A Systematic Evaluation of Solutions for the Final 100m Challenge of Highly Automated Vehicles

MARK COLLEY, Institute of Media Informatics, Ulm University, Germany
BASTIAN WANKMÜLLER, Institute of Media Informatics, Ulm University, Germany
ENRICO RUKZIO, Institute of Media Informatics, Ulm University, Germany

Automated vehicles will change the interaction with the user drastically. While freeing the user of the driving task for most of the journey, the "final 100 meters problem", directing the vehicle to the final parking spot, could require human intervention. Therefore, we present a classification of interaction concepts for automated vehicles based on modality and interaction mode. In a subsequent Virtual Reality study (N=16), we evaluated sixteen interaction concepts. We found that the medially abstracted interaction mode was consistently rated most usable over all modalities (joystick, speech, gaze, gesture, and tablet). While the steering wheel was still preferred, our findings indicate that other interaction concepts are usable if the steering wheel were unavailable.

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Human-centered computing** → **HCI theory, concepts and models**; **Empirical studies in HCI**.

Additional Key Words and Phrases: Systematic comparison; interaction modalities; automated vehicles.

ACM Reference Format:
Mark Colley, Bastian Wankmüller, and Enrico Rukzio. 2022. A Systematic Evaluation of Solutions for the Final 100m Challenge of Highly Automated Vehicles. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 178 (September 2022), 19 pages. https://doi.org/10.1145/3546713

1 INTRODUCTION

Interaction with automated vehicles (AVs) is expected to change drastically compared to today's manually driven vehicles [21]. During the automated journey, users will be able to engage in non-driving related tasks (NDRTs) such as reading, sleeping, or watching a movie [43]. However, technical and interaction challenges remain: inaccurate map data restraining the accuracy of selecting the true destination or uncertain and dynamically changing user needs regarding the destination [45]. These challenges are called the "final 100 meters problem" of AVs [45]. Based on these challenges, human intervention could still (scarcely) be required despite a potential widespread adoption of AVs. Examples of such an interaction could be the determination of the final parking sport (e.g., not too far away from an entrance, not in direct sunlight, not near a tree with birds in it) or altering a driving course required, for example, due to outdated map data. Such an interaction is, thus, defined by its scarcity and the necessity to clearly communicate one's goals related to the driving task. Previous work, in general, has proposed numerous solutions either in general to navigate vehicles via joystick [1], eye-gaze [52], or devices such as tablets [50], smartphones [64] or, with a special focus on the "final 100 meters problem", using gestures [45].

Authors' addresses: Mark Colley, mark.colley@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany; Bastian Wankmüller, bastian.wankmueller@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany; Enrico Rukzio, enrico.rukzio@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany.

2573-0142/2022/9-ART178
https://doi.org/10.1145/3546713

While these interaction concepts pose individual challenges and provide specific advantages, a systematic comparison is missing. Therefore, we first conducted a systematic literature survey on possible interaction *modalities* and *modes*. We screened 609 papers and found 21 concepts in total including chair-, controller-, and joystick-based ones. Based on this survey, we implemented **16** prototypes in Virtual Reality (VR). We employed a steering wheel as a baseline and gestures, speech, a tablet, a joystick, and eye-gaze as modalities. We omitted a chair-based concept as we believe it to be an inappropriate steering mechanism due to missing accuracy and decided to only implement the joystick and not the controller due to the high conceptual overlap. For each of these modalities (except the baseline), we implemented a *direct*, a *taxi-like*, and an *overview* mode with varying degrees of driving task abstraction and technical sophistication required. With these concepts, we conducted a within-subjects study (N=16). We found that participants preferred the steering wheel and that the *taxi-like* mode was rated the most usable *interaction mode*. Task load was highest in the *direct interaction mode*. However, effects were not uniform across all modalities. Our data indicate that while participants preferred the steering wheel, other interaction modalities and modes are possible and usable, therefore, guiding designers of future AVs to introduce AVs safely.

Contribution Statement: This work contributes a (1) systematic literature survey on interaction modalities and modes for directing (automated) vehicles, (2) 15 implementations of previously described and novel interaction concepts for the "final 100 meters problem" of AVs for the *modalities* joystick, gesture, speech, tablet, and eye-gaze and the *interaction modes direct*, *taxi-like*, and *overview*. Finally, this work contributes (3) the findings of a VR study with N=16 evaluating the interaction concepts. Results show that *taxi-like* interactions were rated as most usable. Furthermore, the NASA TLX score showed that *taxi-like* and *overview* were approximately equally demanding. Nonetheless, participants preferred the *baseline* steering wheel.

2 CONCEPT CLASSIFICATION

We classified prior work on steering (automated) vehicles. For this, we conducted a PRISMA [39] based literature survey. We queried the proceedings of the six most cited HCI venues according to Google scholar [36]. Due to their contents regarding HCI research on future mobility, we additionally retrieved publications from the venues named in Table 1. Our search query was:

```
("steer" OR "steering") AND ("vehicle" OR "autonomous vehicle" OR "automated vehicle" OR "robot" OR "conditional driving") AND ("speech" OR "gaze" OR "touch" OR "gesture" OR "haptic")
```

We excluded work that was (1) non-English or non-German, (2) not peer-reviewed work, or (3) when the driving or movement task was not the primary (e.g., driver monitoring was deemed irrelevant) evaluation. Finally, we included only work that proposed an interaction mechanism, therefore, for example, work on steering wheels that solely were enhanced via haptic feedback was excluded. We considered publications from the last 10 years (2011—2021). Two researchers carried out the literature search. After initially coding work together to form a common understanding, we identified and screened 609 publications. **18** papers (5 in CHI, 1 in UIST, 6 in AutoUI, 4 in ETRA, 1 in IJHCI, and 1 in Transportation Research Part F) matched our criteria and provided the basis for our comparison. As our focus was on all possible interaction modalities, we also included racing-based games (e.g., [51]). We classified prior work based on the employed *modality* and the *interaction mode*. Two of the *interaction modes* are based on the work by Funk et al. [25]: input can be either of continuous nature (meaning that input is permanently needed) and directly affects the movement (thus called *direct* from here on), or discrete (or, in our case, *taxi-like*) with pre-defined commands that, however, limit input capabilities (also see possible actions in [9] and the maneuvers in [16]). Additionally, we added a self-defined category *overview*, which represents the highest abstraction. In such *overview* concepts, only the final destination is given to the system, for example,

by looking at the desired spot. Such interaction was, for example, already shown by Tscharn et al. [56]. However, such systems require the highest technical sophistication. Due to the nature of the interaction (non-verbal, e.g., humming) in the work by Funk et al. [25], this *overview interaction mode* was not possible. The codings are shown in Table 2.

Table 1. Retrieved venues and number of publications.

Conference / Venue	**Number of publications** (found)
ACM Conference on Human Factors in Computing Systems (*CHI*)	**5** (152)
ACM Conference on Computer-Supported Cooperative Work & Social Computing (*CSCW*)	**0** (4)
ACM/IEEE International Conference on Human Robot Interaction (*HRI*)	**0** (2)
ACM Symposium on User Interface Software and Technology (*UIST*)	**1** (20)
ACM Conference on Pervasive and Ubiquitous Computing (*UbiComp*)	**0** (10)
ACM International Conference on Human-Computer Interaction with Mobile Devices and Services (*MobileHCI*)	**0** (13)
IEEE Transactions on Affective Computing	**0** (0)
ACM Symposium on Eye Tracking Research and Applications (ETRA)	**4** (10)
ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications (*AutoUI*)	**6** (258)
International Journal of Human–Computer Interaction (IJHCI)	**1** (56)
Transportation Research Part F: Traffic Psychology and Behaviour	**1** (84)
Combined	**18**

Table 2. Categorized eligible work. If multiple modalities were used, these have been categorized separately. X denotes that no prior work was found.

