Effects of Urgency and Cognitive Load on Interaction in Highly Automated Vehicles

MARK COLLEY, Institute of Media Informatics, Ulm University, Germany
CRISTINA EVANGELISTA, Institute of Media Informatics, Ulm University, Germany
TITO DAZA RUBIANO, Institute of Media Informatics, Ulm University, Germany
ENRICO RUKZIO, Institute of Media Informatics, Ulm University, Germany

Fig. 1. Multimodal interface including touch (a), gaze (b), pointing (c), and voice inputs to refer to objects in and outside a vehicle.

In highly automated vehicles, passengers can engage in non-driving-related activities. Additionally, the technical advancement allows for novel interaction possibilities such as voice, gesture, gaze, touch, or multimodal interaction, both to refer to in-vehicle and outside objects (e.g., thermostat or restaurant). This interaction can be characterized by levels of urgency (e.g., based on late detection of objects) and cognitive load (e.g., because of watching a movie or working). Therefore, we implemented a Virtual Reality simulation and conducted a within-subjects study with N=11 participants evaluating the effects of urgency and cognitive load on modality usage in automated vehicles. We found that while all modalities were possible to use, participants relied on touch the most. This was followed by gaze, especially for external referencing. This work helps to further understand multimodal interaction and the requirements this poses on natural interaction in (automated) vehicles.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: Interaction design; automated vehicles; multimodal.

ACM Reference Format:
Mark Colley, Cristina Evangelista, Tito Daza Rubiano, and Enrico Rukzio. 2023. Effects of Urgency and Cognitive Load on Interaction in Highly Automated Vehicles. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 207 (September 2023), 20 pages. https://doi.org/10.1145/3604254

Authors' addresses: Mark Colley, mark.colley@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany; Cristina Evangelista, cristina.evangelista@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany; Tito Daza Rubiano, tito.daza-rubiano@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.

2573-0142/2023/9-ART207
https://doi.org/10.1145/3604254

1 INTRODUCTION

Interaction is expected to alter with the introduction of highly automated vehicles (AVs) [44]. In AVs, a user is not required to intervene in the driving task anymore (i.e., SAE Level 4 or 5). A passenger no longer relevant to the primary driving task (i.e., steering, accelerating, braking) can engage in non-driving related tasks such as reading, watching a movie, or even sleeping [17, 44]. Additionally, with the advance in multimodal sensors in the vehicle as cameras, microphones, or radars, novel interaction techniques and metaphors become feasible [13, 30, 31, 54]. For these interactions, the characteristics of naturalness, intuitiveness, and human likeness become crucial [4, 56]. Additionally, such technology enables novel use cases such as addressing outside objects (i.e., objects outside the AV such as buildings or other road users) with direction-based commands like *"I know back there, there is this little café"* [51, p. 7], simple questions like "What is that?", and still enabling the user to indicate also preferences regarding the driving task (e.g., determining the perfect parking spot [46], modification of the route, the specification of a destination, or changing of the traveling speed [20, 39]). These types of commands can be understood by a multimodal technology without the necessity of unimodal approaches to be very precise (e.g., "What is that in 20 m on the right side close to the tree?´´). Previous works have already evaluated new interaction modalities focusing on a limited number of inputs and specific contexts. Overall, previous studies [2, 25, 50, 51] based in a human-driven car context showed the relevance of new interaction modalities such as eye-gaze and hand-tracking to improve the quality and effectiveness of communication.

While these technologies and these interaction cases are highly probable or can even be seen in current semi-automated vehicles, the way passengers will interact with an AV with all possible interaction techniques was not yet explored.

Therefore, we implemented the voice, gaze, pointing, and touch as input modalities in a Virtual Reality (VR) environment and conducted a within-subjects study with N=11 participants. We designed two scenarios, driving through a city with the objective of referencing different objects or user interfaces in and outside the AV and a scenario in which the effect of urgency (e.g., present in referencing objects outside the vehicle when driving past it) was evaluated. For this, the participant had to reference a building in the line of sight or provide information about driving direction (left or right) in 3, 7, or 10s.

We found that touch, speech, and gaze + touch were preferred and used most frequently. Multimodal approaches were used less both for in- and outside referencing. Participants especially highlighted the need for (visual) feedback if pointing or gaze is used.

In this paper, we will first outline related work, describe the experiment to investigate multimodal interaction, define our interpretation procedure, and report our quantitative and qualitative results. Finally, we discuss our results.

Contribution Statement: Our work provides insights into two scenarios regarding (multimodal) interaction with AVs. The scenarios include referencing objects and user interfaces in and outside the AV and the effects of urgency. Results of a VR study with N=11 participants showed that touch, voice, and gaze + touch were preferred and that, especially for gaze and pointing, (visual) feedback is required. Our work helps guide developers and researchers towards useful and usable multimodal interactions in AVs.

2 RELATED WORK

This work builds on research in interaction modalities in manual and AVs. The interaction modality should satisfy four requisites: *spatial accuracy*, *intuitiveness*, *wide range of possible maneuvers* and *feasibility* [56]. Similar parameters were defined Ataya et al. [4]: *easy to use*, *satisfactory*, *naturalistic*,

controlling, and *useful*. Especially in a dynamic context such as a driving vehicle, the modalities should allow real-time interaction with the external environment [25].

Touch panels for drivers were evaluated to select appropriate maneuvers when reaching an automation limit [10, 58, 59] or to select AV maneuvers (e.g., lane changes) [32]. The position varied: they were attached on the steering wheel [21, 35, 43] or in the center console [3, 42, 51].

The maneuver-based intervention was also achieved via hand gestures [18, 46]. Hand gestures were also used to control AV motions [16, 40, 46]. Free-point gestures [51] and hand-constrained gestures [23] were also used for input.

Eye gaze was used unimodal to reference or select objects [41, 45, 50].

Voice was used to support driver-vehicle cooperation and to select maneuvers [4]. Others studied voice input for the selection of in-vehicle objects [41, 50, 53]. However, voice recognition does not necessarily work adequately in noisy environments, and drivers may be confused about possible commands [7, 18]. For example, Qian et al. [46] report that gestures were favored compared to voice because of the minor influence of recognition failures, noise, and confusion between commands and dialogue context.

Multimodal approaches combined, for example, gaze to localize and hand gestures to coordinate pointing [33, 48]. The advantages of the fusion, such as compensation of the single inputs drawbacks, motivated the work of Gomaa et al. [25], who analyzed the subject's pointing and gaze behavior during the drive. They found that in automated driving (compared to manual driving), gaze accuracy was significantly higher but pointing accuracy showed no significant differences between driving modes. They also found that, in general, gaze accuracy was significantly higher than pointing accuracy. Gomaa et al. [25] acknowledged that a real vehicle could have some influence on these metrics based on vehicle boundaries and movements. Therefore, we employed a vehicle mockup but could not simulate movement. In a later work, Gomaa et al. [26] proposed *ML-PersRef*, a personalized machine learning technique to reference objects inside and outside of a moving vehicle. In line with this work, Aftab and von der Beeck [1] use multiple modalities (eye-gaze, head, and finger as well as a voice command to separate interactions), to reference inside and outside objects. They showed that while finger (91.6%) and eye gaze (96.3%) alone achieve high accuracy, finger plus eye (98.1%) and eye plus head + finger (98.6%) outperform these [1].

Multimodal approaches, however, are still affected by numerous challenges such as alignment, translation, representation, and co-learning [5, 6].

