Teaching Drivers about ADAS using Spoken Dialogue: A Wizard of Oz Study

Luka Rukonić[1][a], Marie-Anne Pungu Mwange[2][b] and Suzanne Kieffer[1][c]

[1]*Institute for Language and Communication, Université catholique de Louvain, Louvain-la-Neuve, Belgium*

[2]*AISIN Europe, Braine-L'Alleud, Belgium*

Keywords: Interactive Driver Tutoring, Mental Models, Car Voice Assistant, Wizard of Oz, Advanced Driver Assistance.

Abstract: Understanding the limitations and capabilities of the advanced driver-assistance systems (ADAS) is a prerequisite for their safe and comfortable use. This paper presents a formative user study on the use of a dialogue-based system, implemented using the Wizard of Oz (WOz) technique, to help drivers learn about the correct use of driving assistance. We investigated whether drivers would build the correct mental model of the driving assistance systems through natural language dialogue. We describe the evolution of the prototype over four iterations of formative evaluation with older and younger drivers. Using a mixed-method approach, combining the WOz, interviews, questionnaires, and a knowledge quiz, we evaluated the prototype of a voice assistant and identified the teaching content objectives. Participants' mental model about ADAS was assessed to evaluate the efficacy of the teaching approach. The results show that the teaching goals need to be clearly communicated to the drivers to ensure the adoption of the VA.

1 INTRODUCTION

Advanced Driver-Assistance Systems (ADAS) are present in almost every new vehicle sold on the market. According to EU regulations, all new vehicles will need to be equipped with at least level 2 automation according to SAE taxonomy (SAE, 2018). ADAS leads to increased driving safety and comfort, however, only if used correctly and consciously. Some car models offer much more than basic lane-keeping and adaptive cruise control systems, such as lane departure warning, automatic emergency braking, obstacle and object detection and driver state monitoring. Although useful, many drivers are not aware of all the systems their car is equipped with, which leads to a low usage and adoption rate of these systems. In addition, many drivers do not receive proper training about the availability and capability of the ADAS systems when purchasing a new car. Besides, learning about automation systems from user manuals is also insufficient to build a correct mental model of the systems (Boelhouwer et al., 2019). Because of that, drivers learn how to use ADAS systems by themselves, on public roads, where they test how the various systems work or when they stop functioning. As a result, driving safety benefits might be waived due to wrongly constructed drivers' mental models about how ADAS functions (Rossi et al., 2020). Further, better understanding of the limitations and boundaries of driving automation systems leads to improvements in drivers' trust calibration which is a crucial aspect to their safe use (Walker et al., 2018)

This study primarily focused on drivers above 50, who are underrepresented in research studies (Young et al., 2017), but are highly interested in the benefits of using ADAS, despite concerns about security issues, system failures or hacking attacks(Schmargendorf et al., 2018). Specifically, we investigated the use of an interactive driver tutoring system with a voice assistant (VA) to teach drivers about the correct use of ADAS. Using a Wizard of Oz (WOz) simulating the car-driver dialogue, we explored how drivers would interact with a VA using natural language. We focused on teaching drivers about the capabilities and limitations of the Lane Keeping Assistant (LKA) and Adaptive Cruise Control (ACC) while driving on the highway. We collected data using a high-fidelity simulator platform. We combined subjective, self-reported data with interviews to evaluate the user experience (UX) with the prototype.

[a] https://orcid.org/0000-0003-1058-0689
[b] https://orcid.org/0000-0003-2971-6347
[c] https://orcid.org/0000-0002-5519-8814

Rukonić, L., Mwange, M. and Kieffer, S.
Teaching Drivers about ADAS using Spoken Dialogue: A Wizard of Oz Study.
DOI: 10.5220/0010913900003124
In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 2: HUCAPP*, pages 88-98
ISBN: 978-989-758-555-5; ISSN: 2184-4321

The specific objectives of this study were: (1) explore how a voice assistant could be used to facilitate learning about ADAS while driving; (2) evaluate the driver's trust in ADAS; (3) evaluate the mental model of drivers about ADAS.

2 RELATED WORK

Research on the use of voice interfaces in cars is plentiful. Politis et al. Politis et al. (2018) evaluated the different types of dialogue-based systems to mediate the takeover in automated driving by increasing the situational awareness of the driver. They found that an interface without additional spoken information was most accepted, followed by the interface where drivers had to respond to questions asked by the system about the driving situation such as hazards, current lane, speed or fuel level. Schmidt et al. (2020) evaluated a proactive and reactive voice assistant in use-cases such as navigating, refueling, and checking the news. Proactivity was generally well accepted, but not for navigation use cases. Voice alerts before automated braking have a positive impact on driver's anxiety, alertness, and sense of control (Koo et al., 2016). Other applications include using voice to guide a takeover from automated to manual driving (Kasuga et al., 2020), or as an auditory reliability display to support and relax drivers in high levels of automation (Frison et al., 2019).

2.1 Driver Tutoring

Merriman et al. (2021) identified driver training as a solution to address four challenges associated with the use automated vehicles, which are: (1) Drivers have a poor mental model of the automation's function, capabilities, and limitations; (2) Automation reduces driver's cognitive workload thus decreasing driver's attention to the road; (3) Over-reliance on automation and reduction of driver's self-confidence to drive well manually without assistance; (4) Degradation of driver's procedural skills to drive manually. Our study focuses on the first challenge by addressing the alignment of the driver's mental model with the capabilities and limitations of the driving-assistance systems. Our prototype serves as the first step towards voice-guided teaching while driving.

