Carvatar: Increasing Trust in Highly-Automated Driving Through Social Cues

Jens Zihsler
Ulm University, Ulm, Germany
jens.zihsler@uni-ulm.de

Denis Schwager
Ulm University, Ulm, Germany
denis-norbert.schwager@uni-ulm.de

Philipp Hock
Ulm University, Ulm, Germany
philipp.hock@uni-ulm.de

Patrick Szauer
Ulm University, Ulm, Germany
patrick.szauer@uni-ulm.de

Marcel Walch
Ulm University, Ulm, Germany
marcel.walch@uni-ulm.de

Enrico Rukzio
Ulm University, Ulm, Germany
enrico.rukzio@uni-ulm.de

Kirill Dzuba
Ulm University, Ulm, Germany
kirill.dzuba@uni-ulm.de

AutomotiveUI'16 Adjunct Proceedings, October 24–26, 2016, Ann Arbor, MI, USA.
ACM 978-1-4503-4654-2/16/10.
http://dx.doi.org/10.1145/3004323.3004354

Abstract
While driving in highly automated new problems occur which are not present in manual driving. Highly automated cars have different internal states which take into account, what was detected, were the car is and to which maneuvers this could lead. One problem is that the driver may not be able to see in which state the car is currently because the functionality is too complex. If drivers are not able to compare the car's actions to the actions they would have performed themselves they might have trouble trusting the system. It is not apparent what the car recognizes and which future actions are planned. Our avatar uses social cues and anthropomorphism to translate the car's state into human behavior and expressions which can be interpreted intuitively by the driver. The driver is therefore more aware of the situation and might gain more trust in the system.

Author Keywords
Avatar;HMI;autonomous driving;highly automated driving;human-computer-interaction.

ACM Classification Keywords
H.1.2 [driver/Machine Systems]: Human factors; H.5.2 [driver Interfaces]: Prototyping, Voice I/O

Introduction

With increasing automation level the complexity of cars increases proportionally. The majority of people does not have the ability and the motivation to fully understand all functions of such cars. Therefore they might have difficulties trusting these cars. Trust is an important aspect for the drivers to accept highly automated cars [7].

One factor for trust is that the driver is able to compare the actions of the car and actions they would perform in different scenarios [7]. Norman describes the important *gulf of evaluation* by three actions the driver has to perform. In context of a highly automated car the driver has to perceive the car's state, understand it, and finally compare if the action was the desired one [6]. This is in line with the three levels of *situation awareness* (perception, understanding and projection) [2]. In complex cars the drivers have difficulties to see the actions and therefore they tend to mistrust them [7].

Another issue is overtrust in highly automated cars [7]. In June 2016, a man was killed in his self-driving Tesla Model S because a crossing truck was not detected as an obstacle[1]. The driver did not react because the car made him feel safe and he stopped paying attention to the road even though the car told him to monitor the system and the surrounding. The Autopilot function in the Tesla Model S requires the drivers to keep their hands on the wheel and will grind to a halt if they do not do so[2]. He overtrusted [7] the car, kept his hands on the wheel but did not pay attention and therefore he could not prevent the accident by taking control of the vehicle back. This accident shows another problem of human computer interaction with cars that provide partly or fully automated driving. If the driving task is taken away from the drivers they start to get distracted by other tasks and stop paying attention [5].

It is not obvious for the driver what the system is able to detect and what the detection rates are of the systems that are currently available on the market. The Tesla Model S for example shows a red cone relative to the car's position if it detects and an obstacle, but there is no information about the obstacle itself except its position. This does not fulfill the requirements of the first and second level of situation awareness [2] and the gulf of evaluation [6]. If the first levels are broken, the third level also cannot be reached. Therefore drivers will loose situation awareness. The driver could conclude as in the crash example that the system is always correct and feel safe while it's not. This is called *overtrust* in automated systems. Another problem is *mistrust* which is the opposite of overtrust. Mistrust in such systems can lead to bad acceptance rates [7].

