving	- Streaming video accessible while driving - Full keyboard available at any speed - Scrolling social media feed accessible on center display while driving

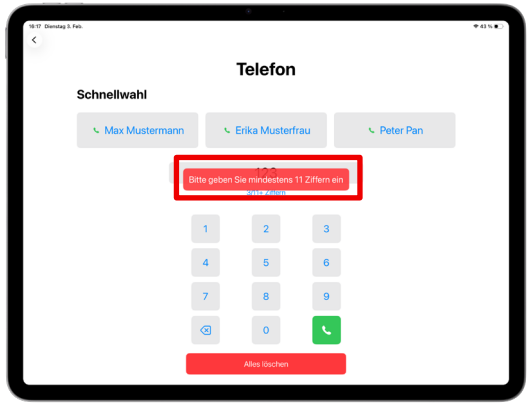
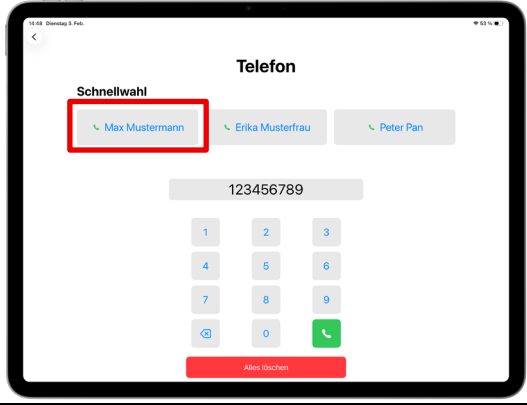
Module II b – Function-Specific HMI Features

Item	Description	Optimal	Suboptimal
M2-TB-1	Susceptibility to Operation Errors	Operation logic and buttons good	Operation logic and buttons unfavorable
	Validation	Buttons optimal	Buttons smaller and unusually arranged (indicator: not on one level and swapped; phone: number pad smaller and upside down)
			
	Example	- Large, well-spaced buttons for media controls - Clear hierarchical menu structure (functions grouped logically) - Sliders with sufficient height for imprecise touch	- Very small arrow buttons for media controls - Identical-looking buttons for different functions in close proximity - Swipe gesture required for a simple on/off toggle

Item	Description	Optimal	Suboptimal
M2-TB-2	Expectation Conformity	Contents and operation logic follow conventions	Strong deviation from conventions
	Validation	Operation logic as with other systems	Unusual controls & symbols: indicator: slider and circles instead of arrows; phone: 4×3 instead of 3×4 dial pad
			
	Example	- Volume slider moves left–right as expected - Temperature controlled with plus and minus elements - Standard play/pause/skip icons matching physical controls	- Swipe gesture replaces tap for simple on/off action - Temperature slider runs bottom-to-top (inverted expectation) - Brand-specific icons for common functions
M2-TB-3	Risk of Operation Errors	No safety risk	Strong distraction
	Validation	No distraction	Feedback very salient – flashscreen (green: max. brightness, red: normal) & error sound; cancellation possible (same as M2-PC-4 for physical controls)
			

Item	Description	Optimal	Suboptimal
	Example	- Accidental tap on seat heating: minor thermal effect only - Accidentally pressing "mute": easily reversed, no distraction	- Accidentally triggering full-volume media (sudden loud noise) - Accidentally initiating video playback (full-screen takeover) - Accidentally disabling driver assistance (safety risk)
M2-TB-4	Correction of Operation Errors	One action, immediately	Multiple actions, slowly
	Validation	Direct cancellation possible	Cancellation requires multiple steps (press cancellation button three times) (same as M2-PC-3 for haptic)
			
	Example	- Persistent "Undo/Cancel" button on every screen - Back button immediately undoes last action - Home button always accessible to exit any function	- Navigation requiring multiple taps to cancel an active route - No undo function: must re-navigate through menu to find original setting - Cancel button only appears after completing all steps
M2-TB-5	Interaction Speed	Waits indefinitely	Waits briefly, returns, gives instructions
	Validation	Function waits indefinitely during number entry	Number entry reset after 3s (after last input) and must be re-entered completely
	Note: Phone only		
	Example	- Phone dialer keeps entered number until driver confirms call - Navigation search keeps query open indefinitely - Number pad stays active waiting for driver input	- Phone dialer clears entered number after 5 s of inactivity - Navigation search auto-cancels after brief pause - Timeout resets entire multi-step process without warning

