HAL
open science

On-Road Evaluation of Autonomous Driving Training

Daniele Sportillo, Alexis Paljic, Luciano Ojeda

To cite this version:

Daniele Sportillo, Alexis Paljic, Luciano Ojeda. On-Road Evaluation of Autonomous Driving Training. 14th ACM/IEEE International Conference on Human Robot Interaction, Mar 2019, Daegu, South Korea. hal-02069508

HAL Id: hal-02069508

https://minesparis-psl.hal.science/hal-02069508

Submitted on 15 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On-Road Evaluation of Autonomous Driving Training

Daniele Sportillo
Center for Robotics, MINES ParisTech
Groupe PSA, PSL Research University
Paris, France
daniele.sportillo@mines-paristech.fr

Alexis Paljic
Center for Robotics, MINES ParisTech
PSL Research University
Paris, France
alexis.paljic@mines-paristech.fr

Luciano Ojeda
Groupe PSA
Vélizy-Villacoublay, France
luciano.ojeda@mpsa.com

***Abstract*—Driver interaction with increasingly automated vehicles requires prior knowledge of system capabilities, operational know-how to use novel car equipment and responsiveness to unpredictable situations. With the purpose of getting drivers ready for autonomous driving, in a between-subject study sixty inexperienced participants were trained with an on-board video tutorial, an Augmented Reality (AR) program and a Virtual Reality (VR) simulator. To evaluate the transfer of training to real driving scenarios, a test drive on public roads was conducted implementing, for the first time in these conditions, the Wizard of Oz (WoZ) protocol. Results suggest that VR and AR training can foster knowledge acquisition and improve reaction time performance in take-over requests. Moreover, participants' behavior during the test drive highlights the ecological validity of the experiment thanks to the effective implementation of the WoZ methodology.**

***Index Terms*—Automated Vehicles; Virtual Reality; Augmented Reality; Transfer of Training; TOR; Wizard of Oz; Human-Vehicle Interaction**

I. Introduction

"The more advanced a control system is, the more crucial may be the contribution of the human operator" [1]. When it comes to autonomous cars, Bainbridge's *ironic* statement could sound inaccurate and, indeed, in the collective imagination, drivers are often considered completely passive during the journey. However, before reaching fully automated or driverless cars, the automation level will actually increase in parallel with the necessity of a human driver ready to take over.

SAE categorizes autonomous driving vehicles into 6 levels, from 0 (no driving automation) to 5 (full driving automation) [2]. Objectives of this study are conditionally automated cars (SAE Level-3), which can perform the entire driving task without human supervision, with the expectation that the driver is receptive to requests to intervene and will respond appropriately. In other words, during the autonomous driving, the driver is enabled to perform secondary activities (reading, using a tablet, watching movies) without monitoring the driving environment. However, when the vehicle reaches a system boundary and thus becomes unable to perform the task at hand, the human driver is expected to take over the driving task within a reasonable amount of time after being prompted with a Take-Over Request (TOR). Regaining control of the car in a safe and attentive manner is crucial for the safety of the driver and the other road users.

Furthermore, according to the Product Liability Directive of the Council of the European Union [3], a product is considered defective when "it does not provide the safety which a person is entitled to expect", including "the use to which it could reasonably be expected that the product would be put". In other words, putting an automated driving system on the market without providing adequate information and formation to the end users would be not allowed. For these reasons, there is the need for car manufacturers to familiarize future drivers with the car interfaces and the interaction modalities. Indeed, the drivers should be aware of the capabilities and limitations of the system and of the actions to perform when their intervention is required. Traditional car handover performed by car dealers may not be sufficient to ensure the acquisition of skills by the customers.

Therefore, a training program is essential to allow safe interaction and to foster the correct acquisition of the operational skills. Our hypothesis is that Virtual and Augmented Reality (VAR) technologies may represent valuable solutions for this purpose, allowing drivers to discover and test car functionalities in a controlled environment without putting them and other road users at risk.

In this paper we address this research topic by conducting a user study with two main purposes. The first objective is the evaluation of VAR systems for the training of drivers of conditionally automated vehicles. The second aim is to assess the transfer of training to the real scenario with a test drive on public roads. Since (at the time of writing) in Europe Level-3 automated cars are not allowed on public roads without special licenses, the test drive was conducted applying the Wizard-of-Oz (WoZ) protocol: it resulted in making the participants believe that the vehicle was driven by an automated driving system when it was actually controlled by a human driver. To the best of our knowledge, this is the first study about autonomous driving which include a WoZ test drive on public road with inexperienced unaware participants.

This research was supported by the French Foundation of Technological Research under grant CIFRE 2015/1392 for the doctoral work of D. Sportillo at Groupe PSA.