Modality \ Mode	*direct*	*taxi-like*	*overview*
Touch	X	4: [60–62, 66]	X
Steering wheel	3: [8, 34, 40]	X	X
Speech	X	2: [60, 65]	X
Eye-gaze	1: [52]	3: [2, 65, 69]	2: [27, 68]
Chair	1: [59]	X	X
Controller	2: [51, 69]	X	X
Tablet	1: [50]	X	X
Joystick	1: [47]	X	X
Gesture	1: [45]	X	X
Combined	**10**	**9**	**2**

We found interactions with (partially) AVs (see SAE J3016 [53]) in various abstractions and with different modalities. For example, Walch et al. [62] used a touchscreen showing a button with which the passenger could interact to let the AV overtake vehicles in front. Wiegand et al. [66] also used a touchscreen but varied the interaction concept (minimalist vs. conversational) for assessing a pedestrian crossing in front of an AV. Ros et al. [47] proposed a concept where the passenger can draw the future trajectory of the AV. Joysticks were already investigated as a possible input device in the early 2000s (e.g., [1]).

Other input devices used were, for example, a tablet for lateral steering [50]. The authors found that this nomadic device, when already in use (e.g., for an NDRT), is superior in reaction times as the change of input modality is omitted. This approach was also well accepted in terms of user

experience. This is in line with work by Wang et al. [64]. They used the pitch and roll motion of a smartphone as an input device for the steering of a vehicle. While city road traveling was slightly inferior, extreme maneuvers were handled better with the smartphone.

Speech was also used to determine relevant objects in a scene [60]. While not used directly for the driving task, Funk et al. [25] evaluated discrete, binary, and continuous non-verbal auditory input, showing that for binary and discrete input, snapping fingers were preferred. For continuous input, the humming was preferred.

Regarding eye-gaze interaction, we predominantly found interaction regarding the movement of vehicles or wheelchairs via a layover where different areas of the screen match certain maneuvers (left, right, ahead, backward) [2, 52, 69]. This input was either continuous [52] or based on commands activated via dwell time (e.g., 500ms [2]). Compared to a traditional controller, this eye-gaze input was found to be 31% slower [69]. Another gaze-based interaction uses waypoints determined by concentrating on a certain location [27, 68]. Wang et al. [65] used the eye-gaze to indicate relevant objects in the scenery the AV should avoid. Therefore, the input was coded as mid-level, meaning that a medium sophistication is required (object detection, gaze recognition, matching).

Qian et al. [45] explored in-air static hand gestures for AVs' "final 100 meters" problem. The "final 100 meters" problem stems from the potential need of AV users to adjust the direction of the AV, for example, to select the desired parking spot via speech or gesture due to, for example, uncertain user needs. They define six user-defined hand shape categories: Palm, fist, thumb, index finger, little finger (i.e., a fist with an outstretched little finger), and roll. Most participants used a "palm-forward, fingers-up gesture – the same one that police use" [45, p. 6]. Other gestures included waving. Participants in their study preferred dynamic gestures, which are gestures including movement. However, the authors provide some evidence that static gestures could be better suited because of less performance time, less distraction, easier standardization, higher learnability, easier detection by technical systems, and less required performance space.

In racing games, traditional controllers [51] and a chair [59] (tilting for acceleration and braking) were used.

3 CONCEPTS AND IMPLEMENTATION

Following the literature review, we found that numerous modalities were used to steer a (partly automated) vehicle. However, we also found that there was no comparison between these modalities in the literature. Therefore, we either implemented existing concepts for a modality or designed novel ones. As different levels of technical maturity can be expected of such vehicles, we also distinguished between three *interaction modes*: *direct* refers to a continuous input, for example, as done today via the steering wheel. In this mode, the user can control every aspect of the driving task. On the other hand, *taxi-like* interactions, as the name indicates, refer to a higher abstraction in which the passenger provides medium-level instructions. We describe these per modality in the following. Finally, in the *overview* mode, the highest sophistication of the AV is assumed. In this mode, the user can only select viable options (e.g., pre-defined parking spots) for the AV to drive to independently.

Based on the literature review, we chose gesture, eye-gaze, speech, tablet, and joystick as modalities. The steering wheel was only used in the *direct* mode as a *baseline*. We omitted the chair as it seems an inappropriate steering mechanism due to missing accuracy. Despite implementing the controller-based interaction, we also omitted the controller both due to high conceptual overlap with the joystick (the controller sticks can be seen as miniature joysticks) and to avoid an overlong study.

For the comparison, we implemented a VR simulation using Unity 2019.4.26f [58] and the asset Suburb Neighborhood House Pack [23]. This asset provides a typical neighborhood located in the

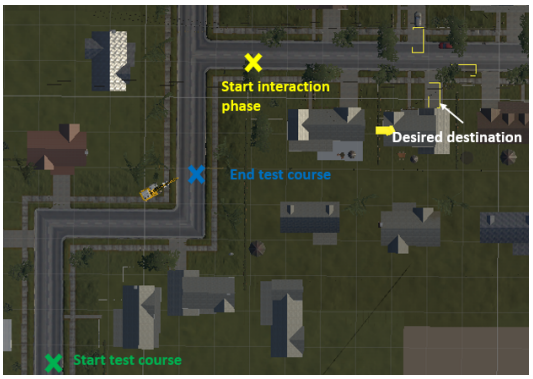


Fig. 1. Overview of the scene. The yellow squares indicated possible parking spots and were only visualized in the center console in the *overview* mode.

USA. In addition, we integrated a test track before the main track (see Figure 1). The participant sat in a simulated Mercedes F015 [3] with all logos removed. A steering wheel was only present in the *baseline* condition where a steering wheel was used (see Section 4). The velocity of the vehicle was restricted to 10 km/h to allow for more accurate input. While, with more training, higher velocities seem possible, we argue that increasing safety is of high importance as users will not interact with the AV on a regular basis. The movement of the vehicle was implemented using Polarith AI Pro [44]. In the *overview* mode, the AV is simulated to determine possible parking spots and displays these on the center screen (see Figure 1 and Figure 2c). We used a Vive Pro Eye and we employed a fan to reduce motion sickness [18].

3.1 Baseline: Steering Wheel

We used a Thrustmaster T150 Pro in combination with the asset Rewired [26]. This interaction modality served as a *baseline*, therefore, no concepts regarding the interaction modes *taxi-like* and *overview* have been implemented. While these would be possible (e.g., selecting a position to drive to via turning the wheel and employing the pedals), these are counter-intuitive as the movement of the steering wheel and the pedals have to be mapped to a 2D map and have, therefore, been discarded.

3.2 Joystick Interaction

We used a Mad Catz F.L.Y.5 joystick again in combination with the asset Rewired [26]. In addition, we designed a custom cap (see Figure 6a) resembling the interaction in the center console of the VISION AVTR [4].

For the *direct* mode, the user either presses the joystick to the front (or the back) to drive on or turns the cap to turn.
In the *taxi-like* mode, this interaction is only necessary at intersections.
In the *overview* mode, the user can move a yellow dot on the map displayed on the center screen. We used a dwell time of 5s for selection. While a selection via a button press would, in general, be possible and potentially quicker, the employed hardware restricted us.

3.3 Gesture Interaction

For gesture interaction, the VIVE Hand Tracking SDK [30] in version 1.0.0 was used for the automatic recognition of the hands and the gestures.

Fig. 2. Three screenshots from the gesture concepts. **(a)** shows the *direct*, **(b)** the *taxi-like*, and **(c)** the *overview interaction mode*.

For *direct* mode, a pointing gesture for-, back-, and sidewards had to be continuously used (see Figure 2a). This static gesture is based on the user elicited gestures by Qian et al. [45]. In their user elicitation study, they found that the pointing gesture was most often used, therefore, we implemented this gesture for the straight-on command. For turning, we also employed static gestures pointing in the desired direction.
In the *taxi-like* mode, the same gestures had to be used solely at intersections (see Figure 2b).
In the *overview* mode, we envisioned two possible concepts:
(1) Point at the desired location on the center screen and select via a fist gesture with the left hand.
(2) Pointing into the world in the direction of the desired parking space and selection via a fist gesture with the left hand (resembling work by Rümelin et al. [49]; see Figure 2c).