While the multimodal approaches showed benefits, it is still unclear how passengers will use the available modalities. The online video-based survey of Ataya et al. [4] addressed this issue, observing which interactions the users chose while performing different non-driving related tasks and which factors could influence this choice. Voice, touch, hand gesture, gaze, and their combination were proposed to resolve maneuver-based (lateral control, longitudinal control, stopping the vehicle, rerouting the navigation) and nonmanoeuvre-based interventions (request information and playing a video) [4] under different cognitive and physical workload such as relaxing, eating, working on a laptop and watching a video. Voice was rated best using the ranking parameters easy to use, satisfaction, naturalness, and usefulness. However, users of the online study were only able to provide their suggestions and impression based on the online videos. Therefore, the study lacks external validity.

Fig. 2. With correct pointing or selection via eye-gaze, the bullet point is green, otherwise red.

Fig. 3. Obscured windshield during the switch between each run in the urgency scenario.

3 EXPERIMENT

3.1 Study Design

Our study focused on interaction in an AV that performs the driving task. To evaluate which and how modalities are used, we designed and implemented a within-subjects study with two scenarios: the "Driving" and the "Urgency" scenario.

In the driving scenario, the participants could freely use every available interaction modality while driving through a city. In this scenario, the cognitive load was altered by displaying a video. Thus, this scenario was experienced twice. The driving scenario aimed at the exploratory research question (RQ):

> *RQ1: What impact does the independent variable "cognitive load" have on passengers in terms of (1) modality usage, (2) task load, (3) usability, and (4) trust?*

In the urgency scenario, we altered the urgency of the interaction by displaying a countdown timer with 3, 7, and 10 seconds duration. We define urgency as the feeling that a given task must be carried out quickly. These timings were chosen as they represent realistic distances for AV users to interact with objects (e.g., assuming 50km/h leads to $\approx$ 41m in 3s, 97m in 7s, and 138m in 10s) and induced sufficient stress in internal pre-trials. In this task, the vehicle remained still to be able to study the effect of urgency independent of the driving. Additionally, we altered the task. Participants either had to ask for the height of a building or indicate where the AV should head towards. Therefore, the urgency scenario followed a 3×2 design. The urgency experiment was performed as the last condition when the users gained more experience with the system. Before each task, the vehicle's windshield was obscured (see Figure 3) to minimize the participant's possibility of preparation. The urgency scenario aimed at the exploratory RQ:

> *RQ2: What impact do the independent variables "urgency" and "task" have on passengers in terms of modality usage?*

3.2 Materials

To investigate the four input modalities, i.e., eye-gaze, pointing, touch, and voice, we implemented a driving course in Unity version 2020.3.19f1.

3.2.1 Interaction Concepts. Our multimodal system provides the users with five input possibilities: touch, voice, voice + gaze, voice + pointing, and voice + pointing + gaze. For voice interaction, we used the Google Speech Recognition Asset [24] in version 4.1.

In the case of pointing and gaze being used, the voice was used to confirm the selection of the object (i.e., as a trigger for the user's activity). When the system detects the combined use of gaze

and pointing, we measured the angle between the ray originated by the eye gaze and the pointing direction line (see Figure 4).

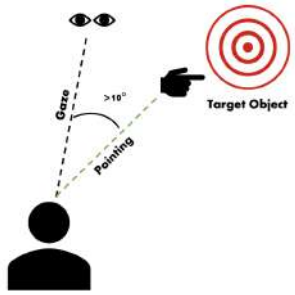


Fig. 4. If the system detects an input from the gaze and pointing modalities, and the direction's vectors have a distance greater than 10°, the pointed object will be selected for the commands.

Based on the findings of Gomaa et al. [25], we considered an offset of 10° between the target POI and the user's gaze. Then, in the case of an angle greater than 10°, the pointed object will be selected. Otherwise, the input will be defined by the user's eye gaze. This behavior is in line with the algorithm of Roider et al. [50], defining that if the distance between two objects of interest is more than one element's position, the gaze does not refer to the target, and the pointing should be taken into consideration.

We also integrated a display into the used model of a Mercedes Benz F015 with all logos removed and added a simulated touch screen at the passenger's door. For the *cognitive load* condition, a second screen was added on the driver's side. For this purpose, we augmented the used virtual hands with an additional collider placed on the top of the right index finger. Thus, no real touch interaction with a physical surface was necessary. While we aligned the virtual dashboard with the physical chassis of the mockup (see Figure 5), this was not always perfect. Thus, haptic feedback was not always possible. It should be noted that for pointing, our implementation included only the detection of right-handed inputs defined with the index finger.

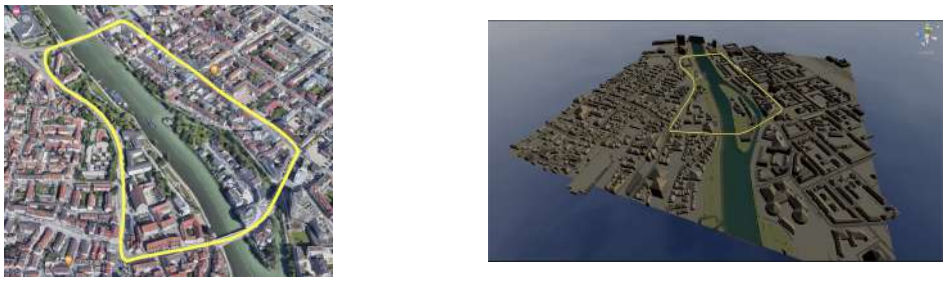
3.2.2 Apparatus. We used the HTC Vive Pro Eye VR headset with its included Tobii eye tracker. The participant's hands needed to perform the pointing action were reproduced in the VR environment using a Leap Motion Controller [57] (version 5.0.0) mounted to the Vive headset.
Finally, to enhance the perceived realism of a drive in a real vehicle and to also accommodate for the physical boundaries that limit gesture interaction, we conducted our study in the driving simulator of the Human Factors department at Ulm University. As this vehicle mockup does not contain physical doors, we added a plastic panel as the door (see Figure 5).

3.3 Scenarios

To analyze the user's preferences on the interactions, participants drove in an AV along the streets of Ulm, generated using CityGen3D [55] (version 1.05), building the environment on real-world data. Figure 6 shows the reproduction of Ulm compared to the existing one; the yellow path, long around 2 km, represents the route used for the learning drive. Below is the scene used for the *Baseline* and *cognitive load* conditions (see Figure 7), covering about 3 km in the real world.

3.3.1 Driving Scenario. In the *cognitive load* condition, participants watched a video due to its moderate but externally valid mental workload required to perform this task [4]. We decided to

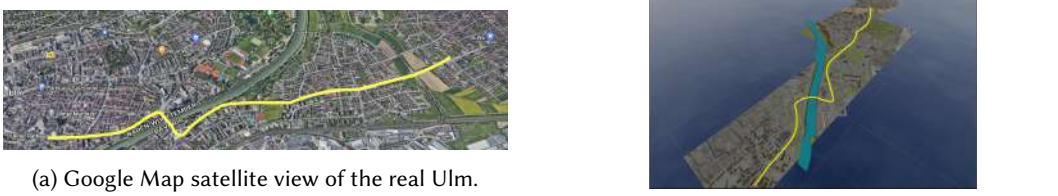
Fig. 5. Participants pointing to a building. Driving simulator of the Human Factors department at Ulm University with a transparent plastic panel.



(a) Google Map satellite view of the real Ulm.