Forster et al. (2019; 2020) evaluated the effect of user education using owner's manuals and interactive tutorials on driver's mental model creation, satisfaction, understanding, and performance with automated systems. Both approaches showed superior operator behavior compared to a generic baseline, however, no difference in satisfaction was found. Boelhouwer et al. (2019) found that informing drivers about car's automation capabilities using owner's manual was not sufficient to build accurate mental models of driving automation. In addition, existing systems for in-car driver tutoring rely on AR overlay on the windscreen and auditory explanations of the ADAS functionalities in low-complexity driving situations (Boelhouwer et al., 2020).

Furthermore, Rossi et al. (2020) showed that people who built their mental model of ADAS based on reading a description about it, use ADAS more effectively and safely compared to those who build the mental model on their own. Past research calls for a structured approach to user education about driving automation that relies on guided learning and building declarative knowledge (Forster et al., 2019). Building a correct or accurate mental model of driving automation and driving assistance technologies is crucial for using it appropriately. Mental models represent what people know and believe about the system they use (Payne, 2012). This approach focuses on the contents of the user's mind, rather than the explanation of cognitive processes for building the representations of systems. The knowledge people possess can be used to explain the behavior with the system in question.

2.2 Trust

Overtrust in ADAS can lead to decreased driving safety and comfort. This is caused by a lack of knowledge about the system's boundaries and capabilities. Walker et al. (2018) found that drivers tend to overtrust Level 2 vehicle's automation capabilities before trying it in real-life situations. After the on-road experience, drivers understood better what the capabilities of the ADAS were, and this led to a decrease in trust in 7 out of 12 scenarios. According to Hoff and Bashir (2015), trust is a multidimensional concept consisting of dispositional, situated, and learned trust. Preconceptions about the system play an important role in forming trust. A study by Lasarati et al. (2020) investigated the effect of four textual explanation styles on users' trust in AI medical diagnostic systems. They found that thorough and contrastive styles were rated the highest. Proactive dialogue strategies, promoting acting to the user before problematic situations happen are found to induce higher levels of trust when interacting with intelligent assistants such as robots (Kraus et al., 2020). We used this assumption when designing our proactive voice assistant for driver tutoring.

(a) Participants drove on L3 and used L2 to avoid obstacles. (b) Route indicated in blue on the loop highway in CARLA.

Figure 1: Lane configuration and road map used in the study.

3 USER STUDY

3.1 Iterative Design

We adopted a formative test-and-refine approach to the design of the teaching assistant similar to Rukonic et al. (2021). We conducted four iterations with a small sample of participants to evaluate the prototypes. Each new iteration saw improvements to the prototype and the teaching content. This study does not elaborate in detail the results and findings in each iteration, but rather provides an overview of what we learned and how we applied that to the incremental development of the prototype. Thus, after a brief description of each iteration we focus on the main qualitative findings and lessons learned. Between iterations, the user profile of participants varied, as well as the data we collected to accommodate for the investigation of aspects we found relevant based on the findings from the previous iteration. We aimed at collecting qualitative data rather than investigating systematic differences between conditions through experimental manipulations. WOz is a great support in the iterative design process that leads the implementation from early phases toward the design of the final version (Dow et al., 2005).

Iteration 1. required participants to drive two different driving scenarios, namely, a baseline scenario (scenario B) and a scenario with on voice assistant (scenario VA). All participants drove both scenarios. First, we ran scenario B to establish a baseline driver behavior and to investigate their use of available ADAS systems. Then, we ran scenario VA to check whether the VA actually supported drivers in their learning about ADAS systems. We designed a use case in which people drove on a highway and were strongly hinted to use driving assistance.

In scenario B no advice was given to the participants, so that they would build a mental model of how to operate the LKA and ACC on their own, i.e., without voice assistance. They drove five laps on the highway (Figure 1b), which lasted around 10 minutes. Drivers encountered two obstacles, both necessitating a takeover. The car detected obstacle#1 (truck), which led to an activation of the emergency braking if they did not react. In contrast, the car did not detect obstacle#2 (container on the road), which led to a collision if they did not perform an emergency maneuver (Figure 1a). In scenario VA, we used two language styles (expert vs. helper) to write the advice for each type of road event. These two styles were modeled after the explanation goals presented by Sørmo and Cassons (2005). They presented four explanation goals for case-based reasoning, of which we selected two and applied them to the design of our advice content. The two explanation styles are based on the Transparency goal (expert style), and Learning goal (helping style). The expert style aims at explaining how the system reached the answer, while the helping style aims at teaching the user about the ADAS domain. In the VA condition, roadworks and a police car were used as obstacles. The police car triggering a similar reaction as the truck of scenario B and a sign announcing the roadworks taking the place of the container of the previous use case.

Iteration 2. improved reactivity and proactivity of the teaching assistant for ADAS. Also, the steering wheel was turning while LKA was on. The goal was, therefore, to explore users' attitude towards a VA that guides them through each ADAS feature. Two scenarios were designed, where Scenario 1 aimed at driver's learning about the ADAS systems, while Scenario 2 aimed at putting their knowledge at test by making the right decisions when avoiding obstacles and usage of safety systems such as forward collision warning (FCW) and automatic emergency braking (AEB). We removed the baseline scenario and made the VA available in both scenarios. Before Scenario 1, we demonstrated the VA by guiding participants through a short conversation with it to build the understanding of how it worked. Instead of having to use the VA immediately in the Scenario 1 as in iteration 1, we learned that participants needed to become familiar with the VA first as well. This way their efforts were less focused on dealing with learning how to use the VA, and more on learning about ADAS. Moreover, we decided to write the teaching content using only one language style, since we could not evaluate its effects on our small sample size.