Item	Description	Optimal	Suboptimal
M2-TB-6	Additional Feedback	On, off, adjust, error – acoustic and visual	Visual feedback only
	Validation	Additional feedback present	Tablet muted
	Example	- Click sound + button highlight on every tap - Confirmation tone + visual state change for each operation step - Error tone + red highlight when blocked option is selected	- Silent touch interface: only color change of button, no sound - No feedback when end of a list or unavailable option is reached - Visual-only feedback (inaudible when music is playing)
M2-TB-7	Legibility	Contents large and high-contrast	Contents too small/large, low-contrast
	Validation	Contents large and high-contrast	Icons & texts: light grey (low contrast) and very small
	Validation		
	Example	- Large text and icons on dark background - White text on dark blue/black background - High-contrast active/inactive states for buttons	- Light grey text on white background - Very small sub-function labels in complex menus - Low-contrast icons requiring close visual inspection

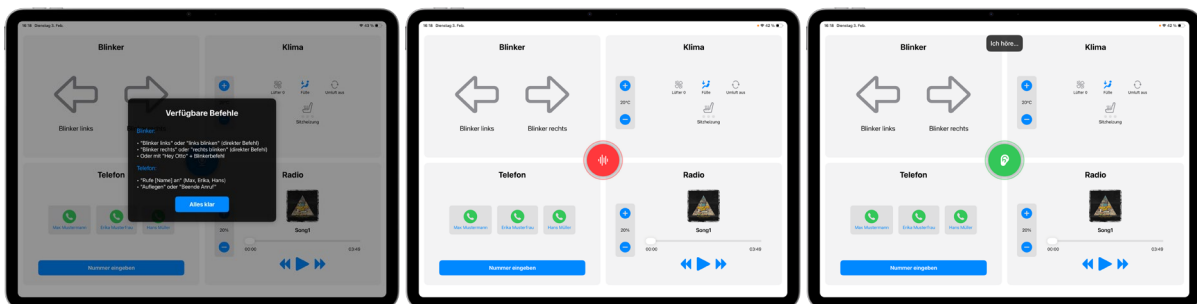
Item	Description	Optimal	Suboptimal
M2-TB-8	In-process Abortion	Possible at any point	Only at some points and only cumbersome
	Validation	Entered number can be partially or fully deleted	Number with min. 11 digits must be entered before other buttons (start call, delete one digit or all digits) can be used
	Note: Phone only		
	Example	- Cancel button (X) always visible during multi-step input - Home button always accessible, exits any process immediately - Back button cancels partial input at any stage	- Minimum number of digits must be entered before canceling is allowed - Full address must be entered before the process can be aborted “Back” must be pressed multiple times to exit nested menus
M2-TB-9	Shortcut	Executable in one step	No shortcuts
	Validation	Call Max via shortcut	Enter full number
	Note: Bonus item Phone only	→	
	Example	- Customizable home screen tile → direct one-tap access to function - Favorite contact button for direct call without navigating - Pinned navigation destination accessible in one tap	- several menu levels required to reach a frequently used function - No home screen customization possible - Every use requires going through full menu hierarchy

SPEECH-BASED HMI

Reference Version:

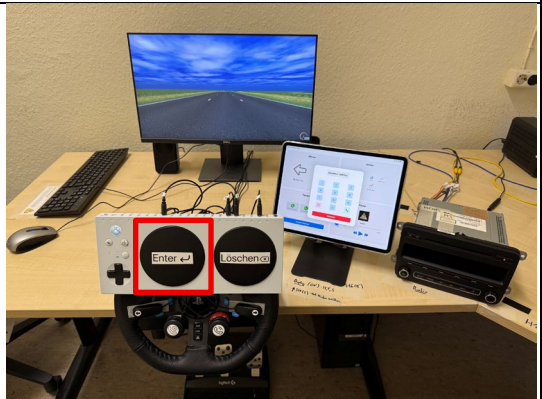


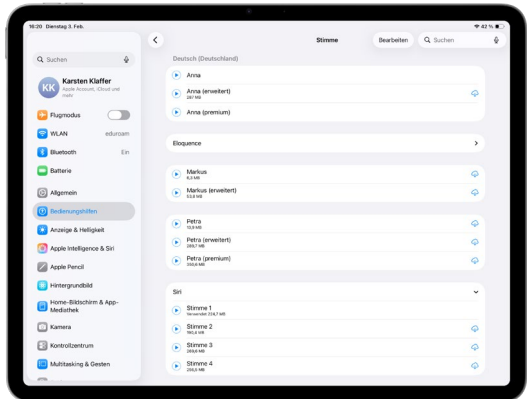
Experimental setup of the speech-based HMI.



Software of the speech-based HMI: Tutorial after startup, standby mode, and listening mode.

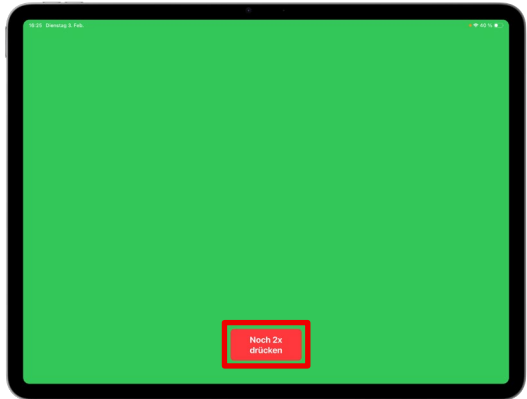
Module I – Shared HMI Features

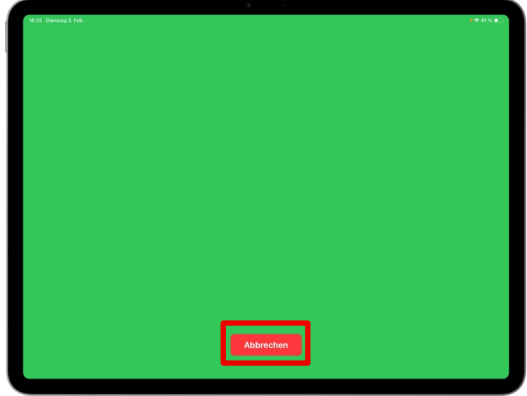
Item	Description	Optimal	Suboptimal
M1-SB-1	Deactivation Option & Alternative Operation Option	Speech-based HMI deactivatable, fallback correct (decision tree)	Speech-based HMI not deactivatable, incorrect fallback (decision tree)
	Validation Note: Correct Fallback tested	Correct fallback: Indicator: phys. controls Phone: touch-based HMI	Speech operation only → reference version measurement used
	Example	- Key phrase (“Hey XYZ”) can be disabled in settings - Full voice assistant deactivatable; all complex functions remain on touchscreen, simple functions on buttons	- Always-listening assistant with no option to deactivate - Voice-only navigation input, no touch alternative
M1-SB-2	Quality of the Speech Recognition	Every voice input recognized	Prone to dialects and background noise
	Validation	Functions reliably controllable	Command processing disrupted; in 70% of cases a different command is processed
	Example	- Modern natural language processing-based system - Works with open windows, radio on, or passenger talking - Understands regional accents and casual phrasing	- Older keyword-recognition systems requiring exact phrasing - System fails when radio or passenger voice is audible - System only reliable for driver with standard accent in quiet cabin
M1-SB-3	Hands-free	Initial activation via key phrase	Push-to-talk button before each input on the steering wheel
	Validation	Activation via “Hey Otto” possible, then voice command	No “Hey Otto” – press Enter button (on steering wheel) each time, then give command; speech-based HMI does not remain active after command ends
			