(b) Bird's-eye view of the generated city.

Fig. 6. Comparison between the real and virtual city of Ulm used for the familiarization drive.



(a) Google Map satellite view of the real Ulm.

(b) Bird's-eye view of the generated city.

Fig. 7. Comparison between the real and virtual city of Ulm used for the base and cognitive drive.

reproduce a neutral video provided by a German national news channel[1]. The video was played on the left display of the car. The same route was traveled in the control condition but with 12 different points of interest. The tasks were displayed on the right screen with an icon and a short text, showing no reference to any interaction modalities (to avoid priming participants to use specific modalities; see Figure 9). Additionally, the text was read by the voice assistant of the car.

[1] https://www.youtube.com/watch?v=pjQRu7m20QA; Accessed 07.01.2023

(a) The lake.

(b) The parking garage.

(c) The train station.

(d) The TV tower.

(e) An artistic fountain.

(f) The church.

Fig. 8. Example of buildings considered for the interactive tasks.

Fig. 9. On the bottom of the right display was shown the visual and textual task description.

Table 6 and Table 7 give an overview of the task descriptions presented in the baseline and *cognitive load* condition. In both scenarios, the interaction choice is based on the user's preferences. At the end of each driving experience, users were questioned about the studied input modalities. In the case of the cognitive condition, to prove the user's attention, we asked the following questions: "Which vaccine did prime minister Ramelow receive?" and "Which was the color of the jacket worn by the journalist presenting the news program?"

3.3.2 Urgency Scenario. Finally, subjects performed the urgency scenario, including the direction decision and building information tasks (see Table 4) that must be performed within 3, 7, and 10 seconds.

3.4 Measurements

Objective Measurements: We logged the eye-tracking features of the Vive Pro Eye and the Leap Motion gesture controller. For each task, we logged the input modality, the selected object, the execution time, and the employed voice command. Further, in the case of the "Urgency" conditions, we considered the gap between the correct position of the point of interest, i.e., the building and the location selected by the participant. Finally, we recorded videos of the participants.
Subjective Measurements only "Driving" scenario: Regarding task load, we employed the NASA

TLX [27] with the six subscales *Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration Level.* The Post-Study System Usability Questionnaire (PSSUQ) [36] was employed for user satisfaction. The PSSUQ is divided into the overall score, the System Usefulness (SYSUSE), Information Quality (INFOQUAL), and Interface Quality (INTERQUAL). The system's usability was assessed with the System Usability Scale (SUS) [9]. Trust in the developed AV was evaluated with the Trust in Automation questionnaire of Körber [34]. This is divided into the subscales Reliability, Understanding, Familiarity, Intention of Developers, Trust, and Propensity to trust.
Finally, participants commented on the tested system and filled out a basic demographic questionnaire, including questions about their driving and VR experiences. Moreover, users were asked to rank their preferences, from most preferred (4) to least favorite (1), on the input modalities based on the question *"Which of the input modalities provided for the system interactions do you prefer?"*. They were also asked which combinations they preferred.

3.5 Procedure

First, participants were introduced to the study procedure and the VR scene. They were informed about the studied conditions and possible interaction modalities. They then signed informed consent and could adjust the seat. The study commenced with a training session, allowing the participants to familiarize themselves with the vehicle and the proposed interactions for as long as necessary (see Table 5). In this session, participants received feedback. In the case of correct selections, a green bullet point was colored in green, otherwise in red (see Figure 2). This was avoided in the later test runs, as such an indication would require an augmented reality windshield which we did not want to prerequire.

Then, following a counterbalanced order, subjects took part in the "Driving Scenario", including 12 tasks with an equal number of POIs inside and outside the AV. After the 12 tasks in the "Driving" scenario, participants performed the "Urgency" scenario with the six conditions each done once in counterbalanced order. Participants answered the stated questionnaires after both conditions (with and without mental workload) in the "Driving" scenario.

The study took approximately 1h. Participants were compensated with 10€. We conducted the study in German. The hygiene concept for studies regarding COVID-19 (ventilation, disinfection, wearing masks) involving human subjects of our university was applied.

3.6 Participants

The experiment was performed by N=11 participants (6 male, 5 female), on average M=25.18 years old (SD=4.19). One participant did not have a driving license. This was no requirement as future AVs might not require users to have the know-how or be allowed to drive manually. Seven participants were drivers or passengers for less than 7.000 km in the last year, two persons for less than 15.000 km, one for about 25.000-33.000 km, and one for more than 33.000 km. The majority of the subjects (5) were not drivers self, and only one subject was reported for the category "every day", "in workdays", "1-3 times per month", and "1 time per week".
Regarding their current occupation, six subjects were students, four were accomplishing training, and one was employed.

4 RESULTS

4.1 Data Analysis

We independently analyzed the two scenarios "Driving Scenario" and "Urgency Scenario".

Before every statistical test, we checked the required assumptions (normal distribution and homogeneity of variance assumption). When comparing two conditions, t-tests were used for parametric, Wilcoxon Signed Rank tests for non-parametric data. For the factorial analysis of non-parametric data, we used the non-parametric ANOVA (NPAV) by Lüpsen [38]. The alpha level was 0.05. R in version 4.2.3 and RStudio in version 2023.03.0 was used. All packages were up to date in April 2023.

4.2 Scenario "Driving"

4.2.1 NASA TLX. A student's t-test found no significant differences in the total NASA TLX score (t(10)=-1.86, p=0.09).
Mental Demand: A Wilcoxon Signed-Rank test found a significant difference for mental demand. With the video (M=10.64, SD=6.19), mental demand was significantly higher than without (M=6.27, SD=4.45) as per our study design.
Physical Demand: A Wilcoxon Signed-Rank test found no significant difference for physical demand (p=0.14).
Temporal Demand: A Wilcoxon Signed-Rank test found no significant difference for temporal demand (p=0.05).
Performance: A Wilcoxon Signed-Rank test found no significant difference for performance (p=0.42).
Effort: A Wilcoxon Signed-Rank test found no significant difference for physical demand (p=0.42).
Frustration: A student's t-test found no significant difference for frustration (t(10)=-0.94, p=0.37).

4.2.2 Post-Study System Usability Questionnaire - PSSUQ. Overall Score: A student's t-test found no significant difference for the overall PSSUQ score (t(10)=-2.26, p=0.05). With *cognitive load*, the overall score (M=3.16, SD=1.19) was slightly higher than without *cognitive load* (M=2.78, SD=1.25).
System Usefulness: A Wilcoxon Signed-Rank test found a significant difference for system usefulness (p=0.03). With *cognitive load*, system usefulness (M=3.23, SD=1.32) was significantly higher than without (M=2.73, SD=1.49).
Information Quality: A student's t-test found a significant difference for the information quality (t(10)=-2.35, p=0.04). With *cognitive load*, information quality (M=2.97, SD=1.41) was significantly higher than without (M=2.62, SD=1.26).
Interface Quality: A student's t-test found no significant difference for the interface quality (t(10)=-1.52, p=0.16). With *cognitive load*, interface quality (M=4.00, SD=1.30) was slightly higher than without (M=3.55, SD=1.58).

4.2.3 System Usability Scale - SUS. A student's t-test found no significant difference for the SUS score (t(10)=0.77, p=0.46). Values with (M=62.05, SD=22.91) and without (M=64.77, SD=22.51) the video were almost the same and medium.