Iteration 3. focused on the transition between autonomous and manual driving modes, precisely on preparing drivers to take over and to hand over the control. These events took place when getting on or leaving the highway, e.g. to change the direction or take the cloverleaf interchange roads (Fig. 1b). ADAS was available only on the highway loop and in case there were no obstacles ahead. ADAS was not available on side roads, interchange loops, and when an obstacle was announced. The VA content was written with these situations in mind. Scenario 1 covered the transition use-case and the availability of autonomous driving depending on the road type. Scenario 2 focused on the interaction with and use of ADAS while the driver was performing a secondary task, also called a non driving-related task (NDRT), i.e. making a phone call. We told participants to use the VA to learn about capabilities and limitations of ADAS, traffic situation, and takeover procedure. We used NDRTs to put the driver out of the loop (OOTL), i.e. in situations where drivers are not monitoring the driving situation, but may or may not be physically controlling the car (Merat et al., 2019). Being OOTL causes a loss of situational awareness and compromises drivers' monitoring or takeover capabilities (Merat et al., 2019). We aimed to explore the potential role of the VA to help maintain driving safety and situational awareness while the driver is making a phone call. We hypothesized that a low situational awareness caused by the NDRT would naturally motivate drivers to use the VA, so as to rebuild it. As research in UX tells us that when people have a goal in mind, they use the appropriate technology to achieve it (Hassenzahl, 2018).

Iteration 4. required participants to execute a series of tasks with the VA. Scenario 1 focused on learning about ADAS. The tasks were presented to them on a tablet placed on their left-hand side next to the steering wheel. We allocated a few minutes before the scenario started to read the tasks thoroughly and get familiar with them. They could ask us for clarifications if needed. All tasks were presented on the screen at the same time. We told participants that they were free to choose when to complete each task and their order. However, we suggested following the order of tasks as shown on the screen. Scenario 2 was designed to collect interaction data with the VA in obstacle avoidance situations.

3.2 Teaching Content Design

The process of learning is divided into three stages: acquiring declarative and procedural knowledge, consolidating the acquired knowledge, and tuning the knowledge towards overlearning (Kim et al., 2013). We designed the teaching content of our study with the aim of building drivers' declarative knowledge. It focuses on facts and things people know about a system, which in our case were ADAS systems such as ACC and LKA. Conversely, procedural knowledge explains how to solve a problem or achieve a task using a system (Forster et al., 2019; Kim et al., 2013). However, declarative knowledge is prone to fast decay in case of infrequent use. Knowing this limitation, we believed that acquiring declarative knowledge while practicing the use of ADAS in parallel might help build the procedural knowledge, which is immune to decay. To that end, we measured driver's learning and retention of declarative knowledge about ADAS in this study.

3.3 Participants

Each iteration involved six participants holding a valid driver's license. All participants were given an incentive for taking part in the study. Iterations 1-3 only involved people above 50 years old, who were active users of ADAS including ACC and LKA, as 50+ drivers were the focus of the research project we were working on. Iteration 4 involved both young and older drivers with no existing experience with ADAS, as VA might be more useful to drivers who begin using it. Drivers with previous experience with ADAS might have built a mental model that is hard to change through the acquisition of declarative knowledge about ADAS and its related features. We considered someone an older driver was a person above 50 and holding their driver's license for more than 20 years. Conversely, a younger driver was a person between 22 and 30 years old, holding their driver's license for a maximum of three years.

(a) Fixed-base simulator setup. (b) HUD automation status icon and speed indicator.

Figure 2: Simulator setup and head-up-display (HUD) icons.

3.4 Materials and Apparatus

The study was conducted using a fixed-base driving simulator. It consisted of three 50-inch screens covering an 180 degrees field of view, a real adjustable car seat, a Fanatec steering wheel with force feedback, gas, brake, and clutch pedals, and a gear lever (Fig 2a). A high-fidelity sound system was installed to produce a realistic and immersive sound, together with a low-frequency speaker placed under the seat to simulate car vibrations. We used the open-source CARLA simulator software (Dosovitskiy et al., 2017) that we adapted to our needs by enabling manual and automated driving. A screen showing the navigation map was placed on the right-hand side of the driver. The driving scenarios were executed in a closed-loop highway environment depicted in Figure 1b. To simulate ADAS systems, LKA and ACC were implemented in a slightly different way compared to real cars. Regarding LKA, it only worked on one of the four highway lanes. Regarding ACC, the minimum speed limit to activate it was 50 km/h. To turn on the driving assistance, a driver had to press a button on the steering wheel. To adjust the speed of ACC, there were two buttons on the steering wheel to increase or decrease the speed by 5 km/h. In sum, the ADAS systems did not replicate those intended for road use, but only resembled them.