3.4 Eye-Gaze Interaction

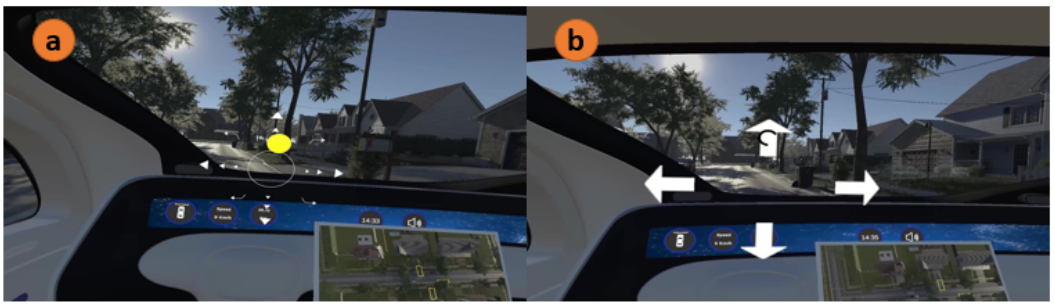


Fig. 3. Two screenshots from the eye-gaze concepts. **(a)** shows the *direct* and **(b)** the *taxi-like interaction mode*.

For eye-tracking, we used the built-in Tobii eye tracker of the HTC Vive Pro Eye (using Tobii Eye Tracking SDK version 2.1.1 and the Tobii Gaze-2-Object-Mapping [55]). Due to cursor jittering, the 1€-Filter [7] was used to smooth cursor movements with a frequency of 120 ms. A head-mounted display or an augmented reality (windshield-)display is required for the eye-gaze concepts to be applicable.

For *direct* mode, we re-implemented the final concept of Stellmach and Dachselt [52] (see Figure 3a). The user directs the vehicle via eye movements on a canvas. In the middle, the space allows for the user not to interact with the automation. Depending on the amplitude of the eye-gaze on the arrows, the vehicle adjusts its velocity.
In the *taxi-like* mode, the four directions straight on, left, right, and backward were shown as arrows and selected via a dwell time of 1s. These had to be used solely at intersections (see Figure 3b).

In the *overview* mode, we again envisioned two possible concepts:
(1) Looking at the desired location on the center screen and selection via blinking (1s to reduce false positive recognitions). While blinking might have negative side effects (e.g., accidentally blinking could lead to a triggering of an action), this was chosen to avoid the necessity for a multimodal approach. In reality, a multimodal approach would most likely be more appropriate.
(2) Looking into the world in the direction of the desired parking space and selection via blinking.

3.5 Speech Interaction

For speech recognition, we used Unity's built-in phrase recognition [57]. As we conducted the study in Germany, we report the possible commands and their English translation.

We considered humming for the *direct* mode but discarded it due to the reported high workload induced by using it for continuous input [25]. Instead, we opted for repeatedly necessary command-based interactions. The possible commands were "vorwärts" (straight on), "links" (left), "rechts" (right), "zurück" (back). For each command, the vehicle drove 5m. We chose 5m as this is still granular enough to be able to adjust the direction and long enough to avoid unnecessarily frequent interactions. Additionally, while street widths alter per country, 2.5m is often the minimum required width [54]. Therefore, a length of 5m per interaction makes it possible to drive over an intersection with one command.
For the *taxi-like* mode, the same commands were possible but only necessary at intersections.
For the *overview* mode, the participant could describe where the vehicle should drive in their own words. Then, the experimenter acted as a Wizard-of-Oz and input the desired location. This method allows the experimenter to manipulate the system with the participant believing the system to be autonomous [15].

3.6 Tablet Interaction

Fig. 4. Three screenshots from the tablet concepts. **(a)** shows the *direct*, **(b)** the *taxi-like*, and **(c)** the *overview interaction mode*.

We used a 10.8 inches WQXGA (2560x1600 pixels) tablet running Android 10. We used a 3D-printed mount to attach an HTC Vive tracker to track the tablet's position in VR (see Figure 6c).

In the *direct* mode, we ported the *direct* eye-gaze concept (see [52]) to the tablet (see Figure 4a). In line with this, we also adapted the *taxi-like* eye-gaze concept for the tablet (see Figure 4b).
For the *overview* mode, the user could click on any space on the map displayed on the tablet that the vehicle can reach (streets and driveways). A yellow point appeared at this location. The user is able to adjust this point or to add additional points. Finally, via a "Start" button at the lower right of the tablet, the user can indicate that the vehicle should start driving (see Figure 4c).

4 USER STUDY

We designed and conducted a within-subject study with N=16 participants to evaluate different concepts designed to aid with the challenges the "final 100 meters problem" poses. The independent

variables were *interaction mode* with three levels: *direct*, *taxi-like*, and *overview*, and *modality* with the five levels speech, gesture, tablet, eye-gaze, and joystick. This resulted in a 3×5 design. The steering wheel was the *baseline* (see Section 3). Therefore, participants encountered 16 conditions.

4.1 Pre-Study

As there were two possible implementations for the *overview* concepts for the modalities gesture and gaze, we conducted a pre-study with N=5 participants (three male, two female) to determine the most appropriate concepts. Participants were, on average, M=25.40 (SD=2.19) years old. We showed participants both implementations of the previously described concepts for the *overview* mode. For the gesture, three participants argued that the pointing into the world, two argued that the center console would be beneficial. For the gaze interaction, four participants argued for the real-world solution while one was undecided. Participants highlighted that there is no need to map the visualization on the center console to the real world as the main benefit of pointing/looking into the world. When arguing for the center console, increased precision was mentioned by both participants.

We used both real-world-based concepts for the *overview* mode in the user study as the reduced mapping necessity is seen as more relevant because, in the *overview* mode, the AV already determines the final destination possibilities.

4.2 Measurements

4.2.1 Objective Measurements: The system logged the position with 2Hz, the number of hits and duration of intersection with the curb, the accuracy of positioning the vehicle perfectly, how straight (relative to the garage door) the vehicle is parked, the distance to the garage door, and the duration.

4.2.2 Subjective Measurements: We measured the task load using the raw NASA-TLX [28] on 20-point scales and usability with the system usability scale (SUS) [6]. These measurements were done in VR after each condition.

After all conditions, participants provided open feedback and rated their preferences of the systems (per interaction level: direct, taxi-like, and overview) from highest (*ranking = 1*) to lowest (*ranking = 5 or 6*).

4.3 Procedure

Fig. 5. German introduction to the eye-gaze *taxi-like* condition at the start of the test course.

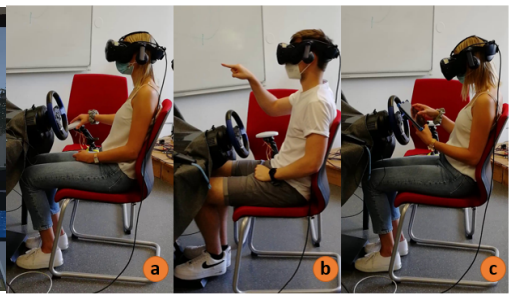


Fig. 6. Participants using (a) the joystick, (b) the gesture, and (c) the tablet.

First, participants provided informed consent and received an overview of the study. We introduced the setting as:

> *"You will drive through a suburb in a virtual reality (VR) environment in a highly automated vehicle. The vehicle takes over lateral and longitudinal guidance (braking, accelerating, steering). As the automated vehicle does not know the desired parking position, it asks you either to approach this position yourself or to tell the car (depending on the condition). First, you have the opportunity to test the interaction on a short test route. The test route ends at the roadblocks on the pavement after a right and left turn. The vehicle then drives itself around the corner until it prompts you to navigate to the destination car park. The destination house is highlighted on the center console with a yellow arrow. You are to park in front of this house directly in front of the garage in the middle. You are then to assess these types of interaction."*

Therefore, this setting represents a scenario in which the AV was not provided with a definitive location to park but was, for example, simply given an address or a change in the desired location.