4.2.4 Trust. One data set had to be excluded for reliability and understanding due to technical issues.
Reliability: A student's t-test found no significant difference for reliability (t(9)=1.49, p=0.17; without: M=3.32, SD=0.73, with *cognitive load*: M=3.06, SD=0.60).
Understanding: A student's t-test found no significant difference for understanding (t(9)=1.41, p=0.19; without: M=3.70, SD=0.86, with *cognitive load*: M=3.41, SD=0.92).
Familiarity: A Wilcoxon Signed-Rank test found no significant difference for familiarity (p=1.00; without: M=3.68, SD=1.06, with *cognitive load*: M=3.64, SD=0.84).
Intention of Developers: A student's t-test found no significant difference for intention of developers (t(10)=-1.00, p=0.34; without: M=3.95, SD=0.91, with *cognitive load*: M=4.14, SD=0.84).

Trust: A student's t-test found no significant difference for trust (t(10)=0.89, p=0.40; without: M=3.50, SD=0.97, with *cognitive load*: M=3.32, SD=1.01).
Propensity to Trust: A Wilcoxon Signed-Rank test found no significant difference for propensity to trust (p=0.23; without: M=3.48, SD=0.70, with *cognitive load*: M=3.21, SD=0.64).

Condition	SpeechBool	In-/Outside	Modality	n
No Cog. Load	No	External	Touch	28
— " —	No	Internal	Touch	72
— " —	Yes	External	Gaze	51
— " —	Yes	External	Gaze (Pointing active)	1
— " —	Yes	External	Pointing (Gaze active)	1
— " —	Yes	External	Speech	2
— " —	Yes	Internal	Gaze	10
— " —	Yes	Internal	Pointing (Gaze active)	3
— " —	Yes	Internal	Speech	21
With Cog. Load	No	External	Touch	62
— " —	No	Internal	Touch	73
— " —	Yes	External	Gaze	52
— " —	Yes	External	Gaze (Pointing active)	1
— " —	Yes	External	Pointing	1
— " —	Yes	External	Pointing (Gaze active)	1
— " —	Yes	External	Speech	1
— " —	Yes	Internal	Gaze	10
— " —	Yes	Internal	Speech	24

Table 1. Employed modalities per condition. The column n represents the number of occurrences for the given interaction combination, e.g., speech + gaze was used 51 times (see row 3) when no cognitive load was induced).

4.2.5 Modality Usage and Task Duration. We summed all the employed modalities as shown in Table 1. For touch, the references to the "Home-Button" were removed. Touch was used most both with and without *cognitive load*. However, for external referencing, gaze was used either more frequently (with no *cognitive load*) or almost equally often (with *cognitive load*). We also plotted the usage count per *cognitive load* with reference to internal and external reference points (e.g., reference to the window for internal or asking for a building for external; see Figure 10). Figure 10 clearly shows that touch or gaze with voice was used mostly both for internal and external referencing. To use touch for external referencing, one could navigate from the navigation page to the attractions page of the dashboard.

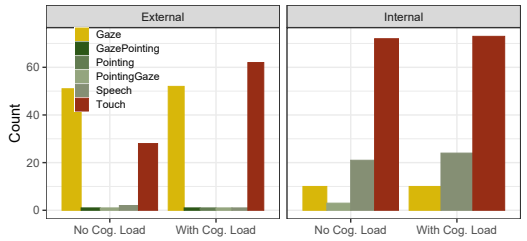


Fig. 10. Modality usage per cognitive load level and in- and outside referencing. GazePointing stands for gaze employed but pointing is also active. PointingGaze stands for pointing employed and gaze is also active.

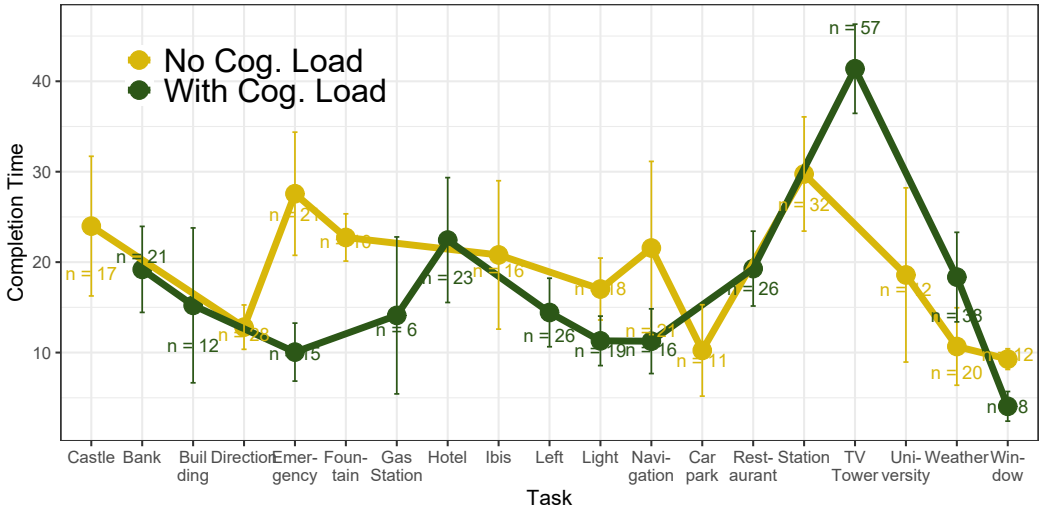


Fig. 11. Completion time for the different tasks.

Regarding the interaction duration, we logged all interactions (see Figure 11). These were only partially available in both rides to avoid learning effects. Therefore, we could not include the task in the statistical analysis. Regarding the *cognitive load*, the NPAV found no significant effects on completion time (with: *cognitive load* M=21.34, SD=17.10; without: M=19.55, SD=14.74).

4.3 Scenario "Urgency"

Task	Time	Voice	Modality	n
Building	3	No		4
— " —	3	Yes	Gaze	5
— " —	7	No		1
— " —	7	No	Touch	1
— " —	7	Yes	Gaze	7
— " —	10	No	Touch	1
— " —	10	Yes	Gaze	8
Direction	3	No		2
— " —	3	No	Touch	3
— " —	3	Yes	Pointing	1
— " —	3	Yes	Voice	3
— " —	7	No		1
— " —	7	No	Pointing	1
— " —	7	No	Touch	1
— " —	7	Yes	Gaze	2
— " —	7	Yes	Pointing	1
— " —	7	Yes	Voice	3
— " —	10	No	Touch	3
— " —	10	Yes	Gaze	2
— " —	10	Yes	Pointing	2
— " —	10	Yes	Voice	2

Table 2. Employed modalities per task and time interval. In yellow, we record the instances where no interaction occurred. n stands for the number of occurences.

4.3.1 Modality, Duration, Gap. A two-way repeated-measures ANOVA was performed to evaluate the effect of the tasks and time intervals on completion time. There was a statistically significant effect of task on completion time ($F(1,1) = 403.76$, p=0.03; see Figure 12). For the building, the duration was significantly shorter (M=0.53, SD=2.70) than for the direction (M=2.12, SD=3.40). In

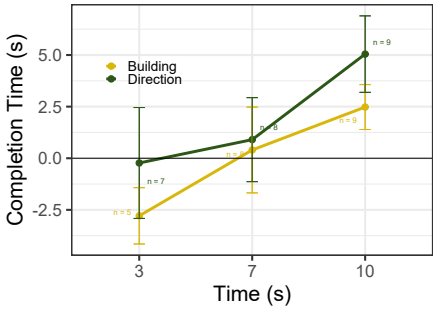


Fig. 12. Completion time. Negative values indicate an entered command after the end of the interval but prior to the next task.