3.5 Wizard of Oz

We used the Wizard of Oz (WOz) method involving two wizards to simulate the VA. Two human wizards worked together on analyzing driver's utterances and generating output. One wizard was in charge of generating feedback utterances after the obstacle avoidance, while the other was in charge of generating utterances related to teaching about the driving assistance. The wizards were located in the same room as the simulator, but separated from it by sound-proof panels. We installed a microphone close to the steering wheel to hear what the participants were saying. Wizards wore headsets during the evaluation and each worked on their own workstation to generate the appropriate output. When in doubt, wizards consulted each other to clear out misunderstandings or to agree on the output that was to be selected.

The wizards listened to participants' utterances and then selected the appropriate response in the wizard's user interface. A text-to-speech (TTS) system produced the spoken outputs using a male voice with a British accent. We followed a proactive approach when designing our VA similarly to Schmidt et al. (2020), where the VA would provide warnings or feedback without participant's request. The goal was to explore how drivers would interact with the VA to learn about the limitations and capabilities of the LKA and ACC. Participants were given the following three tasks explaining what the the voice assistant could do: (1) Explain driving assistance features such as LKA and ACC; (2) Advise how to use the driving assistance systems; (3) Explain the decisions the car is making. In addition, we wrote a set of predefined messages for the two wizards related to the feedback and questions we assumed would be asked the most (see Data Availability section for a full dataset of WOz messages). The types of predefined messages to respond to driver's requests included responses about: ACC/LKA activation/deactivation, ACC/LKA descriptions and purpose, Lane change/steering capabilities, Non-supported questions (e.g. navigation, reminders, appointments, weather).

Upon arrival, participants signed a consent form and were given explanations about study goals and data collection. Then, we administered the demographic survey (age, gender, driving experience, and experience with ADAS systems) and interviewed participants about their experience of learning how to use ADAS. Afterwards, participants completed a 10-minute test drive in manual driving mode to get familiar with the simulator. Then, we used a slideshow presenting the rules to follow while driving and instructions about the simulator and the VA. The slideshow included a prerecorded narration to avoid between-participants variability in the explanation of the rules and instructions about the test procedure. Thus, all participants received the same information before starting the experiments. This consideration ensures the internal validity when studying driving assistance systems (Schnöll, 2021). Engagement and disengagement criteria of the driving assistance, together with braking and steering interventions were explained. However, no detailed instructions about the take-over procedure were given. Then, the Scenario 1 started, followed by Scenario 2. After each scenario, we administered a knowledge quiz followed by the questionnaires assessing trust and quality of the VA. All sessions were video-recorded. Finally, we conducted a semi-structured interview and closed the session.

3.6 UX Measures

We combined relevant questionnaires identified in the literature and a custom-made quiz to collect data about the UX with the VA. We used the SASSI questionnaire (Hone and Graham, 2000) to evaluate the VA, as it provides the highest coverage of UX dimensions and it is a recommended instrument to evaluate speech-based interfaces (Kocaballi et al., 2018) and is best-rated in terms of face validity (Brüggemeier et al., 2020). SASSI consists of 34 items rated on a 7-point rating scale and distributed between six factors: system response accuracy, likeability, cognitive demand, annoyance, habitability and speed.

We evaluated trust using a questionnaire by Madsen & Gregor (2000) initially designed to evaluate trust between humans and automated systems supporting decision-making and providing advice to users. The instrument consists of five dimensions: perceived reliability, perceived technical competence, perceived understandability, faith and personal attachment. Items were rated using a 5-point Likert-type scale.

We created and used a knowledge quiz to assess the accuracy of participants' mental model of the driving assistance system with respect to the amount of declarative knowledge they retained or gained during the simulation. We administered the knowledge quiz after each condition in iterations 1, 2 and 4, the quiz containing 10, 11 and 14 statements respectively (Table 1). The possible answers were "True", "False" and "I don't know". Both wrong answers and "I don't know" were considered incorrect answers. In iteration 4, each statement used a Likert-type response format, ranging from Strongly Disagree (1) to Strongly Agree (6), without the neutral position to encourage precise answers from participants. We treated items individually as interval data, and not as a Likert scale. Thus, we do not compare the quiz results of iterations 1-2 to iteration 4. In iteration 4, we administered the knowledge quiz twice: immediately after the simulation and two weeks later, remotely, without participants coming to drive in the simulator again.