II. Related work

Conditionally automated vehicles represent a fascinating research topic for different communities from a variety of perspectives. What makes these robots appealing for the HRI community is the vastness of the public reached and the crucial role of human intervention.

A. Driver interaction with automated vehicles

The out-of-the-loop problem is a well known potential negative consequence of automation, which leaves "operators of automated systems handicapped in their ability to take over manual operations in the event of automation failure" [4]. In recent years, the behavior of drivers when they resume control of automated vehicles has been extensively studied [5]–[7]. Results from this research have led to important findings about take-over performance [7] depending on the criticality of the situation [8] and the secondary activity [9], [10].

The importance of training for automated car drivers has gained relevance only in recent times. Previously, when an autonomous driving study was conducted, the participants were only briefly informed about the functioning and the interface of the car by the experimenter before starting the test. Recent work instead extensively investigates the impact of training and familiarization with autonomous driving and take-over requests [11]–[13]. Payre et al. [11] analyzed the effectiveness of a training based on simple practice in the simulator with respect to a more elaborated training which included a text, a video and a more complex practice in the simulator. Sportillo et al. [12] compared three different training supports: a user manual displayed on a laptop, a fixed-base driving simulator and a Light Virtual Reality system. Four different familiarization groups (no familiarization, description, experience, description and experience) were instead compared in a driving simulator study by Hergeth et al. [13]. As a result, all the authors agreed that a more elaborated training allows participants to increase trust in the autonomous system and to respond faster to take-over requests.

B. Virtual and Augmented Reality

Virtual and Augmented Reality aim at enhancing user experience by providing enriched perception and interaction in computer-generated scenarios; however, there are several differences which can make one of these technologies more suitable to the research question we address. In particular, AR provides the possibility to naturally interact with the real car equipment enhanced with additional information overlaid on top of it [14]. The narrow field of view of current AR headsets, however, may limit the possibility of providing a fully encompass experience to the users.

Virtual Reality instead creates an artificial surrounding environment, enabling total isolation in the virtual environment and providing the possibility to simulate fully immersive driving scenarios allowing drivers to be trained in an environment that enables correction, repetition and non-dangerous failure [15].

The application of Virtual and Augmented Reality to HRI is a powerful concept which is gaining more and more interest [16] in the last years. In this field, successful utilizations of VAR have been proven in particular for applications related to human-robot collaboration [17], teleoperation [18], [19] and robot skill acquisition [20]. Besides HRI, VAR platforms have been historically adopted in education [21], [22], training [23], [24], aviation [25] with important benefits in terms of performance [24], users' involvement and motivation [21] and transfer to real settings [25].

If until a few years ago VAR platforms were destined to a niche audience, the democratization and the spread on a large scale of *prêt-à-porter* immersive devices, such as headsets, made their adoption more accessible in terms of cost, ease of implementation and setup [26]. This radical innovation has allowed researchers to address not only experts and professionals [23], [25], but also general public.

C. Wizard of Oz for Autonomous Driving

All the studies about automated vehicles mentioned in Section 2.1, which included a test drive, were conducted in high-end driving simulators usually consisting of a real vehicle surrounded by panoramic displays. The few studies involving an on-road driving experience concerned a lower level (typically SAE Level-2) of automation [27], [28].

Real-life driving experiences can lead to a better understanding of vehicles' limitations and to improvements in trust calibration of automated vehicles [28]. To date, conducting an experiment with automated vehicles requires strict authorizations; furthermore, for security and legal issues, only drivers with a special license are usually allowed to drive in predetermined and controlled stretches of road. All these limitations are necessary and legitimate, but they confine user tests to experts in the field and keep final users distant from them. Driving simulators reduce this gap, but, although their effectiveness has been proven for traditional cars [29], very little is known about their validity when the level of automation increases.

The Wizard of Oz protocol represents a suitable research methodology for allowing subjects to interact with a system they believe autonomous but which is actually controlled by a human. Although in some HRI studies [30] the use of this protocol has been judged controversial for what concerns ethical issues and embarrassment related to participants' deception, when it comes to autonomous vehicle interaction, the potential benefits of the WoZ would make this protocol appropriate for conducting valid experiments with the general public in real driving scenarios. In fact, a robust illusion would make hard for the subjects to guess that someone is controlling the car; as a consequence, it would allow participants to behave more similar to the real case.

The use of the WoZ to simulate autonomous driving is not a novelty [31]–[35]. Most of the implementation of the WoZ in vehicles includes a dummy steering wheel with no function. In a recent study, Wang et al. [32] presented Marionette, a system built on the RRADS platform [31] able to simulate Level 3 and 4 autonomous driving. The interest of this system is the ease of implementation in commercial vehicles in terms of cost

and effectiveness. However, the driving task is still performed by the *pilot wizard* at all time by interpreting participants' input on the steering wheel and the pedals; this can introduce lags between the input and the action and break the deception. The main difference between these implementations and ours is that our participants used fully functional controls and they actually drove the vehicle when it was in manual mode. Moreover, they were completely unaware that the person next to them was controlling the car.