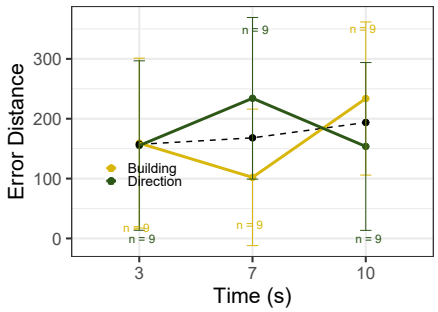


Fig. 13. Error distance in Unity unit.

five trials of the building task and 3 of the direction task, participants were not able to provide information in time. Six of these happened in the 3s time interval and 2 in the 7s time interval.

The NPAV found no significant effects on error distance (see Figure 13).

4.4 Reasonability Assessment and General Remarks

Variable	n	Min	q_1	$\tilde{x}$	$\bar{x}$	q_3	Max	s	IQR
Gesture	11	1	2.5	5	4.4	6.0	7	2.0	3.5
Gaze	11	2	5.0	5	5.5	6.5	7	1.4	1.5
Voice	11	5	5.0	6	5.9	7.0	7	0.9	2.0
Touch	11	3	5.0	6	5.6	7.0	7	1.5	2.0

Table 3. Table of reasonability of the interaction modalities. n stands for the number of participants rating each interaction modality.

We asked participants about their impressions of using different modalities. The gesture received the lowest ratings. Voice was rated best, closely followed by touch and gaze (see Table 3).

After all conditions, we asked participants about their assessment of the advantages and disadvantages of each of the modalities and how these could be improved. Most improvement proposals focused on the technical limitations of the systems, that is gesture, voice, and eye gaze detection. As we employed state-of-the-art software and hardware, this indicates that further improvements seem necessary here. Regarding the interaction modalities' general applicability, most agreed that besides voice (which was rated best) and also touch, gestures and eye gaze could become viable options (e.g., "if you look at the building any way you can get information directly", "[gesture] easy to use"). However, the participants stressed that visualization is necessary ("When looking or pointing, show a point on, e.g., the target, which the system has recognized as focus.") as one participant stated, "It is difficult to see if the system correctly recognizes the gesture and when pointing, recognizes the correct, e.g., building."

5 DISCUSSION

We designed two scenarios to evaluate (1) the effect of cognitive load on interaction modalities in AVs and (2) the effect of urgency. In a within-subjects study with N=11 participants, we found, in line with previous work [4, 19], that touch and voice-only were preferred, with gaze being used for external referencing. However, while other previous work suggested and made a case for multimodal approaches, our data suggest that besides gaze + voice, other multimodal approaches are less desired and will be used less frequently.

5.1 Modality Usage

In line with previous studies [4, 19], the voice-only input was confirmed to be the most preferred method for in-car activities. However, as already noted by Qian et al. [46], drawbacks of voice are recognition failures, confusion between specified instructions, and the influence of ambient noise. As noted by our work, these issues can be compensated by the fusion with other input modalities. We agree that providing the users with a multimodal interface will support a more comfortable and less stressful driving experience, increasing the system's reliability, robustness, and performance [1, 25]. This seems especially relevant when AVs enable passengers to reference external objects (see Figure 10).

The menu was designed hierarchical despite attempts from vehicle manufacturers (e.g., see Mercedes Benz's adaptive Hyperscreen [8]) and academia [22] to provide contextual data as hierarchical menus are still state-of-the-art. Despite this downside, touch was still mostly used. We assume that this is due to the habituation effect. As noted by Detjen et al. [19], touch seems more reliable and easy to understand, but it implies a specific hardware location. Gestures imply a higher physical load and effort [19] that could negatively influence the interaction quality compared to the voice or touch inputs. Our data support lower usage of this modality. Another possible reason could be that participants were allowed to point only with the index of their right hand, which could limit their preferences. However, we believe this to be, at most, a minor reason as the index finger, while not used universally [61], seems most useful for pointing. In contrast with other studies [2, 49, 51, 56] which describe pointing as a useful, intuitiveness, naturalness, and simple interaction modality, applicable without training, our data and open feedback suggest that pointing is less relevant.

Gomaa et al. [25] demonstrated that pointing does not suffer from significant effects of distance and environment density. However, in our cognitive experiment, we could not exclude an influence on the system's performance and interaction choice. To address this, we analyzed the interaction preferences in a static situation, i.e., stopping at an intersection. Our data suggest that users need at least 7 seconds to complete such activities.

5.2 Effects on NASA TLX, Trust, and Usability Evaluations in the "Driving" Scenario

As planned, the induced workload led to increased mental demand, showing the appropriateness of our method. Interestingly, this did not influence other demands in a significant fashion. Additionally, we found no significant differences for trust. We also found no significant differences for the SUS and only a barely significant difference (p=0.04) for information quality. We did find differences regarding the usage of touch for referencing outside objects: with cognitive load, this was used more than twice as often compared to no cognitive load. Our data suggest that the interaction modalities used are not significantly affected by the presence of cognitive load but that users do default to known interaction modalities such as touch.

5.3 Practical Implications

Our scenarios resembled potential real-world applications of using (multi-)modal referencing interactions. We opted for high external validity by allowing users to employ several modalities and their combinations. Our data suggest that a focus on adaptive touch interfaces and improved voice recognition should be the main focus of vehicle manufacturers. While referencing via gaze was also well-received, (visual) feedback seems crucial for understanding the currently selected object, despite advances in multimodal referencing accuracy [1]. This could be done via augmented reality windshield displays (which were suggested in numerous previous works [11, 12, 15, 28, 37, 52, 60, 62]). However, challenges such as parallax effects remain.

5.4 Limitations

One limitation is the moderate number of participants in the study (N=11). As mostly younger participants who were probably more proficient in novel technologies took part, the transferability of the study findings to other age groups is also unclear. Future work should also consider the demographics and characteristics (Quantity of experience, Perceptual sensitivity, Situation selectivity, Reflexivity, and Willingness) proposed by Rainer and Wohlin [47] for the selection of additional participants. For example, older participants or participants with little or much experience with AR should be recruited. Additionally, the study setting limited the possibility of simulating vection influence on interactions [14]. While we used VR to add immersion, the vection was already shown to have an influence on the perceived usefulness of interaction modalities [14, 29].

Regarding technical aspects, we implemented the interactions with state-of-the-art hardware and software. Nonetheless, there were technical limitations in the detection of interactions. However, this is also likely to happen in a moving vehicle. Therefore, these aspects, while reducing internal validity, improved external validity. Additionally, wearing masks will most likely have negatively influenced recognition rates for voice and potentially eyes.

Finally, the chosen interaction tasks in both scenarios were chosen carefully and resemble relevant and realistic future tasks. However, this is not an exhaustive list, and the interactions are not necessarily comparable (e.g., because of the size of the objects).

6 CONCLUSION

This work evaluated a multimodal interactive system to interact with objects in and outside an AV. In a within-subjects study with N=11 participants, in two scenarios, we investigated interaction preferences considering differences depending on the user's mental workload and the effect of perceived urgency and interaction task. Despite the availability of other modalities and combinations, we found that touch and voice were mostly used and preferred. Our work helps determine the relevance of multimodal interaction in a more ecologically valid setting as the use of modalities was not enforced. Therefore, it helps developers and designers in future interaction design.