Afterward, participants were able to adjust the VR headset. Then, participants were assigned to the conditions via a balanced Latin Square. Participants were, in every condition, first introduced to the interaction concept via an in-game description (see Figure 5). They were also able to test the concept in a test course (see Figure 1). After every condition, participants filled out the subjective questionnaires in VR described in Section 4.2. Participants were able to take a break at any time. Finally, participants filled out a demographic questionnaire.

Participants were compensated with 15€. Each session lasted approximately 80 min. The study was conducted in German. The hygiene concept for studies regarding COVID-19 (ventilation, disinfection, wearing masks) involving human subjects of our university was applied.

5 RESULTS

In the following, we present the results of the study.

5.1 Data Analysis

For non-parametric data, we used the non-parametric ANOVA (NPAV; function np.anova) as described by Lüpsen [37]. For post-hoc tests, Bonferroni correction was used. We calculated effect sizes using the formula proposed by Rosenthal et al. [48]. We used the R package ggstatsplot [42] for the figures, which include statistical details including the effect size, mean values, and a distribution curve. R in version 4.2.0 and RStudio in version 2022.02.3 was employed. All packages were up to date in June 2022.

5.2 Participants

We determined the required sample size via an a-priori power analysis using G*Power in version 3.1.9.7 [22]. To achieve a power of .8 with an alpha level of .05, 16 participants should result in an anticipated low to medium effect size (0.2 [24]) in a within-factors repeated measures ANOVA. Therefore, we recruited N=16 participants (4 female, 12 male). Participants were, on average, M=25.63 (SD=2.28) years old. Most participants indicated that their highest educational level was College (11) followed by High School (5). Regarding their employment status, 10 participants reported to be students, and 6 are employees. In terms of driving experience, one person had no driving licence (thus, representing no driving experience which could be common when little intervention is necessary in the future), one 1-3 years, two had their licence between 3-5 years, nine had their license between 5 and 10 years, and three between 10 and 20 years. Seven participants drove less than 7.000 km/year, eight between 15.000 and 25.000 km, and one between 25.000 and 33.000 km. Regarding the frequency, two participants stated that they never or almost never drive, four stated

that they drive less than once a month, one stated 1-3 times per month, one stated to drive once per week, four stated 3-4 times per week, one on work days, and one stated to drive daily.

On 7-point Likert scales (*1 = Strongly Disagree — 7 = Strongly Agree*), participants reported to have used VR (M=4.88, SD=2.55). Still, most participants do not use these regularly (11 participants), less than two hours per week (3), or between 2 and 8 hours per week (2).

5.3 System Usability

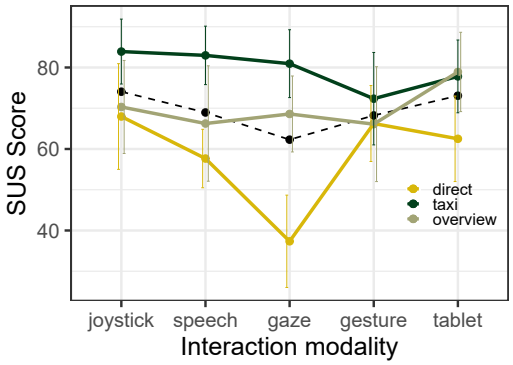
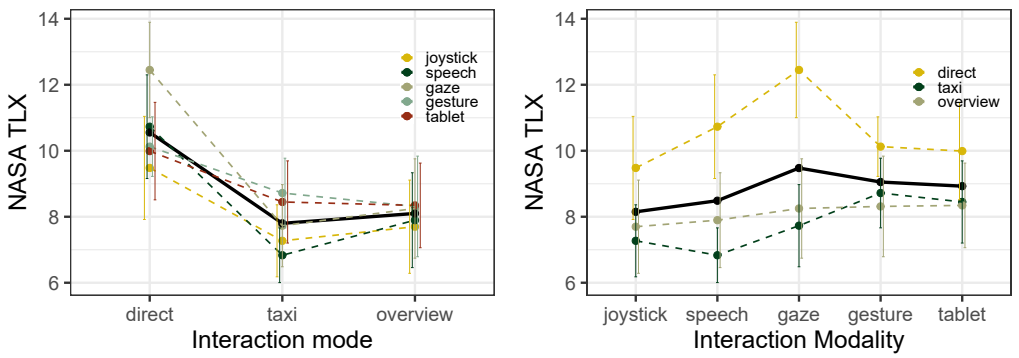


Fig. 7. Interaction effect of *interaction mode* × *modality* on SUS.

The NPAV found a significant main effect of *interaction mode* on SUS Score ($F(2, 30) = 29.02$, p<0.001). The NPAV also found a significant main effect of *modality* on SUS Score ($F(4, 60) = 3.91$, p=0.007). The NPAV additionally found a significant interaction effect of *interaction mode* × *modality* on SUS Score ($F(8, 120) = 2.26$, p=0.027; see Figure 7). While for most *interaction modes*, the usability was rated approximately equal, the direct eye-gaze concepts were rated worse. The *taxi-like* mode was rated best for almost all modalities but the tablet, where it was rated almost equally compared to the *overview* mode.

5.4 NASA-TLX



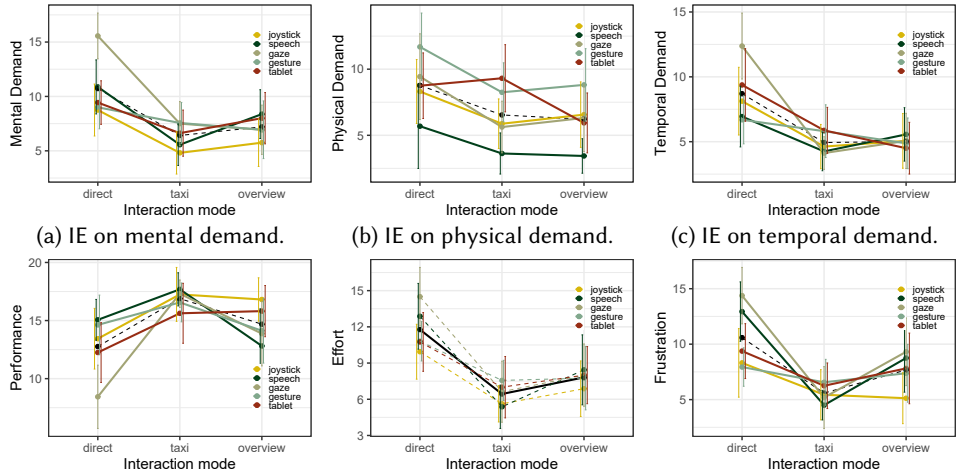
(a) Main effect of *interaction mode* on NASA TLX.

(b) Main effect of *modality* on NASA TLX.

Fig. 8. NASA TLX main effects.

The NPAV found a significant main effect of *interaction mode* on NASA TLX Score ($F(2, 30) = 51.64$, $p<0.001$; see Figure 8a). Post-hoc tests using Dunn's test showed that the *direct interaction mode* required significantly more load than the *taxi-like* and the *overview interaction modes*.
The NPAV also found a significant main effect of *modality* on NASA TLX Score ($F(4, 60) = 3.55$, $p=0.011$; see Figure 8b). Post-hoc tests using Dunn's test, however, showed these differences not to be significant.

In the following, the results for each NASA-TLX subscale are reported.



(a) IE on mental demand. (b) IE on physical demand. (c) IE on temporal demand.

(d) IE on performance. (e) Effect on effort. (f) IE on frustration.

Fig. 9. NASA TLX subscale effects.

Mental Demand. The NPAV found a significant main effect of *interaction mode* on Mental Demand ($F(2, 30) = 30.97$, $p<0.001$). The NPAV found a significant main effect of *modality* on Mental Demand ($F(4, 60) = 6.48$, $p<0.001$). The NPAV found a significant interaction effect of *interaction mode* × *modality* on Mental Demand ($F(8, 120) = 2.82$, $p=0.007$; see Figure 9a). While mental workload was rated highest for all modalities in the *direct interaction mode* and joystick, speech, and tablet were rated as requiring the lowest mental workload for the *taxi-like interaction mode*, gesture and gaze were best rated for the *overview interaction mode*.