People trust other people more if they act similar to themselves, because they compare their own actions with actions of others. Therefore social cues could be a way to abstract and visualize the complex functionality of these cars and create trust in such systems. The target is to create a system which enables the driver to trust the information presented and use these functionalities [7].

Klien et al. [3] describe ten challenges for an automated system incorporating with other agents like a human. One challenge is that the agent and the human have to have a *common ground* [3]. In case of an automated car this common ground can include the knowledge about the surrounding traffic, the traffic rules, what the car anticipates will happen and what it plans to do. The common ground is achieved by indicating important objects and actions

[1] http://www.telegraph.co.uk/news/2016/07/01/tesla-autopilot-crash-driver-was-watching-harry-potter-movie-wit/ (accessed 08/2016)

[2] https://www.tesla.com/presskit/autopilot (accessed 08/2016)

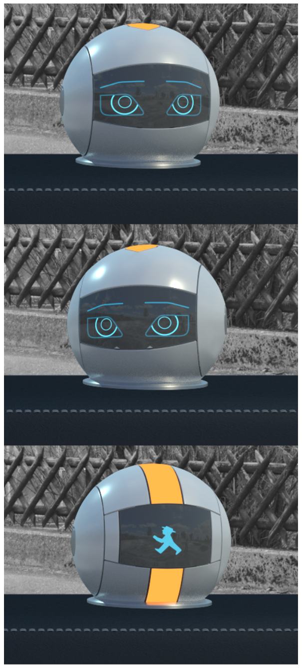

Figure 1: Different rotations of the avatar's head. In the top image the avatar is looking at the back of the car while in the center image it looks at the driver. The bottom image shows the potential of the back screen by displaying symbols for recognized objects. The person symbol can be used when a pedestrian crosses the street in front of the car.

through the gaze direction and additionally through supporting symbols. Because the system should be collaborating with the driver Klien et al. [3] suggest to always maintain information and visualization of the current situation and the overall plans. A common ground could be crucial for the drivers to increase trust in highly automated cars [3].

We propose the usage of an avatar to overcome these issues. The main purpose of the avatar is to facilitate the driver's perception of the system's states and actions. AIDA (Affective Intelligent Driving Agent) has shown that the use of robotic avatars can help to interact more socially with the driver [1]. Our motivation was the assumption that an avatar using social cues could visualize complex functionality more intuitively, especially when it acts like a real driver. We present a concept of a hardware avatar which can be used to visualize different states of the automated system. There is broad range of opportunities and scenarios in which the avatar can be used and we propose that it has the potential to create a higher level of trust between the driver and a highly automated car. For example the rotation of the avatar can show that important objects or other traffic participants are recognized. Another possibility is the visualization of the car's internal action plans like the driving direction. But also interactions with the driver can be realized more socially.

Concept

This concept is intended for highly automated cars where situations might arise in which the driver needs to take over the vehicle and therefore needs to be kept in the loop. To create a common ground it is important that the drivers are able to perceive and understand the current situation and the overall plans [3]. Therefore the avatar imitates the actions of a real driver to make that comparison possible. The main features of the avatar are the gaze direction, internal state representation and dialogues with the driver.

Recognized Objects

The gaze direction of the avatar can be used to imitate a real driver while steering a car. Humans look at an object for a specific amount of time. Such objects could be other traffic participants, traffic signs or the traffic lane. If passengers see the driver looking at a person they are sure that the driver saw the crossing person. The passengers can compare the drivers intention with the action they would perform. If the driver behaves similar to the passenger, the passenger will trust the driver [7].

An automated car does not have a specific gaze direction, its sensors can always detect the complete surrounding without focusing on a specific point. But that is not intuitive for the driver. If the car detects an important object like another traffic participant, person or traffic sign, the avatar rotates so that it is facing the object for a specific amount of time. The driver can see the direction the avatar is facing and is therefore able to see what the car has detected. Klien et al. [3] describe this as *predictability*.