ACKNOWLEDGMENTS

We thank all study participants and the Human Factors department at Ulm University. This work was supported by the project 'SEMULIN' (**se**lbstunterstützende, **mu**ltimodale **In**teraktion) funded by the Federal Ministry for Economic Affairs and Energy (BMWi).

A INTERACTION TASKS

Task Description	Translation
Entscheide, ob du nach Rechts oder nach Links fahren möchtest.	Decide if you want to turn right or left.
Informiere dich über die Höhe des größten Gebäudes vor dir.	Ask about the height of the tallest building in front of you.

Table 4. Overview of the interaction activities presented during the urgency condition.

Task Description	Translation
Bitte mache das Fenster der linken, vorderen Türe auf. Verwende dafür die Sprachsteuerung (sage z.B.: "Mach das Fenster auf") und zeige oder schaue in Richtung des Fensters.	Open the window of the left front door, using the voice command (e.g. say: "Open the window") and pointing or looking toward the window.
Bitte mach das Fenster auf. Benutze dafür den im Display angezeigten Fensterschalter.	Open the window. Use the window button shown in the display.
Bitte mach das Fenster zu. Benutze dafür nur die Sprachsteuerung (sage z.B.: "Mach das Fenster der linken vorderen Türe zu").	Close the window, using the voice commands (e.g. say: "Close the window of the left front door").
Bitte ändere die Temperatur auf 19°. Benutze dafür die Sprachsteuerung oder das Touchdisplay (sage z.B.: "Setze die Temperatur auf 19°")	Set the temperature to 19°, using the voice commands or the touch display (e.g. say: "Set the temperature to 19°").
Benutze das Touchdisplay, um Zugriff auf die Speisekarte des nächsten Restaurants zu bekommen.	Use the touch display to access the menu of the nearest restaurant.
Lass dir die Speisekarte des nächsten Restaurants anzeigen. Verwende dafür die Sprachsteuerung (sage z.B.: "Speisekarte des nächsten Restaurants").	Ask for the menu of the nearest restaurant (e.g. say: "Menu of the nearest restaurant").
Die Kirche, die im Display erscheint, wird bald zu sehen sein. Schau hin und frage über einen Sprachbefehl was das ist (frage z.B.:"Wie heißt die Kirche?" oder "Was ist das?").	Soon you will see the church displayed on the screen. Look at it and using your voice ask what that is (e.g. ask: "What's the name of the church?" or "What is that?"
Erkundige dich wozu der Knopf auf der Türe dient. Zeige oder schaue gezielt in Richtung des Knopfes und frage über einen Sprachbefehl was das ist (frage z.B.: "Was ist das?").	Find out what the functionalities of the button on the door. Point or look at the button, and ask what it is using the voice commands (e.g. ask: "What is that?").
Biege bitte nach rechts ab, sobald die Brücke überquert ist. Verwende dafür die Sprachsteurung (sage z.B.: "Gehe nach rechts").	Turn right once you have crossed the bridge, using the voice control (e.g. say: "Go to the right").
Biege bitte nach rechts ab, sobald die Brücke überquert ist. Betätige dafür den rechten Blinker.	Turn right once you have crossed the bridge. Use the right turn button to do this.
Informiere dich über die Öffnungszeiten des nächsten Parkhauses. Zeige mit dem Finger auf das Parkhaus und frage über einen Sprachbefehl nach den Öffnungszeiten.	Find out the opening hours of the nearest parking garage. Point your finger at the parking garage and ask for the opening time, using the voice commands.
Schalte die Innenlichtfarbe auf rot. Dafür kannst du unter "Einstellungen" die Lichterseite aufrufen. Um Sie zurückzusetzen, schalte die Farbe durch einen Sprachbefehl auf weiß.	Switch the ambient lighting color to red. To do this, navigate to the lights view of the car's Settings. To reset it, use the voice command to turn the color to white.
Informiere dich über die nächste Kirche. Zeige mit dem Finger oder schau auf das Gebäude und frag über einen Sprachbefehl wie sie heisst.	Discover information about the nearest church. Point or look at the building and ask its name using a voice command.
Informiere dich über die nächste Kirche. Dafür kannst du auf die "Attraktionen" Seite des Navigationssystem navigieren.	Discover the nearest church. For this you can navigate to the "Attractions" page of the navigation system.
Benutze die Sprachsteuerung, um dich über das Wetter zu informieren (frage z.B.: "Wie wird das Wetter?").	Ask about the weather using the voice commands (e.g. say: "What's the weather like?").
Benutze das Touchdisplay, um dich über das Wetter zu informieren.	Get some weather information using the touch display.
Lass dich nach München fahren. Benutze dafür das Touchdisplay.	Let you drive to Munich. Use the touch display for this.
Lass dich nach München fahren. Definiere das Ziel über Sprachbefehl (frage z.B.: "Fahre mich nach München.")	Let you drive to Munich. Use the voice command to set the destination (e.g. say: "Navigate to Munich").
Frag wie hoch ein zufälliges Wohngebäude ist. Zeige oder schaue gezielt in Richtung des Gebäudes und frage über einen Sprachbefehl nach der Höhe (frage z.B.: "Wie hoch ist das Gebäude?"	Choose a building randomly and ask how tall it is. Point or look at the building and use a voice command to ask for the height (e.g. ask: "How tall is the building?").

Table 5. Overview of the interaction activities presented during the training condition.

Task Description	Translation
Finde heraus wozu der Knopf an der Türe dient.	Find out for what the button on the door is used for.
Informiere dich über die Öffnungszeiten der nächsten Bank.	Find out the opening hours of the nearest bank.
Erkundige dich nach der Hohe des Ulmer Maritim Hotels.	Ask about the height of the Ulm Maritim Hotel.
Biege bitte nach links, vor der nächsten Kreuzung ab.	Turn left before the next intersection.
Lasse dir die Speisekarte des nächstes Restaurants anzeigen.	Ask for the menu of the next restaurant.
Bitte mache das Fenster auf.	Open the window.
Informiere dich über den Turm, der im Display anzeigt wird.	Inform about the tower displayed in the screen.
Informiere dich über das Wetter.	Ask about the weather.
Informiere dich über das nächste Gebäude.	Inform you about the next building.
Schalte die Farbes des Innenlichtes auf rot.	Switch the color of the ambient light to red.
Informiere dich über die Preise der nächsten Elektroladestation.	Ask about the price of the next electric charge station.
Lass dich nach München fahren.	Let you drive to Munich.

Table 6. Overview of the interaction activities presented during the cognitive condition.

Task Description	Translation
Der Brunnen, der im Display erscheint, wird bald zu sehen sein. Informiere dich über den Namen.	The fountain that appears in the display will appear soon. Find out about the name.
Bitte ändere die Temperatur auf 19°.	Set the temperature to 19°.
Informiere dich über das nächste Gebäude.	Inform yourself about the next building.
Biege bitte nach rechts, vor der nächste Kreuzung	Turn right before the next intersection.
Informiere dich über das Wetter.	Ask about the weather.
Informiere dich über das Ibis Hotel.	Ask about the Ibis Hotel.
Informiere dich über den Namen des nächsten Bahnhofes.	Ask about the name of the next train station.
Frage, wozu dient der Knopf an der Ture.	Find out for what the button on the door is used for.
Informiere dich über die Öffnungszeiten des nächsten Parkhauses.	Ask about the opening time of the next parking garage.
Schalte die Farbes des Innenlichtes auf blau.	Set the color of the ambient light to red.
Informiere dich über das nächste Gebäude.	Inform yourself about the next building.
Lass dich nach München fahren.	Get a ride to Munich.