4 RESULTS

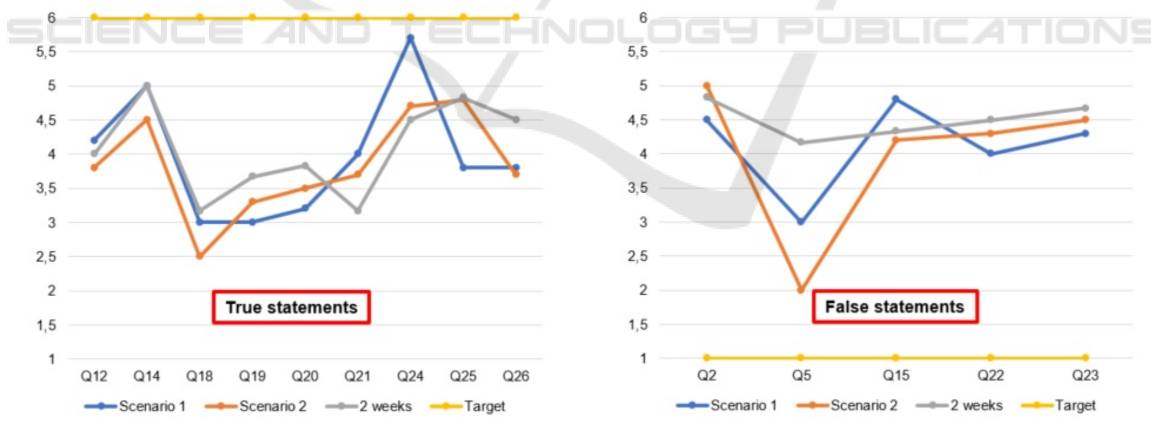
4.1 Driver's Learning - Mental Model

Knowledge quizzes in iterations 1 and 2 had 10 and 11 questions respectively, thus we calculated the mean scores of correct answers and converted them to percentages. Overall, participants scored better on the quiz after Scenario 2 (Table 2). The quiz scores were higher in iteration 2, indicating a more accurate mental model. Due to the nature of formative evaluations with a small participants sample, we cannot directly attribute the higher score to the presence of the VA. Q3, Q6, and Q8 had the lowest number of correct answers in iteration 1. Regarding Q3 and Q6, participants could not answer correctly because they did not see those events happening, nor did they ask the VA about it. Although AEB was implemented to stop before a car and a truck, no participant let that happen. Regarding Q8, none of the participants asked the VA regarding the minimum speed to activate ADAS, nor did we implement a specific indication to warn the drivers about that. All other questions covered situations that happened in test scenarios and participants could draw conclusions based on that.

To evaluate the mental model of participants in iteration 4, we divided the statements from the quiz in two categories: true statements and false statements. The differences in the answers to the quiz between two scenarios were very small (Figure 3). The paired t-test also showed no statistically significant differences. The mean quiz scores in Scenario 1 ($M=4.02$) and scenario 2 ($M=3.89$) indicate that there was no difference in the understanding of the ADAS between two scenarios ($t(13) = 0.822$, $p=.42$, two-tailed). However, for better interpretation of quiz results, we calculated the closeness score of participants' mental model representation of ADAS (perceived representation) to the actual functioning of ADAS (actual representation). The closer the average score for each question in the quiz was to the target score of 1 or 6, the more accurate we considered their mental model representation to be.

Table 1: Knowledge quiz statements used in Iteration 1, 2 and 4 and related correct answers.

ID	IT	Question	Answer
Q1	1	When ADAS is activated, the steering wheel moves when the car is turning	False
Q2	1,2,4	The system can detect (static) objects on the road (barriers, trash, etc.)	False
Q3	1	The system can detect other cars	True
Q4	1	When the assistant is activated, the system can handle turns	True
Q5	1,2,4	I can activate LKA and ACC separately	False
Q6	1	The system can detect trucks	True
Q7	1	When the assistant is activated, the car is not able to avoid road works	True
Q8	1	The assistant will work only if I drive faster than 50 km/h	True
Q9	1	When the assistant is activated, the icon turns white	False
Q10	1	The ACC helps me to stay in the lane	False
Q11	1,2	To deactivate the driving assistance, I can brake	True
Q12	2,4	The Automated Emergency Braking activates with stationary and moving vehicles	True
Q13	2	The car can detect road works	False
Q14	2,4	The car is able to avoid accidents by braking if the car in front of it is too close	True
Q15	2,4	The ACC is not able to adapt the car's speed with stationary vehicles	False
Q16	2	The driving assistant can be turned on at 30kph	False
Q17	2	The Forward Collision Warning is able to only detect moving vehicles	False
Q18	4	There are potential problems to stay in lane in sharp turns	True
Q19	4	Automated driving might not work well in the rain	True
Q20	4	Automated driving might not work well in the fog	True
Q21	4	There are potential problems to recognize pedestrians on the road	True
Q22	4	The warning alerts are given for any type of obstacle on the road	False
Q23	4	The car will come to a safe stop if I do not react to a takeover request	False
Q24	4	I need to take over every time the take over request occurs	True
Q25	4	Automated driving system requires me to constantly monitor the road	True
Q26	4	Automated driving system is only available on the highway with multiple lanes	True



(a) Scores for true quiz statements.

(b) Scores for false quiz statements.

Figure 3: Mental model assessment mean scores. 2 weeks=mean score after 2 weeks.

Table 2: Knowledge quiz results for iterations 1 and 2. IT=iteration; SC=Scenario; IDNK=I do not know.

IT	SC	Mean	SD	IDNK	Score/100	N
1	S1	5,8	1,8	2,4	53	6
1	S2	6,3	0,8	1,5	57	6
2	S1	6,2	1,79	1,2	62	5
2	S2	7,4	1,34	0	74	5

We developed this measure to check the accuracy of participants knowledge in relation to the target responses based on the truthfulness of the statements. Thus, the target for true statements was 6 (Strongly Agree) and the target for false statements was 1 (Strongly Disagree). The lower the closeness value the better. The closeness of true statements was

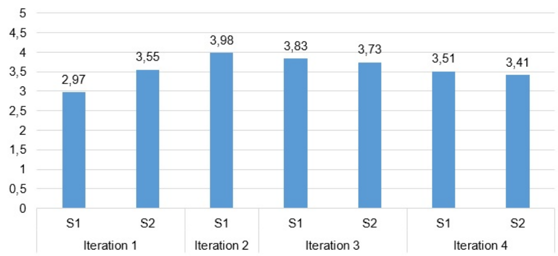


Figure 4: Mean trust scores per iteration and scenario.

slightly lower in Scenario 1 (2.03) than in Scenario 2 (2.17). Overall, Fig. 3a shows the mean scores between the two scenarios were relatively unchanged. It appeared to be more difficult for the participants to determine whether false statements were true or false (Fig. 3b). The closeness score after Scenario 2 (3.00) shows higher accuracy compared to Scenario 1 (3.12).