The plans of the car are shown by using the head rotation. For example, the direction in which the car is maneuvering can be easily seen if the avatar looks in the according direction. This could be used for the visualization of the car's future trajectory, the intention to drive forward or backward or to show for example that the car will turn right at an intersection. Another thing that can be shown through the avatar's rotation is that the car is aware of the current traffic situation and the actual right of way. A look at a traffic sign or person shows that the car has detected them and is aware of them. In case of a roundabout traffic and intersections the rotation displays that the car has checked if it has the right of way. This corresponds with the *revealing status and intentions challenge* of Klien et al. [3]. Our avatar

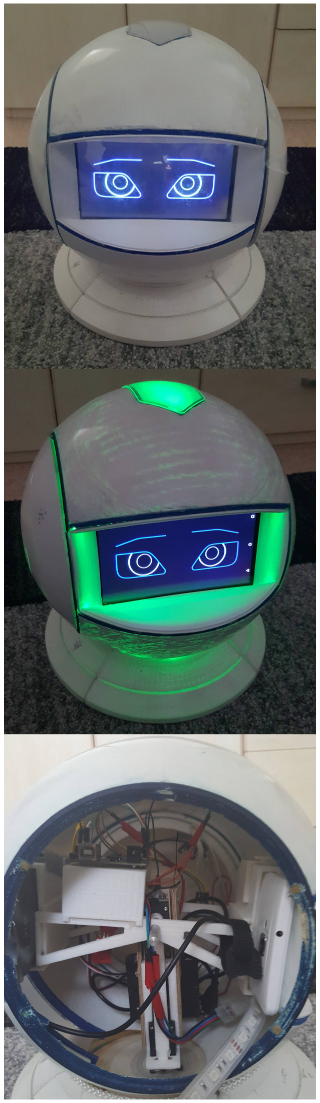

Figure 2: Hardware Prototype Carvatar

makes the actions of the automation predictable, shows the next action and symbolizes the car's state in a way the driver can understand easily.

A display on the back of the avatar can be used to clarify situations. For instance, if the avatar looks at a person crossing the street the display can show that the car has detected the pedestrian. This communicates that the car has detected the person and the driver can follow the car's actions more easily.

Internal State Representation

The color of a light strip can change between colors to show different internal states. It could show how reliable the sensors' data is at the moment, if the travel plan is fulfilled or how probable it is that the driver has to take over.

Dialogues with the Driver

Apart from just visualizing the internal car states the avatar has the possibility to socially interact with the drivers to keep them in the loop. The driver's attention to the surrounding traffic situation could be increased by using small and direct interactions. Moreover, these dialogues could be put inside little games. One such game could for example be asking the driver to find a specific object in the range of the driving path like in the game "I spy with my little eye". Information about points of interest while the car passes them could also help to enhance the driver's attention to their surroundings.

A lot more dialogues are conceivable, like small and clear dialogues to prepare the drivers to take over, to keep them in the loop or to explain features of the vehicle like an assistant that allows automated parking.

Carvatar

A rendered version of Carvatar is shown in Figure 1. In the image the avatar is placed in a car to show three different states. Figure 2 shows our hardware prototype Carvatar that could be placed in a highly automated car to visualize the complex functionality and to interact with the driver socially.

The avatar has an abstract shape modeled after the human head. The avatar has the ability to rotate its head 360 degrees horizontally and 40 degrees vertically. It has one display in the front to display the eyes and one in the back for different symbols like traffic signs or a pedestrian symbol. Additionally a light strip inserted in the head indicates the back of the avatar and can show different colors. For instance, green stands for the car's ability to handle the situation on its own. If the color changes to yellow drivers can prepare themselves for a possible takeover. This could be caused by difficult weather conditions or laws which prevent automated driving in a specific area. The red color stands for critical situations like a system error or situations in which the system is not able to handle a situation.