III. The User Study

This study aimed to make people ready to drive a Level-3 (Conditional Driving Automation) automated car. To do so, in a between-subject study, 60 participants with valid driving licenses, were randomly assigned to one of three groups and trained with different methods: an on-board video tutorial, an Augmented Reality training program and a Virtual Reality simulator. After the training, all the participants drove a prototype of an automated car on a public freeway. For security reasons, it was not possible to perform a test with an actual autonomous car on public roads. Consequently, the car was not actually autonomous, but it was controlled by a human pilot unbeknownst to the subjects. Since the test involved a modified version of a commercial car, a license plate for prototype vehicles was obtained, and an authorization for the study was issued by the local ethics committee. The study lasted about 2 hours for each participant.

A. The target vehicle

The vehicle used for the test drive was a Citroën Grand C4 Picasso, suitably modified for the experiment. The car was a right-handed driving vehicle with automatic transmission to which fully functional steering wheel, pedals and gear shift were added on the left-hand (Fig.4a). To detect the presence of the hands and the foot on the steering wheel and the pedals respectively, these controls were equipped with force sensors. The vehicle had 5 possible states which are described in Fig.1.

During Autonomous Driving, the vehicle respected the speed limit, adapting the speed in order to maintain the safety distance from the preceding car; however, the vehicle did not perform overtaking or lane changes.

The Human-Machine Interface inside the car included an on-board computer with two screens, a sound system and the Autonomous Driving button. The two screens (behind the steering wheel, and in the central console), were used to display information about the car and to provide the possibility, during the autonomous driving, to perform secondary activities such as watching a movie and playing games. The Autonomous Driving button was placed in the central console near the gear shift. When the Autonomous Driving mode was available, the driver could push the button to activate it. All the car's state changes were notified to the driver with visual-auditory alerts which consisted of displaying an icon on the screens and playing a sound and a vocal message.

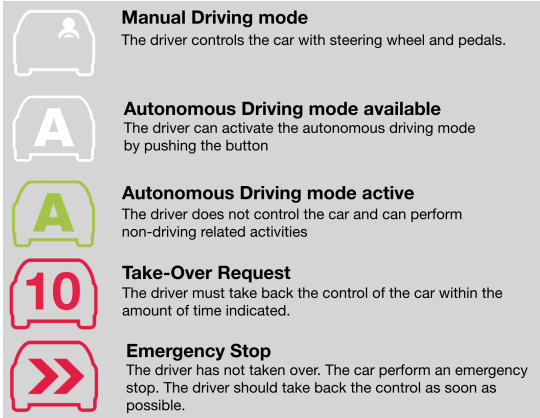


Fig. 1: The description of each vehicle state and the associated icon

B. The training

The main objective of the training was to provide the participants with the essential knowledge to interact, in a safe and effective manner, with the target vehicle by themselves. The training was designed to not require the intervention nor the presence of the experimenter; however, for the purpose of the study, an experimenter was always present.

In this study, three different self-administered training methods were designed and evaluated: an on-board video tutorial (*Video*), a Virtual Reality simulator (*VR*) and an Augmented Reality training program (*AR*).

The training included a common video for the three conditions (displayed according to the system) in which the following information was given: the purpose of the training was explained, and the main characteristics of the Level 3 autonomous driving were introduced; afterward, the 5 states of the car were presented along with the icon and the visual-auditory alerts representative of each state. In addition to the video, the AR and VR groups included simple simulated driving scenarios in which the trainee could practice the activation and the deactivation of the automated driving system and experience two TORs (roadwork and exit from the highway).

In what follows, the three training systems are presented and the motivation for their choice is explained.

1) The on-board Video tutorial: The training based on the on-board video tutorial consisted in simply displaying the informative video on the central screen of the car. The only interaction required of the participant was to localize the *autonomous driving button* and to push it when required.

This condition represented the baseline for the study. The training provided the participants with the bare essential knowledge needed to interact with the car. This training provided the lowest level of interaction: in fact, the localization of the "Autonomous driving button" was the only active part. Including a practice of the take-over procedure, given that this condition does not allow to display a driving scenario from the viewpoint of the user, was considered counter-intuitive. Also, it represents the easiest system to implement and, at the