At this stage, we assume false statements were not answered as accurately because participants were not always able to experience those situations in the driving scenarios. Although it was theoretically possible to test those situations, false statements covered borderline situations that would require participants to compromise driving safety, e.g. waiting to see whether the car will detect static objects on the road (Q2). Similarly, for Q22 and Q23, it was not possible to build the accurate mental model. Interestingly, Q5 has the highest closeness discrepancies. To sum up, the high closeness score of false statements indicates the importance of addressing untrue behaviors and borderline situations in the teaching content to improve the users' mental model representation. These issues represent the work for future developments. Two weeks later, we measured participants' mental model accuracy again. True statements were rated with even better closeness to the target, although just for a fraction, scored at 1.93. The false statements were rated similarly as well compared to the original score, with the closeness score of 3.5. Therefore, we may assume that the accuracy of the mental model is robust in people's minds two weeks after the use without being exposed to the system at all.

4.2 Trust in ADAS

Trust increased in Scenario 2 in iteration 1 across all subscales (Figure 4), compared to Scenario 1 when VA was not available. Perceived Understandability, measuring how well a user can form a mental model of the system, scored particularly well in both conditions. However, the results of the quiz (2) confirm that they reflect participants' perceived understandability and not the actual one. This could be due to participants' learned trust affected by their preexisting knowledge, as explained by Hoff and Bashir (2015). Most participants formed expectations from the system based on their previous interactions with ADAS in real-world cars. In our study, some driving assistance features were not present, such as that the steering wheel was not moving in turns. In other iterations, the trust scores did not change between scenarios, indicating that the choice of tasks to complete or goals to achieve does not affect trust significantly.

4.3 Voice Assistant Evaluation

The SASSI system response accuracy score indicates that the system scored well on recognizing correctly users' input and that it performed as they intended and expected (Table 3). This is expected since human operators (wizards) interpreted the user's input. Likeability had the highest score amongst all factors, which indicates that users had a positive attitude towards the system. Regarding cognitive demand, it seems that driving and interacting with the VA at the same time was quite demanding for drivers. This could potentially be improved by revising the length of the teaching content or by introducing another modality to convey the knowledge, i.e. visuals. Annoyance was rated low (2.96), which implies a positive user experience. Habitability tells us that there is a relatively good match between the user's mental model of the VA and the actual system. Finally, speed was rather low (4.31), presumably because wizards were not always able to produce the output in a short time frame. Participants confirmed that in interviews.

4.4 Discussion and Lessons Learned

We used the log files from the wizard's TTS interface and manually extracted participants' utterances from the video recordings to build the interaction corpus. Table 4 shows the summary of learning objectives that the teaching content for the VA should address. We believe these guidelines will be helpful for VA designers and future studies related to teaching about ADAS. From each video recording, we aggregated participants' statements in interviews and extracted the relevant comments and reflections from them. Using the affinity diagramming technique, we categorized these statements. Below we give short summaries of our findings across all four iterations.

Regarding teaching content, participants agreed that the content was clear and useful. They emphasized the usefulness of the VA when driving a car for the first time, such as when buying a new one or using a rental car. Nevertheless, there should be a way to skip some parts of the learning process. Some par-

Table 3: Mean scores per construct and overall from SASSI questionnaire rated on a 7-point rating scale.

Iteration	Condition	SRA	LKA	CD	ANO	HAB	SP	Overall SASSI	N
1	S2	4,18	5,11	4,68	3,32	4,9	3,9	4,35	6
2	S1	3,82	5,57	4,96	3,1	4,27	3,75	4,25	5
3	S1	4,6	6,2	5,08	3,4	4,25	4,9	4,74	5
3	S2	4,51	5,89	5,36	3,36	4,85	4,9	4,81	5
4	S1	4,42	5,78	5	2,4	4,4	4,2	4,37	6
4	S2	4,27	5,91	4,92	2,2	4	4,2	4,25	6
All	All	**4,3**	**5,74**	**5**	**2,96**	**4,45**	**4,31**	**4,46**	

Table 4: List of learning goals to address in the design of the teaching content.

Domain	Category	Description
ADAS	Explanation	Provide descriptions about ADAS systems available
ADAS	Automation	Explain how driving can be automated using LKA and ACC
ADAS	Takeover	Describe when the takeover request is issued
ADAS	Activation	Explain how to activate driving automation
ADAS	Limitations	Describe the operational boundaries of the driving automation system
ADAS	Takeover	Describe the takeover procedure and how to disable driving automation
ADAS	Speed regulation	Explain how the car speed is set and regulated
ADAS	Controls	Explain the use of buttons to control ADAS to the driver
ADAS	Availability	Justify why driving automation is/is not available
ADAS	Availability	Inform the driver when driving automation is/is not available
ADAS	Speed	Explain how the ACC speed adaptation works
Driver	Monitoring	Explain what happens when driver does not react to takeover requests
Driver	Role	Explain the driver's role while automation is on
User Interface	Icons	Explain the placement and meaning of automation status icons
VA	Usefulness	Inform the driver about what the VA is capable of doing
VA	Navigation	Provide explanation for why a certain route is calculated as best

ticipants thought that this information could be distributed without having to drive or at the red light.