The three main components are two android based smartphones and an Arduino with a Bluetooth Low Energy (BLE) shield. The smartphone in the front is used as the main controller and communicates the different actions to the second smartphone in the back and to the Arduino. The smartphone in the back can then show different images or play sound. The Arduino controls the horizontal and vertical rotations by using two 360 degree servos. The light strip color is also set by the Arduino.

Interaction Examples

In this section we present some interaction examples to clarify the benefits of our approach using the example of a journey to Ulm University. Figures 3 and 4 show different situations and the according visualization through the avatar. For each part of the journey an according image is referenced in these figures.

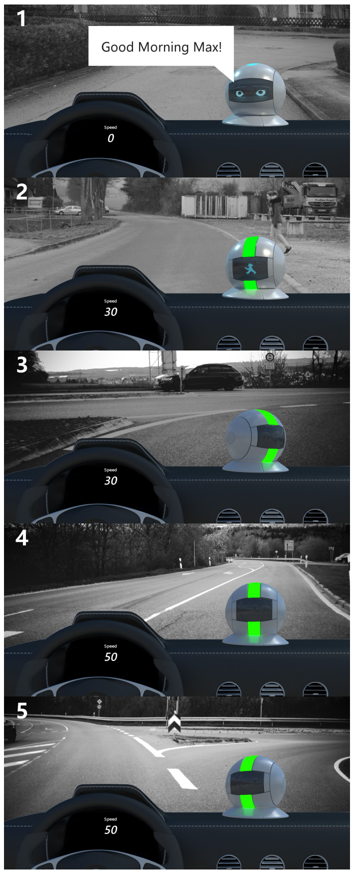

Figure 3: Different Visualizations of car states through the avatar

1 Max takes place in his car. The avatar faces him and welcomes him warmly. Its light strip is blue because they are not driving yet. He then can tell it simply that he wants to start by saying: 'Drive me to the Ulm University'. The avatar confirms and changes its lights to green and looks forward.

2 After leaving the parking spot a person wants to cross the street. The avatar looks at the person on the side, shows a pedestrian symbol on the back and the car slows down and stops. While the person crosses, the avatar follows the person by rotating horizontally. The driver can therefore be sure that the car has recognized the pedestrian. When the person has passed the symbol disappears and the avatar looks forward again.

3 At an intersection they have to turn right. But the cars approaching from the left have the right of way. To show that the car is aware of that, the avatar rotates to left. If the road is clear it rotates to the right and the car turns right.

4 While driving on the road the avatar looks at the car's future trajectory to show that the car knows which road to follow and the driver is notified about the driving direction.

5 After a while they need to turn again. The avatar looks to the right were the car is going to drive to show the driver that a turn is coming.

6 After the intersection the speed limit is 60 km/h. The avatar fixates the sign for a short time and displays the according speed limit symbol on the back.

7 The car has to yield right of way to vehicles already in the upcoming roundabout. The avatar turns left, looking inside the roundabout until the road is free to show that it knows this rule, then it faces slightly right to where the car will go.

8 While driving through a small city suddenly a car pulls out of a street ignoring their right of way. The car brakes, the avatar changes the light to yellow and shows an exclamation mark on the back. The color has changed to yellow to tell the driver that the car might not be able to handle the situation on its own. For instance, that could happen if an appearing car caused an accident with another traffic participant. Then the driver has to take over. If the situation is cleared, the symbol dissolves, the light changes to green again, and the car continues the drive.

9 Just before they arrive at the university they have to drive onto a two-lane road. All cars driving on this street have the right of way which have to be respected. The avatar looks left to check if there is a car in the back on the left hand side (see Figure 4, image 9). If the road is clear the car can change to the other lane. While the car is changing lane the avatar looks to the front and slightly left showing the direction the car will go.

10 An emergency vehicle is now approaching from the rear. The avatar looks to the right side of the street and displays an emergency vehicle symbol while the car is slowing down and clearing the way for the emergency vehicle. While the car performs this action

the color turns to yellow to make the driver aware of the possibility that he might have to take control and drive the car to another position if necessary.