Item	Description	Optimal	Suboptimal
	Example	- Single key phrase activates voice assistant for entire conversation “Hey XYZ” → no further button presses - Stays in listening mode across multiple commands	- No persistent listening; button must be re-pressed each time - Push-to-talk button not on steering wheel (requires reaching)
M1-SB-4	Robustness of Activation	Key phrase and button not too sensitive	Too sensitive to noise and inputs
	Validation	Activation via “Hey Otto” reliable	Randomized transition from “Hey Otto” to listening status – timer between 5 and 10 seconds
	Example	- Key phrase sufficiently distinct from common conversation words - System ignores passengers and radio/podcast content - Push-to-talk button recessed / not easily pressed accidentally	- Key phrase close to everyday words (accidental activation in conversations) - System activates when passenger uses similar phrase - Radio host or podcast accidentally triggers assistant
M1-SB-5	Intelligibility / Quality of Voice Output	Perfectly natural language	Rather artificial language
	Validation	Voice “Viktor”	Voice “Anna”
			
	Example	- Neural text-to-speech with natural intonation - Voice clearly distinguishable from ambient noise	- Monotone, robotic voice from older systems (choppy cadence) reducing comprehension - Unnatural stress patterns making instructions hard to follow

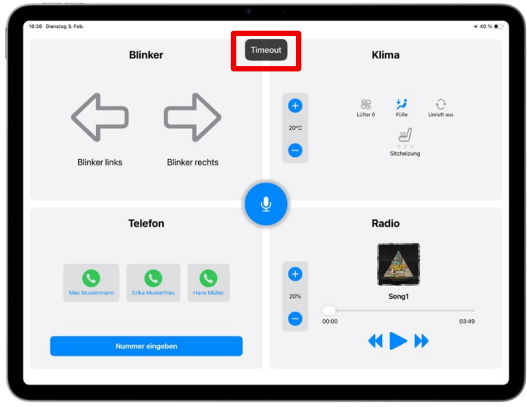
Item	Description	Optimal	Suboptimal
M1-SB-6	Feedback Dimensions	Speech output with chimes and visual feedback	Speech output only, no chimes or visual feedback
	Validation	Speech, chimes, visual	Speech only – no notification chimes & no visual overlay (feedback dimensions refer only to speech-base HMI, not to the functions itself)
	Example	• Activation chime + spoken response + instrument cluster overlay • Listening state indicator on display + tone when ready • Confirmation sound + brief text on HUD	• Voice response only, no visual indicator of listening state • No signal whether command was understood or failed • Only audio feedback (inaudible if music is playing)
M1-SB-7	Volume	Default volume appropriate, adjustable, automatic	Default volume too loud/quiet, adjustable, no automatic
	Validation	Volume appropriately loud (50%)	Volume set very low (10%)
	Example	• Automatic volume increase at highway speeds • Music ducked automatically when assistant speaks • Volume adjustable via voice command (“speak louder”)	• Assistant barely audible with windows down at speed • No adaptation to background noise level • Volume not separately adjustable from media volume

Module II b – Shared HMI Features

Item	Description	Optimal	Suboptimal
M2-SB-1	Fallback	Correct according to decision tree	Incorrect according to decision tree
	Validation Note: Incorrect Fallback tested	Indicator: phys. controls Phone: touch-based HMI (already measured with M1-SB-1)	Incorrect fallback: Indicator: touch-based HMI Phone: physical controls
	Example	- All complex voice-controlled functions also on touchscreen - Simple functions also available via physical controls	- Voice-only phone dialing with no touch alternative - Navigation input only possible via voice - Fallback uses wrong HMI type (e.g., turn signal via touchscreen)
M2-SB-2	Correction of Operation Errors	One command, barge-in, immediately, verbal feedback	Multiple commands, wait for assistant, no feedback
	Validation	Direct cancellation via button and speech	Multiple-step cancellation (press button 3×) or elongated dialogue (same as M2-PC-3 / M2-TB-4)
			