Physical Demand. The NPAV found a significant main effect of *interaction mode* on Physical Demand ($F(2, 30) = 14.23$, $p<0.001$). The NPAV found a significant main effect of *modality* on Physical Demand ($F(4, 60) = 13.85$, $p<0.001$). The NPAV found a significant interaction effect of *interaction mode* × *modality* on Physical Demand ($F(8, 120) = 2.21$, $p=0.031$; see Figure 9b). For all modalities except the tablet and the speech, the *direct* mode was the most physically demanding, followed by the *overview*. For the tablet, the *taxi-like* mode was the most physically demanding. For speech and tablet, the *overview* mode was the least physically demanding.

Temporal Demand. The NPAV found a significant main effect of *interaction mode* on Temporal Demand ($F(2, 30) = 28.66$, $p<0.001$). The NPAV found a significant interaction effect of *interaction mode* × *modality* on Temporal Demand ($F(8, 120) = 2.73$, $p=0.008$; see Figure 9c). For all modalities, the *direct* mode was the most temporally demanding. For the tablet and the gesture, the *overview* was least temporally demanding. For the other modalities, the least temporally demanding was the *taxi-like* mode.

Performance. The NPAV found a significant main effect of *interaction mode* on Performance ($F(2, 30) = 25.05$, $p<0.001$). The NPAV found a significant interaction effect of *interaction mode* × *modality* on Performance ($F(8, 120) = 3.07$, $p=0.004$; see Figure 9d). Performance was assessed to be worst in the *direct* and best in the *taxi-like* mode for all modalities other than the tablet. For the tablet, performance was rated best in the *overview* mode.

Effort. The NPAV found a significant main effect of *interaction mode* on Effort ($F(2, 30) = 45.65$, $p<0.001$; see Figure 9e). Post-hoc tests using Dunn's test showed that the *direct interaction mode* required significantly more effort than the *taxi-like* and the *overview* modes.

Frustration. The NPAV found a significant main effect of *interaction mode* on Frustration ($F(2, 30) = 27.97$, $p<0.001$). The NPAV found a significant main effect of *modality* on Frustration ($F(4, 60) = 5.32$, $p<0.001$). The NPAV found a significant interaction effect of *interaction mode* × *modality* on Frustration ($F(8, 120) = 3.09$, $p=0.003$; see Figure 9f). Frustration was highest in the *direct* mode for all modalities and lowest in the *taxi-like* mode besides the joystick. Interestingly, the frustration was higher for gaze and speech in the *overview* mode than joystick and gesture in the *direct* mode.

5.5 Duration, Curb Hits, and Position Accuracy

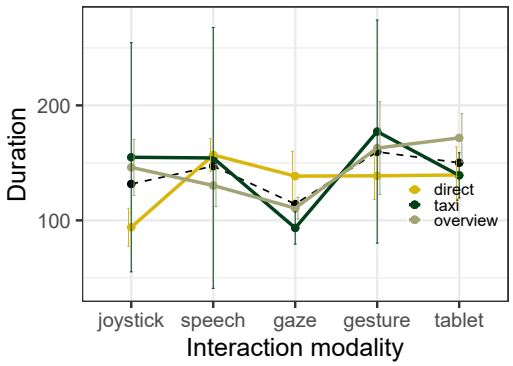


Fig. 10. Interaction effect of *interaction mode* × *modality* on duration.

The NPAV found a significant main effect of *interaction mode* on duration ($F(2, 30) = 10.37$, $p<0.001$). The NPAV found a significant main effect of *modality* on duration ($F(4, 60) = 8.66$, $p<0.001$). The NPAV found a significant interaction effect of *interaction mode* × *modality* on duration ($F(8, 120) = 8.16$, $p<0.001$; see Figure 10). Duration to the final location was lowest for joystick *direct* and eye-gaze *taxi-like* and highest for gesture *taxi-like*. The *direct* mode was, besides the joystick, consistent over modalities. Participants needed, on average, M=151.50s (SD=189.54).

We only analyzed Curb Hits and Position Accuracy for the direct mode as the participant only had full control over the AV in these conditions.

A Friedman's ANOVA found no significant differences in the number of curb hits (p=0.141). However, a Friedman's ANOVA found a significant difference (χ^2 (4)=19.97, p=.001) in the duration of curb hits. Post-hoc tests found that eye-gaze (M=8.36) had a significantly longer curb hit duration than gesture (M=1.34) and speech (M=1.30). The duration was also significantly higher for joystick (M=3.67) compared to gesture and speech.

A Friedman's ANOVA showed a significant difference in the mean values for the accuracy of the parking spot coverage (χ^2 (4)=9.58, p=.048). However, post-hoc tests showed these not to be

significant. Percentages ranged from 63% (eye-gaze) over 70% (joystick and tablet) to 71% (gesture and speech).

5.6 Reasonability and Preference

After all conditions, participants were asked to assess the reasonability of the *interaction modes* and *modalities* on 7-point Likert scales (1=*Totally disagree* to 7=*Totally agree*; see Table 3 and Table 4). Participants rated the steering wheel as the most reasonable modality and the *overview* as the most reasonable *interaction mode*.

Table 3. Reasonability measurement per modality.

Variable	Min	q_1	$\tilde{x}$	$\bar{x}$	q_3	Max	sd	IQR
gesture	1	3.0	5.0	4.4	6.0	7	1.9	3.0
eye-gaze	1	1.8	2.5	3.2	4.2	7	2.1	2.5
speech	1	4.5	5.0	4.9	6.0	7	1.9	1.5
tablet	2	3.0	5.0	4.8	6.0	7	1.7	3.0
joystick	1	4.0	6.0	5.4	7.0	7	1.7	3.0
steering wheel	2	5.8	6.5	5.9	7.0	7	1.6	1.2

Table 4. Reasonability measurement per interaction mode.

Variable	Min	q_1	$\tilde{x}$	$\bar{x}$	q_3	Max	sd	IQR
direct	1	2	3	3.2	4.2	7	1.9	2.2
taxi-like	1	4	5	4.9	6.0	7	1.8	2.0
overview	3	5	6	5.6	6.0	7	1.2	1.0

Participants rated their preference for all modalities and then for each *interaction mode* individually. A Friedman's ANOVA found a significant difference (χ^2 (5)=19.46, p=.002) difference for modality preference. The most preferred (i.e., lowest mean) was the steering wheel (M=2.44) closely followed by the joystick (M=2.50). Both were significantly better ranked than the eye-gaze-based interaction (M=4.81).

A Friedman's ANOVA also found a significant difference (χ^2 (4)=33.75, p<.001) difference for modality preference for the *direct interaction mode*. The joystick (M=1.38) was most and significantly more preferred than any other *interaction mode* other than the tablet (M=2.44). Eye-gaze (M=4.38) was also ranked significantly worse than gesture (M=3.19) and tablet. Finally, speech (M=3.62) was also rated significantly worse than the tablet. A Friedman's ANOVA showed a significant difference in the mean rankings for the *taxi-like interaction mode* (χ^2 (4)=12.35, p=.015). Post-hoc tests showed the joystick was significantly more preferred than the eye-gaze. The joystick also received values indicating the highest preference (lowest mean) of M=2.00. A Friedman's ANOVA also showed a significant difference in the mean rankings for the *overview interaction mode* (χ^2 (4)=12.85, p=.012). Post-hoc tests showed the tablet was significantly more preferred than the eye-gaze. For the *overview interaction mode*, the tablet received the lowest (i.e., best) ratings of M=2.12.

Asked about possible combinations, participants included speech (13 times), the tablet (12), the joystick (11), the eye-gaze (8), and gestures (6). In the text field, the participants described the envisioned interactions. Joystick and tablet were envisioned as standalone. The other combinations included eye-gaze and gesture to select a location both on the center console or the real world and speech for the selection.