Regarding participants' mental model, we concluded that most of them did not build an accurate understanding of ADAS. In iterations 1 and 2 participants thought LKA was not available or that it could not handle turns because the steering wheel did not physically rotate in turns. There was no trait that would provide a cue in the user's environment to signal the functioning of the LKA. This resulted in a behavior where participants kept turning the steering wheel despite the driving assistance being switched on. To illustrate, P4 said: *"There was only cruise control and there was no lane assist. Or I don't find it or I don't understand it but I don't feel it"*. Furthermore, in iteration 1, some participants claimed that ACC was not adaptive because they did not get to try it while driving. Since the lane of the driver's vehicle was empty (Fig.1a), participants did not try to position their car on the left lanes that carried traffic (L1 and L2). Participants could not determine the capabilities of obstacle detection and avoidance accurately. The reason could be that there were not enough situations to analyse the car's behavior regarding obstacle detection as we only had two obstacles per scenario.

The most common complaint was about the VA's response time. We expected this to occur when the wizards faced a request for which they did not have the answer prepared. However, half of the participants expressed that they had issues identifying what to ask the VA, although they received instructions. In interviews, some participants revealed that it was difficult to understand what the assistant could help them with. For example, P3 said *"It was not easy to find questions let's put it that way"*. P4 said: *"Yes, very useful but we have to know what kind of information we can ask the system"*. It may be that participants felt they were not supposed to learn about the system while driving. As P1 said: *"I don't think that when you are in the car, you should learn the theory about how it works. It is about the practical part when you are driving"*. This suggests that learning by doing might be a better approach than providing the list of tasks the VA supports.

Participants often questioned the usefulness of the VA in the long-term. Although they generally liked the availability of such assistance, they raised the question of its long-term usage. Obviously, the application of the VA to learning is a short-term goal and its features should be extended to other domains, such as traffic information as suggested by P1. P3 said *"Something that came through my mind, it's a little bit to replace the manuals, which are boring"*, which is an existing idea and one of the goals of this study too. P6 said that receiving feedback about the use of driving assistance was not needed; *"But when I avoided the accident, the voice told me "ok you managed to avoid the thing". This was obvious. What is the added value of telling me I did a good job?"*.

In iteration 2, participants seldom interacted with the VA. Conversations were simple and were constructed as a short question and answer. The VA started conversations most times and drivers would shortly respond. Although participants were provided with learning goals, they rarely completed them, in which case wizards decided to proactively start conversations or provide teaching content without demand. P4 said it was easy to focus both on the road and the spoken content, while P5 commented that talking to the car was not a usual practice. In iteration 3, however, participants talked even less with the VA, stating a lack of explicit tasks or reason to do so as a cause. All six participants said it was unclear that they had to start the conversation and think of questions to ask about ADAS. Most times, the VA initiated the interaction and consisted of warning alerts and warning messages about takeover requests. However, when participants did speak to the VA, they tried to request the VA to remind them when to take the turn or to activate ADAS. In iteration 4, the introduction of tasks produced more interaction between drivers and the VA and it seems a promising teaching approach.

5 CONCLUSION

This paper contributes to the topic of the design of conversational agents aimed at driver tutoring about driving assistance systems in vehicles. We compiled a list of learning objectives a teaching assistant should support. Our findings suggest that drivers welcome the idea of a VA for such purposes. However, the purpose, tasks, and functions of such assistants have to be clear to drivers to increase its adoption. Although we provided explanations and limitations of the VA, participants still did not know how exactly they should interact with the assistant, despite their insufficient understanding of the implemented driving assistance. Nevertheless, participants seemed to be eager to explore the possibilities of the VA and were favorable towards the idea of having such a system in their car. However, we believe that this was due to unclear specifications of the purpose and functionalities of the VA. Additionally, the knowledge retention rates remained high two weeks after the learning took place. This indicates that using speech to provide declarative knowledge to drivers might be effective. However, further research is necessary to establish clear design guidelines as well as to focus on the learning and retention of procedural knowledge in various driving situations. Finally, the transfer of knowledge in real-world scenarios is another challenge that needs to be addressed.

ACKNOWLEDGEMENTS

This research has been supported by the Service Public de Wallonie (SPW) under the reference 7982.

REFERENCES

Boelhouwer, A., van den Beukel, A. P., van der Voort, M. C., and Martens, M. H. (2019). Should I take over? Does system knowledge help drivers in making take-over decisions while driving a partially automated car? *Transportation Research Part F: Traffic Psychology and Behaviour*, 60:669–684.

Boelhouwer, A., van den Beukel, A. P., van der Voort, M. C., Verwey, W. B., and Martens, M. H. (2020). Supporting drivers of partially automated cars through an adaptive digital in-car tutor. *Information*, 11(4).

Brüggemeier, B., Breiter, M., Kurz, M., and Schiwy, J. (2020). User Experience of Alexa when controlling music: Comparison of face and construct validity of four questionnaires. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, CUI '20, New York, NY, USA. Association for Computing Machinery.

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., and Koltun, V. (2017). CARLA: An Open Urban Driving Simulator. (CoRL):1–16.

Dow, S., MacIntyre, B., Lee, J., Oezbek, C., Bolter, J. D., and Gandy, M. (2005). Wizard of oz support throughout an iterative design process. *IEEE Pervasive Computing*, 4(4):18–26.