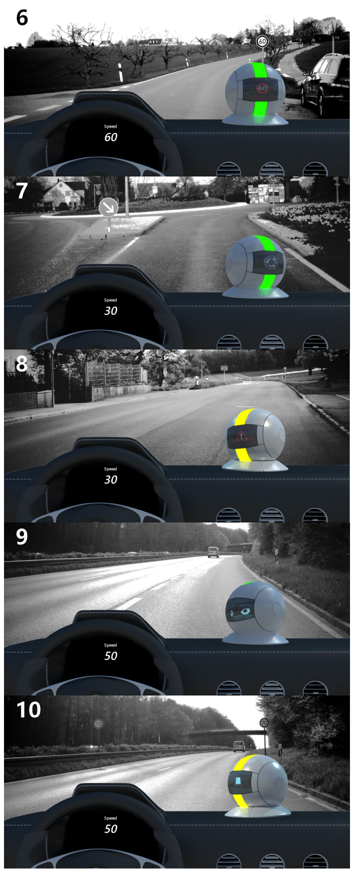

Figure 4: Different Visualizations of car states through the avatar

Conclusion

We presented Carvatar, an avatar that is supposed to facilitate the driver's perception of the automation state by mimicking a human driver and visualizing according information. For the visualization we use human-like movements and reactions which can be understood intuitively (comprehension of data). Therefore drivers could be able to comprehend the car's intentions (projection to the near future) and they might gain more trust in automated cars [7]. Another advantage of this interaction concept is that it is possible to interact with the driver more naturally.

To demonstrate our concept we integrated our avatar in the autonomous driving prototype of the Ulm University [4] and a driving simulation. The simulation is based on the real track which is also currently used to develop the autonomous driving prototype.

The next step is to evaluate the concept in a car and the simulation. For future work several studies are planned which focus on the interaction with the avatar and how this interaction concept compares to commonly used displays in terms of user experience and trust. It is also intended to create new concepts based on the existing Carvatar concept.

Acknowledgment

The authors thank Manuel Stübler and colleagues from the Institute of Measurement, Control and Microtechnology at Ulm University for their valuable feedback and the cooperation regarding the integration of Carvatar in their experimental vehicle.

REFERENCES

1. G. Di Lorenzo, F. Pinelli, F. C. Pereira, A. Biderman, C. Ratti, and C. Lee. 2009. An Affective Intelligent Driving Agent: Driver's Trajectory and Activities Prediction. In *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*. IEEE, 1–4.

2. M. R. Endsley and E. R. Kiris. 1995. Situation awareness global assessment technique (SAGAT) TRACON air traffic control version user's guide. (1995).

3. G. Klien, D. D. Woods, J. M. Bradshaw, R. R. Hoffman, and P. J. Feltovich. 2004. Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems* 19, 6 (Nov. 2004), 91–95.

4. F. Kunz, D. Nuss, J. Wiest, H. Deusch, S. Reuter, F. Gritschneder, A. Scheel, M. Stübler, M. Bach, P. Hatzelmann, C. Wild, and K. Dietmayer. 2015. Autonomous driving at Ulm University: A modular, robust, and sensor-independent fusion approach. In *2015 IEEE Intelligent Vehicles Symposium (IV)*. 666–673.

5. R. E. Llaneras, J. Salinger, and C. A. Green. 2013. Human Factors Issues Associated with Limited Ability Autonomous Driving Systems: Drivers' Allocation of Visual Attention to the Forward Roadway. In *Driving Assessment 2013: Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*.

6. Donald A. Norman. 2013. *The Design of Everyday Things*. Basic Books, Inc., New York, NY, USA.

7. R. Parasuraman and C. A. Miller. 2004. Trust and Etiquette in High-criticality Automated Systems. *Commun. ACM* 47, 4 (April 2004), 51–55.