5.7 Speech Overview Commands and Open Feedback

While the interaction was predefined in most concepts, the commands possible in the speech *overview* interaction mode were not restrained as we employed the Wizard-of-Oz protocol. Therefore, we noted the commands given. For the test course, all 16 participants stated some form of "Drive to the excavator". For the study course, the answers were also location-based. Most participants (15 times) stated something along the lines of "Drive to [location] on the right" where the location was either the garage, the driveway, or the house. The other participant gave a more driving-related order: "After 30m turn right". All of the answers indicate that there is a rather high assumed intelligence of the AV.

Participants could also give feedback on the advantages and disadvantages per modality and in general.
For the tablet, especially the necessity of an additional device was seen as a disadvantage. However, the clear concept and the feedback related to the chosen point were seen as benefits.
For the gesture, the intuitive and natural interaction was well received. However, the tiring interaction, as well as technical limitations, were named as drawbacks.
Regarding the interaction with eye-gaze, participants stated that it is not physically tiring. However, as indicated by the preference, the concepts were seen as unsafe as sometimes parts of the field of view are obscured, and a simple glance could lead to accidents.
While the speech was very intuitive, participants named the unclear message requirements and the speech recognition capabilities with an accent as a drawback.
Regarding the joystick, participants highlighted the ease of use, the known interaction, and the "fun factor" [P1]. The necessity to use a hand and the imprecise input were seen as drawbacks.

6 DISCUSSION

We presented 15 interaction concepts using the modalities gesture, eye-gaze, speech, tablet, joystick, and the steering wheel as a baseline. Our VR study with N=16 participants revealed the usability, required workload, and accuracy of the concepts. The steering wheel was rated most reasonable. However, other interaction modalities also lead to high usability ratings, especially in the *taxi-like* mode.

6.1 Steering Wheel Predominance

Our results indicate that participants preferred the steering wheel. We attribute this to the familiarity participants had with this interaction modality. While there are both *familiarity* and *novelty* preferences [35], today, it is required to take several lessons to receive a driver's license. Therefore, users are very accustomed to this device. While participants were able to test the concept in a test course, this exposure was rather short. We expect the preference differences to become less with more exposure time. Nonetheless, this short exposure is beneficial for the external validity of our study as such interventions will most likely be scarce and short. This should be evaluated in a further study.

6.2 Abstraction Level of Vehicle Steering

The different *interaction modes direct*, *taxi-like*, and *overview* require different levels of sophistication of an AV. While the passenger takes over full control in the *direct* mode and, therefore, the system does not have to be involved in the driving task anymore, in the other two modes, the AV still at least has to steer the vehicle, recognize intersections, and avoid collisions. In the *overview* mode, the AV additionally has to be able to derive appropriate parking spots. Therefore, we assumed it likely that the usability and demand were best with less required interaction. However, the *interaction mode taxi-like* was rated as most usable (see Figure 7) and least demanding (see Figure 8a). This is interesting as the *overview* concepts were developed with the assumption that they would be easiest to use. Based on the open feedback, we assume that this is due to the unknown interaction and the potentially non-immediate feedback. For example, when describing the location in the *overview* mode, the passenger has to trust that the vehicle correctly understood the command. In the *direct* and *taxi-like* modes, the feedback is immediate as the vehicle will alter its course immediately. System transparency was shown to increase trust in AVs [10, 11, 13, 14, 17, 33] and we believe that the transparency is relevant for the perceived usability.

6.3 Opportunities, Drawbacks, and Multimodality of Novel Interaction Concepts

As the steering wheel was shown to be most reasonable (see Table 3), the question arises why novel *interaction modalities* could be useful and how they could generate opportunities. Already, post-automation effects negatively influence driving performance have been shown in the context of takeovers [5, 19, 20, 38, 46]. With further automation, passengers of AVs will become less skilled in using steering wheels. Therefore, we believe that more intuitive interaction concepts that put more burden on the automation will be beneficial as parts of the automation regarding safety stay active. As we were able to show, already today, some of these concepts are rated highly usable (see Figure 7). Additionally, steering wheels limit the users to non-disabled passengers. However, as AVs could increase mobility for people with disabilities, other interaction concepts for the "final 100 meters problem" can be useful. Also, the usage of seat-independent modalities allows for greater flexibility in the design and usage of AVs. Another potential advantage is the integration of the interaction concept into the already prevalent environment. For example, when using a tablet to watch a movie, the usage of this tablet seems adequate to avoid a context switch. The potential of novel interaction concepts here has not yet been fully realized [31].

Drawbacks lay mostly in the required technical capabilities and the ease of use. Our results showed passengers' concerns about using technology that might be prone to errors such as gesture recognition. For these driving-relevant tasks, high accuracy is necessary. Also, the interaction must be designed highly intuitive as such interactions could be scarce.

We have designed each interaction concept only using one modality. For example, the gesture *overview* concept used a fist to select the location indicated with the other hand. Here, multimodal interaction should probably be used, as also indicated by the participants. The most promising combinations for this have to be found.

6.4 Technical Considerations and Practical Implications

For the driving-relevant interaction, high accuracy in recognizing gestures, speech, and eye-gaze is necessary. While these are already rather good (hand recognition [41], speech recognition approximately 98% [67], and 1 degree in automotive eye-tracking use cases [32]), the applicability for the entire spectrum of potential users has to be considered.

Practically, our data showed that the highest sophisticated designs are not necessary to solve the "final 100 meters problem". Our data show that the *taxi-like* approach is seen as highly usable. This is in line with previous work showing that cooperation with an AV is feasible and usable [63].

The exposure time regarding the interaction concepts was low. While we do believe that with sufficient practice, some of these will become more usable, we argue that in future automated traffic, the actual required number of interactions will be low and short. Therefore, the design of our study actually benefits from an externally valid comparison as users will not be well-trained if they have to use these concepts.

6.5 Limitations

The number of participants in the study was of moderate size (N=16). As mostly younger male participants (on average 25.63 years old, only four women) took part, it is unclear whether this work's findings are transferable to other age groups. However, we believe the preference towards the steering wheel will be even more prominent in an older sample due to the increased experience with it. Regarding the chosen interaction concepts, we selected commonly found input modalities. However, as these have specific strengths and weaknesses, comparability for all the *interaction modes* was difficult. This can be seen in the *overview* implementation of the gesture and the eye-gaze.

Based on the pre-study with five participants, however, we opted for the more ecologically valid interaction leveraging the specific strengths of the modalities.

Finally, the implementation of the concepts was not perfect. While we used state-of-the-art eye-tracking and gesture recognition, especially the gesture recognition did not perform perfectly. Therefore, the results might not be perfectly transferable to real-world applications. However, these technical limitations apply here. Regarding the usage of a VR setup, especially the usage of the tablet was difficult. The tablet was heavier than usual due to the necessity to employ an HTC Vive tracker. Additionally, the mapping of the recognized hand and the subsequent potential misalignment of VR and the real world could have negatively influenced the evaluation. Furthermore, immersion and realism of the interaction could be enhanced by employing simulators with more degrees of freedom (e.g., [12] or [29]).

The duration measurements were influenced by the dwell times for blinking (1s) and tablet *overview* (5s). Therefore, these values are directly dependent on the implementation and their values have to be interpreted accordingly.

7 CONCLUSION

Overall, this work presented a concept classification of interaction concepts targeting AVs. Based on this classification, 15 interaction concepts were designed, implemented, and evaluated. As a baseline, a steering wheel was used. The results of the VR study with N=16 participants showed high usability of the *taxi-like interaction mode* in general and especially for the joystick, speech, and eye-gaze. This work provides insights for AVs to be successfully introduced into traffic with a special focus on the "final 100 meters problem" even without perfect driving capabilities.

ACKNOWLEDGMENTS

The authors thank all study participants. This work was conducted within the project 'SEMULIN' funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK).

REFERENCES

[1] Brian Andonian, William Rauch, and Vivek Bhise. 2003. Driver steering performance using joystick vs. steering wheel controls. *SAE transactions* (2003), 1–12.

[2] Jacopo M Araujo, Guangtao Zhang, John Paulin Paulin Hansen, and Sadasivan Puthusserypady. 2020. Exploring eye-gaze wheelchair control. In *ACM Symposium on Eye Tracking Research and Applications*. ACM, New York, NY, USA, 1–8.