Forster, Y., Hergeth, S., Naujoks, F., Krems, J., and Keinath, A. (2019). User education in automated driving: Owner's manual and interactive tutorial support mental model formation and human-automation interaction. *Information (Switzerland)*, 10(4).

Forster, Y., Hergeth, S., Naujoks, F., Krems, J. F., and Keinath, A. (2020). What and how to tell beforehand: The effect of user education on understanding, interaction and satisfaction with driving automation. *Transportation Research Part F: Traffic Psychology and Behaviour*, 68:316–335.

Frison, A.-K., Wintersberger, P., Oberhofer, A., and Riener, A. (2019). ATHENA: Supporting UX of conditionally automated driving with natural language reliability displays. In *Proceedings of the 11th International*

Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings, pages 187–193.

Hassenzahl, M. (2018). The thing and I: understanding the relationship between user and product. In *Funology 2*, pages 301–313. Springer.

Hoff, K. A. and Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3):407–434.

Hone, K. S. and Graham, R. (2000). Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). *Natural Language Engineering*, 6(3-4):287–291.

Kasuga, N., Tanaka, A., Miyaoka, K., and Tatsuro, I. (2020). Design of an HMI System Promoting Smooth and Safe Transition to Manual from Level 3 Automated Driving. *International Journal ITS Research*, 18:1–12.

Kim, J. W., Ritter, F., and Koubek, R. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science*, 14(1):22–37.

Kocaballi, A. B., Laranjo, L., and Coiera, E. (2018). Measuring User Experience in Conversational Interfaces: A Comparison of Six Questionnaires. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference*, HCI '18, Swindon, GBR. BCS Learning; Development Ltd.

Koo, J., Shin, D., Steinert, M., and Leifer, L. (2016). Understanding driver responses to voice alerts of autonomous car operations. *International journal of vehicle design*, 70(4):377–392.

Kraus, M., Wagner, N., and Minker, W. (2020). Effects of Proactive Dialogue Strategies on Human-Computer Trust. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '20, page 107–116, New York, NY, USA. Association for Computing Machinery.

Larasati, R., Liddo, A. D., and Motta, E. (2020). The effect of explanation styles on user's trust. In *2020 Workshop on Explainable Smart Systems for Algorithmic Transparency in Emerging Technologies*.

Madsen, M. and Gregor, S. (2000). Measuring Human-Computer Trust. In *Proceedings of Eleventh Australasian Conference on Information Systems*.

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., Green, C. A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., and Keinath, A. (2019). The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications. *Cognition, Technology and Work*, 21(1):87–98.

Merriman, S. E., Plant, K. L., Revell, K., and Stanton, N. (2021). Challenges for automated vehicle driver training: A thematic analysis from manual and automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 76:238–268.

Payne, S. J. (2012). Mental models in human–computer interaction. In *The Human–Computer Interaction Handbook*, pages 41–54. CRC Press.

Politis, I., Langdon, P., Adebayo, D., Bradley, M., Clarkson, P. J., Skrypchuk, L., Mouzakitis, A., Eriksson, A., Brown, J. W. H., Revell, K., and Stanton, N. (2018). An Evaluation of Inclusive Dialogue-Based Interfaces for the Takeover of Control in Autonomous Cars. In *23rd International Conference on Intelligent User Interfaces*, IUI '18, page 601–606, New York, NY, USA. Association for Computing Machinery.

Rossi, R., Gastaldi, M., Biondi, F., Orsini, F., De Cet, G., and Mulatti, C. (2020). A driving simulator study exploring the effect of different mental models on adas system effectiveness. In De Paolis, L. T. and Bourdot, P., editors, *Augmented Reality, Virtual Reality, and Computer Graphics*, pages 102–113, Cham. Springer International Publishing.

Rukonic, L., Mwange, M.-A., and Kieffer, S. (2021). UX Design and Evaluation of Warning Alerts for Semi-autonomous Cars with Elderly Drivers. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* - Volume 2: HUCAPP, pages 25–36.

SAE (2018). *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems*. SAE Standard J3016. USA.

Schmargendorf, M., Schuller, H.-M., Böhm, P., Isemann, D., and Wolff, C. (2018). Autonomous driving and the elderly: Perceived risks and benefits. In Dachselt, R. and Weber, G., editors, *Mensch und Computer 2018 - Workshopband*, Bonn. Gesellschaft für Informatik e.V.

Schmidt, M., Minker, W., and Werner, S. (2020). User Acceptance of Proactive Voice Assistant Behavior. In Wendemuth, A., Böck, R., and Siegert, I., editors, *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2020*, pages 18–25. TUDpress, Dresden.

Schnöll, P. (2021). Influences of Instructions about Operations on the Evaluation of the Driver Take-Over Task in Cockpits of Highly-Automated Vehicles. In *5th International Conference on Human Computer Interaction Theory and Applications*, Vienna, Austria. INSTICC, SCITEPRESS.

Sørmo, F., Cassens, J., and Aamodt, A. (2005). Explanation in Case-Based Reasoning–Perspectives and Goals. *Artificial Intelligence Review*, 24:109–143.

Walker, F., Boelhouwer, A., Alkim, T., Verwey, W. B., and Martens, M. H. (2018). Changes in Trust after Driving Level 2 Automated Cars. *Journal of Advanced Transportation*, 2018.

Young, K. L., Koppe, S., and Charlton, J. L. (2017). Toward best practice in human machine interface design for older drivers: A review of current design guidelines. *Accident Analysis & Prevention*, 106:460 – 467.