	Example	“Cancel!” barge-in interrupts assistant mid-sentence - Cancellation confirmed verbally (“Okay, cancelled”) - Single command sufficient to abort any action	- Must wait for assistant to finish speaking before canceling - Requires repeating “Cancel” three times to abort - No verbal confirmation that cancellation was accepted
M2-SB-3	Susceptibility to Operation Errors	Function reliably operable	Unreliable, other functions are activated
	Validation	Function reliably controllable	Randomization of subfunction activation (indicator or phone); in 70% of cases a different subfunction is activated
	Example	“Play Playlist XYZ” always activates media playback reliably - Ambiguous commands prompt clarification rather than wrong action	“Radio” misrecognized as “Call Mario” - Similar-sounding function names causing frequent wrong activations

Item	Description	Optimal	Suboptimal
M2-SB-4	Shortcut	Own ideal command possible	Long predefined command, voice control gives hints
	Validation Note: Non-shortcut tested for indicator, shortcut for phone	Direct function activation without keyword "Hey Otto" and by skipping steps	Elongated function operation with keyword and extra operation steps
	Example	• "Drive home" starts navigation home directly • "Call Max" without specifying function prefix • User-defined voice shortcuts configurable in settings	• Must say full phrase: "Hey XYZ, Navigation, Drive to home address, Confirm" • No custom shortcut creation possible • Every use requires full menu dialog regardless of frequency
M2-SB-5	Risk of Operation Errors	No safety risk	Strong distraction
	Validation	Function works normally	Feedback very salient – flashscreen (green: max. brightness, red: normal) & error sound; direct cancellation via touch button possible (same as M2-PC-4 / M2-TB-3)
			
	Example	• Accidentally activating seat heating → mild thermal effect only • Wrong media track selected → easily corrected, no distraction	• Accidentally initiating video playback (full-screen takeover) • Wrong call initiated → full-screen outgoing call overlay appears • Accidentally disabling lane-keeping assist (safety-critical)

Item	Description	Optimal	Suboptimal
M2-SB-6	Additional Feedback	Speech, chimes, visual if multi-step	Speech too long, visual too long
	Validation	Additional feedback	No notification chimes, only speech & visual feedback (feedback overlay) – unnecessarily long due to filler phrase “command has been processed”
	Example	• “Done” + chime + small display indicator • “Heating on” + checkmark on cluster display	• “Your request to activate seat heating has been successfully processed” (too long) • System reads full navigation instructions before allowing input • Long visual overlay persisting on screen for 10+ seconds
M2-SB-7	Diverse Commands	Many common expressions recognized	One keyword must always be included
	Validation	Variety of commands possible	Keyword must always be included: turn signal commands must start with “Turn signal – ...”; telephone commands with “Telephone – ...”
	Example	• “I’m cold” → seat heating on • “warmer please” → temperature up	• “I’m cold” not understood; only “Seat heating on” works • Every command requires a exact predefined keyword
M2-SB-8	In-process Abortion	At every point	At some points, only cumbersome
	Validation	Call setup can be cancelled during process – via touch & speech	No cancellation possible
	Note: Phone only		
	Example	• “Cancel” works at any point during phone dialing dialog • Starting a new function automatically cancels the current one	• Full phone number must be spoken before process can be aborted • System continues dialing even after “Stop” is said

Item	Description	Optimal	Suboptimal
M2-SB-9	Interaction Speed	Waits indefinitely	Waits briefly, gives instruction, cancels
	Validation	Listening status after “Hey Otto” indefinitely long	Listening window after “Hey Otto” activation only 3 seconds, then timeout
	Note: Phone only		
	Example	• Phone dialing keeps number entry open indefinitely until confirmed • Navigation keeps destination query active until driver confirms • System explicitly prompts and waits at each step	• Listening window closes after 3 s with no explanation • Quick timeout + error sound without retry prompt • Entire multi-step process resets if driver pauses briefly



Gesamtverband der Deutschen Versicherungswirtschaft e. V.
(GDV - German Insurance Association)
Wilhelmstraße 43/43 G, 10117 Berlin
Postfach 08 02 64, 10002 Berlin
Tel. 030 2020-5000, Fax 030 2020-6000
www.gdv.de, berlin@gdv.de