[3] Mercedes Benz. 2015. Der Mercedes-Benz F 015 Luxury in Motion. https://www.mercedes-benz.com/de/innovation/forschungsfahrzeug-f-015-luxury-in-motion/. [Online; accessed: 07-DECEMBER-2019].

[4] Mercedes Benz. 2021. Inspired by the future: The VISION AVTR. https://www.mercedes-benz.com/en/vehicles/passenger-cars/mercedes-benz-concept-cars/vision-avtr/. [Online; accessed: 07-AUGUST-2021].

[5] S. Brandenburg and E. M. Skottke. 2014. Switching from manual to automated driving and reverse: Are drivers behaving more risky after highly automated driving?. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, New York, NY, USA, 2978–2983. https://doi.org/10.1109/ITSC.2014.6958168

[6] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[7] Géry Casiez, Nicolas Roussel, and Daniel Vogel. 2012. 1 € Filter: A Simple Speed-Based Low-Pass Filter for Noisy Input in Interactive Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2527–2530. https://doi.org/10.1145/2207676.2208639

[8] Włodzimierz Choromański, Iwona Grabarek, and Maciej Kozłowski. 2019. Research on an innovative multifunction steering wheel for individuals with reduced mobility. *Transportation research part F: traffic psychology and behaviour* 61 (2019), 178–187.

[9] Mark Colley, Ali Askari, Marcel Walch, Marcel Woide, and Enrico Rukzio. 2021. ORIAS: On-The-Fly Object Identification and Action Selection for Highly Automated Vehicles. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 79–89. https://doi.org/10.1145/3409118.3475134

[10] Mark Colley, Christian Bräuner, Mirjam Lanzer, Walch Marcel, Martin Baumann, and Enrico Rukzio. 2020. Effect of Visualization of Pedestrian Intention Recognition on Trust and Cognitive Load. In *Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '20)*. ACM, Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3409120.3410648

[11] Mark Colley, Benjamin Eder, Jan Ole Rixen, and Enrico Rukzio. 2021. Effects of Semantic Segmentation Visualization on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445351

[12] Mark Colley, Pascal Jansen, Enrico Rukzio, and Jan Gugenheimer. 2022. SwiVR-Car-Seat: Exploring Vehicle Motion Effects on Interaction Quality in Virtual Reality Automated Driving Using a Motorized Swivel Seat. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 150 (dec 2022), 26 pages. https://doi.org/10.1145/3494968

[13] Mark Colley, Svenja Krauss, Mirjam Lanzer, and Enrico Rukzio. 2021. How Should Automated Vehicles Communicate Critical Situations? A Comparative Analysis of Visualization Concepts. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 94 (Sept. 2021), 23 pages. https://doi.org/10.1145/3478111

[14] Mark Colley, Max Rädler, Jonas Glimmann, and Enrico Rukzio. 2022. Effects of Scene Detection, Scene Prediction, and Maneuver Planning Visualizations on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 49 (jul 2022), 21 pages. https://doi.org/10.1145/3534609

[15] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies—why and how. *Knowledge-based systems* 6, 4 (1993), 258–266.

[16] Henrik Detjen, Sarah Faltaous, Stefan Geisler, and Stefan Schneegass. 2019. User-Defined Voice and Mid-Air Gesture Commands for Maneuver-Based Interventions in Automated Vehicles. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC'19)*. Association for Computing Machinery, New York, NY, USA, 341–348. https://doi.org/10.1145/3340764.3340798

[17] Na Du, Jacob Haspiel, Qiaoning Zhang, Dawn Tilbury, Anuj K. Pradhan, X. Jessie Yang, and Lionel P. Robert. 2019. Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload. *Transportation Research Part C: Emerging Technologies* 104 (2019), 428–442. https://doi.org/10.1016/j.trc.2019.05.025 ID: 271729.

[18] Joep Eijkemans. 2019. Motion sickness in a Virtual Reality cycling simulation. http://essay.utwente.nl/78690/

[19] Alexander Eriksson and Neville A. Stanton. 2017. Takeover Time in Highly Automated Vehicles: Noncritical Transitions to and From Manual Control. *Human Factors* 59, 4 (2017), 689–705. https://doi.org/10.1177/0018720816685832 arXiv:https://doi.org/10.1177/0018720816685832 PMID: 28124573.

[20] Alexander Eriksson and Neville A Stanton. 2017. Takeover time in highly automated vehicles: noncritical transitions to and from manual control. *Human factors* 59, 4 (2017), 689–705.

[21] Daniel J Fagnant and Kara Kockelman. 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice* 77 (2015), 167–181.

[22] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods* 41, 4 (2009), 1149–1160.

[23] Finward Studios. 2021. *Suburb Neighborhood House Pack (Modular)*. Finward Studios. https://assetstore.unity.com/packages/3d/environments/urban/suburb-neighborhood-house-pack-modular-72712

[24] David C. Funder and Daniel J. Ozer. 2019. Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science* 2, 2 (2019), 156–168. https://doi.org/10.1177/2515245919847202 arXiv:https://doi.org/10.1177/2515245919847202

[25] Markus Funk, Vanessa Tobisch, and Adam Emfield. 2020. *Non-Verbal Auditory Input for Controlling Binary, Discrete, and Continuous Input in Automotive User Interfaces*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376816

[26] Guavaman Enterprises. 2021. *Rewired (version 1.1.37.0.U2019)*. Guavaman Enterprises. https://assetstore.unity.com/packages/tools/utilities/rewired-21676

[27] John Paulin Hansen, Alexandre Alapetite, Martin Thomsen, Zhongyu Wang, Katsumi Minakata, and Guangtao Zhang. 2018. Head and Gaze Control of a Telepresence Robot with an HMD. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (Warsaw, Poland) *(ETRA '18)*. Association for Computing Machinery, New York, NY, USA, Article 82, 3 pages. https://doi.org/10.1145/3204493.3208330

[28] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, Amsterdam, The Netherlands, 139–183.

[29] Philipp Hock, Mark Colley, Ali Askari, Tobias Wagner, Martin Baumann, and Enrico Rukzio. 2022. Introducing VAMPIRE - Using Kinaesthetic Feedback in Virtual Reality for Automated Driving Experiments. In *14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '22)*. Association for Computing Machinery, New York, NY, USA, 9 pages. https://doi.org/10.1145/3543174.3545252 Accepted.

[30] HTC Corporation. 2021. *Vive Hand Tracking SDK (version 1.0.0)*. HTC Corporation. https://developer.vive.com/resources/vive-sense/sdk/vive-hand-tracking-sdk/

[31] Pascal Jansen, Mark Colley, and Enrico Rukzio. 2022. A Design Space for Human Sensor and Actuator Focused In-Vehicle Interaction Based on a Systematic Literature Review. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 56 (jul 2022), 51 pages. https://doi.org/10.1145/3534617

[32] Anuradha Kar and Peter Corcoran. 2017. A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms. *IEEE Access* 5 (2017), 16495–16519.

[33] Jeamin Koo, Jungsuk Kwac, Wendy Ju, Martin Steinert, Larry Leifer, and Clifford Nass. 2015. Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 9, 4 (2015), 269–275.

[34] Key Jung Lee, Yeon Kyoung Joo, and Clifford Nass. 2014. Partially Intelligent Automobiles and Driving Experience at the Moment of System Transition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3631–3634. https://doi.org/10.1145/2556288.2557370

[35] Hsin-I Liao, Su-Ling Yeh, and Shinsuke Shimojo. 2011. Novelty vs. familiarity principles in preference decisions: task-context of past experience matters. *Frontiers in psychology* 2 (2011), 43.

[36] Google LLC. 2021. Google Scholar's top HCI venues and publications. https://scholar.google.com/citations?view_op=top_venues&hl=de&vq=eng_humancomputerinteraction (Accessed: August 2021).

[37] Haiko Lüpsen. 2020. R-Funktionen zur Varianzanalyse. http://www.uni-koeln.de/~luepsen/R/. [Online; accessed 25-SEPTEMBER-2020].