Table 7. Overview of the interaction activities presented during the control condition.

REFERENCES

[1] Abdul Rafey Aftab and Michael von der Beeck. 2022. Multimodal Driver Referencing: A Comparison of Pointing to Objects Inside and Outside the Vehicle. In *27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 483–495. https://doi.org/10.1145/3490099.3511142

[2] Abdul Rafey Aftab, Michael von der Beeck, and Michael Feld. 2020. You have a point there: object selection inside an automobile using gaze, head pose and finger pointing. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. Association for Computing Machinery, New York, NY, USA, 595–603.

[3] Bashar I. Ahmad, Patrick M. Langdon, Simon J. Godsill, Robert Hardy, Lee Skrypchuk, and Richard Donkor. 2015. Touchscreen usability and input performance in vehicles under different road conditions: an evaluative study. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '15)*. Association for Computing Machinery, New York, NY, USA, 47–54. https://doi.org/10.1145/2799250.2799284

[4] Aya Ataya, Won Kim, Ahmed Elsharkawy, and SeungJun Kim. 2021. How to Interact with a Fully Autonomous Vehicle: Naturalistic Ways for Drivers to Intervene in the Vehicle System While Performing Non-Driving Related Tasks. *Sensors* 21, 6 (2021), 2206.

[5] Pradeep K Atrey, M Anwar Hossain, Abdulmotaleb El Saddik, and Mohan S Kankanhalli. 2010. Multimodal fusion for multimedia analysis: a survey. *Multimedia systems* 16, 6 (2010), 345–379.

[6] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2018. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence* 41, 2 (2018), 423–443.

[7] Klaus Bengler, Michael Rettenmaier, Nicole Fritz, and Alexander Feierle. 2020. From HMI to HMIs: Towards an HMI Framework for Automated Driving. *Information* 11, 2 (Feb. 2020), 61. https://doi.org/10.3390/info11020061 Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[8] Mercedes Benz. 2021. Creating the MBUX Hyperscreen. https://supplier-portal.daimler.com/docs/DOC-2653. [Online; accessed 07-APRIL-2022].

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

[10] 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* (Leeds, United Kingdom) *(AutomotiveUI '21)*. Association for Computing Machinery, New York, NY, USA, 79–89. https://doi.org/10.1145/3409118.3475134

[11] Mark Colley, Christian Bräuner, Mirjam Lanzer, Marcel Walch, Martin Baumann, and Enrico Rukzio. 2020. Effect of Visualization of Pedestrian Intention Recognition on Trust and Cognitive Load. 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, 181–191. https://doi.org/10.1145/3409120.3410648

[12] 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*. Association for Computing Machinery, New York, NY, USA, Article 155, 11 pages. https://doi.org/10.1145/3411764.3445351

[13] Mark Colley, Sebastian Hartwig, Albin Zeqiri, Timo Ropinski, and Enrico Rukzio. 2023. AutoTherm: A Dataset and Ablation Study for Thermal Comfort Prediction in Vehicles. *arXiv preprint arXiv:2211.08257* (2023).

[14] 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

[15] 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

[16] 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 (sep 2022), 19 pages. https://doi.org/10.1145/3546713

[17] Mark Colley, Dennis Wolf, Sabrina Böhm, Tobias Lahmann, Luca Porta, and Enrico Rukzio. 2021. Resync: Towards Transferring Somnolent Passengers to Consciousness. In *Adjunct Publication of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 1, 6 pages. https://doi.org/10.1145/3447527.3474847

[18] 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 (MuC'19)*. Association for Computing Machinery, New York, NY, USA, 341–348. https://doi.org/10.1145/3340764.3340798

[19] Henrik Detjen, Stefan Geisler, and Stefan Schneegass. 2020. Maneuver-based Control Interventions During Automated Driving: Comparing Touch, Voice, and Mid-Air Gestures as Input Modalities. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, New York, NY, USA, 3268–3274. https://doi.org/10.1109/SMC42975.2020.9283431

[20] Henrik Detjen, Bastian Pfleging, and Stefan Schneegass. 2020. A Wizard of Oz Field Study to Understand Non-Driving-Related Activities, Trust, and Acceptance of Automated Vehicles. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 19–29.

[21] Tanja Döring, Dagmar Kern, Paul Marshall, Max Pfeiffer, Johannes Schöning, Volker Gruhn, and Albrecht Schmidt. 2011. Gestural Interaction on the Steering Wheel: Reducing the Visual Demand. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 483–492. https://doi.org/10.1145/1978942.1979010

[22] Michael Feld, Gerrit Meixner, Angela Mahr, Marc Seissler, and Balaji Kalyanasundaram. 2013. Generating a Personalized UI for the Car: A User-Adaptive Rendering Architecture. In *User Modeling, Adaptation, and Personalization*, Sandra Carberry, Stephan Weibelzahl, Alessandro Micarelli, and Giovanni Semeraro (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 344–346.

[23] Kikuo Fujimura, Lijie Xu, Cuong Tran, Rishabh Bhandari, and Victor Ng-Thow-Hing. 2013. Driver queries using wheel-constrained finger pointing and 3-D head-up display visual feedback. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '13)*. Association for Computing Machinery, New York, NY, USA, 56–62. https://doi.org/10.1145/2516540.2516551

[24] Frostweep Games. 2022. Speech Recognition using Google Cloud [VR AR Mobile Desktop] Pro. https://assetstore.unity.com/packages/add-ons/machinelearning/speech-recognition-using-google-cloud-vr-ar-mobile-desktop-pro-72625. [Online; accessed 07-APRIL-2022].

[25] Amr Gomaa, Guillermo Reyes, Alexandra Alles, Lydia Rupp, and Michael Feld. 2020. Studying person-specific pointing and gaze behavior for multimodal referencing of outside objects from a moving vehicle. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. Association for Computing Machinery, New York, NY, USA, 501–509.

[26] Amr Gomaa, Guillermo Reyes, and Michael Feld. 2021. ML-PersRef: A Machine Learning-Based Personalized Multimodal Fusion Approach for Referencing Outside Objects From a Moving Vehicle. In *Proceedings of the 2021 International Conference on Multimodal Interaction* (Montréal, QC, Canada) *(ICMI '21)*. Association for Computing Machinery, New York, NY, USA, 318–327. https://doi.org/10.1145/3462244.3479910

[27] 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.

[28] Renate Häuslschmid, Max von Bülow, Bastian Pfleging, and Andreas Butz. 2017. Supporting Trust in Autonomous Driving. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) *(IUI '17)*. Association for Computing Machinery, New York, NY, USA, 319–329. https://doi.org/10.1145/3025171.3025198

[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 *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Seoul, Republic of Korea) *(AutomotiveUI '22)*. Association for Computing Machinery, New York, NY, USA, 204–214. https://doi.org/10.1145/3543174.3545252

[30] Pascal Jansen, Julian Britten, Alexander Häusele, Thilo Segschneider, Mark Colley, and Enrico Rukzio. 2023. AutoVis: Enabling Mixed-Immersive Analysis of Automotive User Interface Interaction Studies. In *Proceedings of the 2023*

CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 378, 23 pages. https://doi.org/10.1145/3544548.3580760

[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] M. Kauer, M. Schreiber, and R. Bruder. 2010. How to conduct a car? A design example for maneuver based driver-vehicle interaction. In *2010 IEEE Intelligent Vehicles Symposium*. IEEE, New York, NY, USA, 1214–1221. https://doi.org/10.1109/IVS.2010.5548099 ISSN: 1931-0587.