[38] Natasha Merat, A. Hamish Jamson, Frank C.H. Lai, Michael Daly, and Oliver M.J. Carsten. 2014. Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour* 27 (2014), 274 – 282. https://doi.org/10.1016/j.trf.2014.09.005

[39] David Moher, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G Altman. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 151, 4 (2009), 264–269.

[40] Brian Mok, Mishel Johns, Stephen Yang, and Wendy Ju. 2017. Reinventing the Wheel: Transforming Steering Wheel Systems for Autonomous Vehicles. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 229–241. https://doi.org/10.1145/3126594.3126655

[41] Munir Oudah, Ali Al-Naji, and Javaan Chahl. 2020. Hand gesture recognition based on computer vision: a review of techniques. *journal of Imaging* 6, 8 (2020), 73.

[42] Indrajeet Patil. 2021. Visualizations with statistical details: The 'ggstatsplot' approach. *Journal of Open Source Software* 6, 61 (2021), 3167. https://doi.org/10.21105/joss.03167

[43] Bastian Pfleging, Maurice Rang, and Nora Broy. 2016. Investigating User Needs for Non-Driving-Related Activities during Automated Driving. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia* (Rovaniemi, Finland) *(MUM '16)*. Association for Computing Machinery, New York, NY, USA, 91–99. https://doi.org/10.1145/3012709.3012735

[44] Polarith. 2021. *Polarith AI Pro | Movement with 3D Sensors (version 1.7.1)*. Polarith. https://assetstore.unity.com/packages/tools/ai/polarith-ai-pro-movement-with-3d-sensors-71465

[45] Xiaosong Qian, Wendy Ju, and David Michael Sirkin. 2020. Aladdin's magic carpet: Navigation by in-air static hand gesture in autonomous vehicles. *International Journal of Human–Computer Interaction* 0, 0 (2020), 1–16. https://doi.org/10.1080/10447318.2020.1801225 arXiv:https://doi.org/10.1080/10447318.2020.1801225

[46] Fabienne Roche and Stefan Brandenburg. 2018. Should the urgency of auditory-tactile takeover requests match the criticality of takeover situations?. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, IEEE, New York, NY, USA, 1035–1040.

[47] Felix Ros, Jacques Terken, Frank van Valkenhoef, Zane Amiralis, and Stefan Beckmann. 2018. Scribble Your Way Through Traffic. In *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Toronto, ON, Canada) *(AutomotiveUI '18)*. Association for Computing Machinery, New York, NY, USA, 230–234. https://doi.org/10.1145/3239092.3267849

[48] Robert Rosenthal, Harris Cooper, and L Hedges. 1994. Parametric measures of effect size. *The handbook of research synthesis* 621, 2 (1994), 231–244.

[49] Sonja Rümelin, Chadly Marouane, and Andreas Butz. 2013. Free-Hand Pointing for Identification and Interaction with Distant Objects. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Eindhoven, Netherlands) *(AutomotiveUI '13)*. Association for Computing Machinery, New York, NY, USA, 40–47. https://doi.org/10.1145/2516540.2516556

[50] Clemens Schartmüller, Andreas Riener, and Philipp Wintersberger. 2018. Steer-by-wifi: Lateral vehicle control for take-overs with nomadic devices. In *Adjunct proceedings of the 10th international conference on automotive user interfaces*

and interactive vehicular applications. ACM, New York, NY, USA, 121–126.
[51] Brian A. Smith and Shree K. Nayar. 2018. *The RAD: Making Racing Games Equivalently Accessible to People Who Are Blind*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174090
[52] Sophie Stellmach and Raimund Dachselt. 2012. Designing Gaze-Based User Interfaces for Steering in Virtual Environments. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 131–138. https://doi.org/10.1145/2168556.2168577
[53] SAE Taxonomy. 2014. *Definitions for terms related to on-road motor vehicle automated driving systems*. Technical Report. Technical report, SAE International.
[54] Dušan Teodorović and Milan Janić. 2017. Chapter 11 - Transportation, Environment, and Society. In *Transportation Engineering*, Dušan Teodorović and Milan Janić (Eds.). Butterworth-Heinemann, 719–858. https://doi.org/10.1016/B978-0-12-803818-5.00011-1
[55] Tobii. 2021. Tobii G2OM. https://vr.tobii.com/sdk/solutions/tobii-g2om. [Online; accessed: 24-JULY-2021].
[56] Robert Tscharn, Marc Erich Latoschik, Diana Löffler, and Jörn Hurtienne. 2017. "Stop over There": Natural Gesture and Speech Interaction for Non-Critical Spontaneous Intervention in Autonomous Driving. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (Glasgow, UK) *(ICMI '17)*. Association for Computing Machinery, New York, NY, USA, 91–100. https://doi.org/10.1145/3136755.3136787
[57] Unity. 2021. Unity Phrase Recognition System. https://docs.unity3d.com/ScriptReference/Windows.Speech.PhraseRecognitionSystem.html. [Online; accessed: 12-AUGUST-2021].
[58] Unity Technologies. 2021. *Unity*. Unity Technologies. https://unity.com/
[59] Julius von Willich, Dominik Schön, Sebastian Günther, Florian Müller, Max Mühlhäuser, and Markus Funk. 2019. VRChairRacer: Using an Office Chair Backrest as a Locomotion Technique for VR Racing Games. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3290607.3313254
[60] Marcel Walch, Mark Colley, and Michael Weber. 2019. CooperationCaptcha: On-the-fly object labeling for highly automated vehicles. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–6.
[61] Marcel Walch, Tobias Sieber, Philipp Hock, Martin Baumann, and Michael Weber. 2016. Towards Cooperative Driving: Involving the Driver in an Autonomous Vehicle's Decision Making. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Ann Arbor, MI, USA) *(Automotive'UI 16)*. Association for Computing Machinery, New York, NY, USA, 261–268. https://doi.org/10.1145/3003715.3005458
[62] Marcel Walch, Marcel Woide, Kristin Mühl, Martin Baumann, and Michael Weber. 2019. Cooperative Overtaking: Overcoming Automated Vehicles' Obstructed Sensor Range via Driver Help. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 144–155. https://doi.org/10.1145/3342197.3344531
[63] Marcel Walch, Marcel Woide, Kristin Mühl, Martin Baumann, and Michael Weber. 2019. Cooperative Overtaking: Overcoming Automated Vehicles' Obstructed Sensor Range via Driver Help. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications*. ACM, New York, NY, USA, 144–155.
[64] Chengshi Wang, Kim Alexander, Philip Pidgeon, and John Wagner. 2019. *Use of Cellphones as Alternative Driver Inputs in Passenger Vehicles*. Technical Report. SAE Technical Paper.
[65] Chao Wang, Matti Krüger, and Christiane B. Wiebel-Herboth. 2020. "Watch out!": Prediction-Level Intervention for Automated Driving. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) *(AutomotiveUI '20)*. Association for Computing Machinery, New York, NY, USA, 169–180. https://doi.org/10.1145/3409120.3410652
[66] Gesa Wiegand, Kai Holländer, Katharina Rupp, and Heinrich Hussmann. 2020. The Joy of Collaborating with Highly Automated Vehicles. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) *(AutomotiveUI '20)*. Association for Computing Machinery, New York, NY, USA, 223–232. https://doi.org/10.1145/3409120.3410643
[67] Linlin Xia, Gang Chen, Xun Xu, Jiashuo Cui, and Yiping Gao. 2020. Audiovisual speech recognition: A review and forecast. *International Journal of Advanced Robotic Systems* 17, 6 (2020), 1729881420976082.
[68] Guangtao Zhang and John Paulin Hansen. 2020. People with Motor Disabilities Using Gaze to Control Telerobots. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3382939
[69] Guangtao Zhang, John Paulin Hansen, and Katsumi Minakata. 2019. Hand- and Gaze-Control of Telepresence Robots. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 70, 8 pages. https://doi.org/10.1145/3317956.3318149

Received February 2022; revised May 2022; accepted June 2022