[33] Myeongseop Kim, Eunjin Seong, Younkyung Jwa, Jieun Lee, and Seungjun Kim. 2020. A Cascaded Multimodal Natural User Interface to Reduce Driver Distraction. *IEEE Access* 8 (2020), 112969–112984. https://doi.org/10.1109/ACCESS.2020.3002775 Conference Name: IEEE Access.

[34] Moritz Körber. 2019. Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, Sebastiano Bagnara, Riccardo Tartaglia, Sara Albolino, Thomas Alexander, and Yushi Fujita (Eds.). Springer International Publishing, Cham, 13–30.

[35] Shunsuke Koyama, Yuta Sugiura, Masa Ogata, Anusha Withana, Yuji Uema, Makoto Honda, Sayaka Yoshizu, Chihiro Sannomiya, Kazunari Nawa, and Masahiko Inami. 2014. Multi-Touch Steering Wheel for in-Car Tertiary Applications Using Infrared Sensors. In *Proceedings of the 5th Augmented Human International Conference* (Kobe, Japan) *(AH '14)*. Association for Computing Machinery, New York, NY, USA, Article 5, 4 pages. https://doi.org/10.1145/2582051.2582056

[36] James Lewis. 1992. Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ, In Proceedings of the human factors society annual meeting. *Proceedings of the Human Factors Society* 2, 1259–1263.

[37] Patrick Lindemann, Tae-Young Lee, and Gerhard Rigoll. 2018. Catch my drift: Elevating situation awareness for highly automated driving with an explanatory windshield display user interface. *Multimodal Technologies and Interaction* 2, 4 (2018), 71.

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

[39] Udara E Manawadu, Mitsuhiro Kamezaki, Masaaki Ishikawa, Takahiro Kawano, and Shigeki Sugano. 2016. A hand gesture based driver-vehicle interface to control lateral and longitudinal motions of an autonomous vehicle. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, New York, NY, USA, 001785–001790.

[40] Udara E. Manawadu, Mitsuhiro Kamezaki, Masaaki Ishikawa, Takahiro Kawano, and Shigeki Sugano. 2016. A hand gesture based driver-vehicle interface to control lateral and longitudinal motions of an autonomous vehicle. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, New York, NY, USA, 001785–001790. https://doi.org/10.1109/SMC.2016.7844497

[41] Robert Neßelrath, Mohammad Mehdi Moniri, and Michael Feld. 2016. Combining Speech, Gaze, and Micro-gestures for the Multimodal Control of In-Car Functions. In *2016 12th International Conference on Intelligent Environments (IE)*. IEEE, New York, NY, USA, 190–193. https://doi.org/10.1109/IE.2016.42 ISSN: 2472-7571.

[42] Alexander Ng and Stephen A. Brewster. 2016. Investigating Pressure Input and Haptic Feedback for In-Car Touchscreens and Touch Surfaces. 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, 121–128. https://doi.org/10.1145/3003715.3005420

[43] Max Pfeiffer, Dagmar Kern, Johannes Schöning, Tanja Döring, Antonio Krüger, and Albrecht Schmidt. 2010. A Multi-Touch Enabled Steering Wheel: Exploring the Design Space. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 3355–3360. https://doi.org/10.1145/1753846.1753984

[44] 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

[45] Tony Poitschke, Florian Laquai, Stilyan Stamboliev, and Gerhard Rigoll. 2011. Gaze-based interaction on multiple displays in an automotive environment. In *2011 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, New York, NY, USA, 543–548. https://doi.org/10.1109/ICSMC.2011.6083740 ISSN: 1062-922X.

[46] 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

[47] Austen Rainer and Claes Wohlin. 2022. Recruiting credible participants for field studies in software engineering research. *Information and Software Technology* 151 (2022), 107002. https://doi.org/10.1016/j.infsof.2022.107002

[48] Florian Roider and Tom Gross. 2018. I See Your Point: Integrating Gaze to Enhance Pointing Gesture Accuracy While Driving. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive*

Vehicular Applications (AutomotiveUI '18). Association for Computing Machinery, New York, NY, USA, 351–358. https://doi.org/10.1145/3239060.3239084

[49] Florian Roider and Tom Gross. 2018. I see your point: Integrating gaze to enhance pointing gesture accuracy while driving. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 351–358.

[50] Florian Roider, Sonja Rümelin, Bastian Pfleging, and Tom Gross. 2017. The effects of situational demands on gaze, speech and gesture input in the vehicle. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 94–102.

[51] 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*. Association for Computing Machinery, New York, NY, USA, 40–47.

[52] Tobias Schneider, Joana Hois, Alischa Rosenstein, Sabiha Ghellal, Dimitra Theofanou-Fülbier, and Ansgar R.S. Gerlicher. 2021. ExplAIn Yourself! Transparency for Positive UX in Autonomous Driving. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 161, 12 pages. https://doi.org/10.1145/3411764.3446647

[53] Tevfik Metin Sezgin, Ian Davies, and Peter Robinson. 2009. Multimodal inference for driver-vehicle interaction. In *Proceedings of the 2009 international conference on Multimodal interfaces (ICMI-MLMI '09)*. Association for Computing Machinery, New York, NY, USA, 193–198. https://doi.org/10.1145/1647314.1647348

[54] Annika Stampf, Mark Colley, and Enrico Rukzio. 2022. Towards Implicit Interaction in Highly Automated Vehicles - A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 191 (sep 2022), 21 pages. https://doi.org/10.1145/3546726

[55] CityGen Technologies. 2022. CityGen3D. https://assetstore.unity.com/packages/tools/terrain/citygen3d-162468. [Online; accessed 07-APRIL-2022].

[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] Ultraleap. 2022. Leap Motion Controller. https://www.ultraleap.com/product/leap-motion-controller/. [Online; accessed 07-APRIL-2022].

[58] Marcel Walch, Lorenz Jaksche, Philipp Hock, Martin Baumann, and Michael Weber. 2017. Touch Screen Maneuver Approval Mechanisms for Highly Automated Vehicles: A First Evaluation. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct (AutomotiveUI '17)*. Association for Computing Machinery, New York, NY, USA, 206–211. https://doi.org/10.1145/3131726.3131756

[59] 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 (Automotive'UI 16)*. Association for Computing Machinery, New York, NY, USA, 261–268. https://doi.org/10.1145/3003715.3005458

[60] Gesa Wiegand, Christian Mai, Kai Holländer, and Heinrich Hussmann. 2019. InCarAR: A Design Space Towards 3D Augmented Reality Applications in Vehicles. 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, 1–13. https://doi.org/10.1145/3342197.3344539

[61] David Wilkins. 2003. Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). *Pointing: Where language, culture, and cognition meet* (2003), 171–215.

[62] Philipp Wintersberger, Anna-Katharina Frison, Andreas Riener, and Tamara von Sawitzky. 2019. Fostering User Acceptance and Trust in Fully Automated Vehicles: Evaluating the Potential of Augmented Reality. *PRESENCE: Virtual and Augmented Reality* 27, 1 (2019), 46–62. https://doi.org/10.1162/pres_a_00320 arXiv:https://doi.org/10.1162/pres_a_00320

Received January 2023; revised May 2023; accepted June 2023