(a) On-board Video Tutorial

Real vehicle ✓
Driving scenario **X**

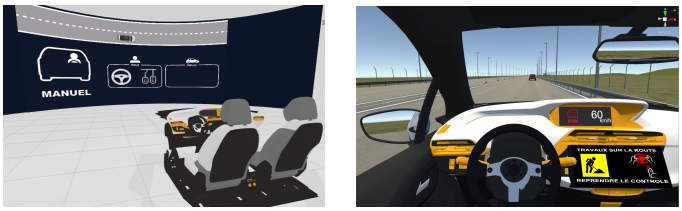
(b) Augmented Reality Program

Real vehicle ✓
Driving scenario ✓

(c) Virtual Reality Simulator

Real vehicle **X**
Driving scenario ✓

Fig. 2: The three training systems and the characteristics: the on-board video, (b) the Augmented Reality training program (manipulated image), (c) the VR system



(a) (b)

Fig. 3: The VR simulator: (a) the learning environment; (b) a practice driving scenario.

time of writing, the most commonly implemented; in fact, car companies are more and more providing embedded manuals in the form of electronic documents and videos to allow their customers to get easier access to the information about their car.

2) The Virtual Reality simulator:
The interest of Virtual Reality for this study is given by the fact that it would allow to train drivers everywhere (in car dealerships as at home) without the need for the real vehicle. The VR hardware setup included a HTC Vive and a Logitech G27 racing wheel and pedals (Fig.2c).

At the beginning of the experience, the user was virtually placed in a reality-virtuality airlock aimed at smoothing the transition from the real to the virtual world (Fig.3a). This environment included a car cockpit (similar to the real one) and the two front seats; a curved screen was placed in order to display the introductory videos.

In order to enable direct and natural interaction between the user and the virtual car, a finger tracking device (a Leap Motion controller) was added to the headset: this allowed the participants to use and see a representation of their hands while operating the steering wheel and the HMI inside the virtual car.

After the presentation of each car state, a short driving simulation session was performed by the participants in order to apply what they learned in the previous steps. In the first driving session, to familiarize users with the car controls, they were required to drive the car in manual mode in a very simple driving environment: a single lane with no traffic. In the subsequent session, the users were required to drive on a highway and to activate the Autonomous Driving when it was available. The last two driving sessions concerned take-over requests. In the first one, the users had to take-over because of the presence of roadwork on the lane. In the second one, they had to take-over to exit the highway.

When dealing with Virtual Reality, it is important to take into account *simulator sickness*. In driving simulation in particular, longitudinal and lateral accelerations contribute to the incoherence between what we see and what our body feels (visuo-vestibular conflicts). In the implementation of the simulator, these conflicts were reduced by designing straight or slightly curved roads with low optic flow in the peripheral vision.

3) The Augmented Reality training program: The Augmented Reality training took place inside the car. The headset used for this training was a Microsoft HoloLens. Also in this case, the training was totally self-administrated. To enable the communication between the HoloLens and the car, the headset was connected to the car's WiFi network; this allowed the HMI inside the car to be updated according to the training content and allowed the training to advance when the correct user action was detected.

The interior of the car was augmented with (i) a virtual screen in which the training content was displayed, (ii) a blinking sphere to highlight the activation button, (iii) an arrow as gaze guidance toward the activation button. Since AR allows first-person view content, similar to the VR condition, the same driving scenarios were displayed when a user action was required (autonomous driving activation and take-over requests).

To summarize, the main differences between the three conditions were related to the coherence between the training and the operational environment (real vehicle vs 3D model of the car) and the presentation of driving situations in which the trainee could practice activation/deactivation of the autonomous system and experience TORs.

C. The Wizard of Oz setup

To validate the effectiveness of a training program, it is important to assess to what extent trainees are able to transfer skills acquired from the training environment to the real case. With this purpose, and to satisfy security and liability requirements, the presented study proposes the implementation of the Wizard of Oz protocol to simulate autonomous driving in order to make drivers believe that they were actually interacting with an automated system.

The experimental setup included the presence of two *Wizards* in the car: (1) the *pilot wizard* who drove the car when the *Autonomous Driving Mode* was active and (2) the *Interaction Wizard* who performed the sensing part, analyzed the driving environment and sent appropriate notification to the participant.

The participants were told that since they were driving a prototype, a test engineer (the *pilot wizard*) was legally required to sit on the passenger seat to ensure the correct functioning of the car and to intervene in case of emergency. To hide the pilot's controls and to help participants believe that the car was actually autonomous, a panel was placed between the two seats. The panel did not cover the entire height, so the participant could still have a partial view of the right side. The participants were in charge of the driving task only when the vehicle was in *manual mode*, but they could take over the control at any time.

1) The Pilot Wizard: The Pilot Wizard was in charge of the driving task when the vehicle was in *autonomous mode*. From the display placed behind his steering wheel, he could know the current state of the car and if the participant was touching the steering wheel or the pedal. When the car was in manual mode, the Pilot Wizard paused to control the car, but he was always ready to intervene in case of emergency. Prior to the study, he was adequately trained to behave as a Level-3 automated driving system.

2) The Interaction Wizard: The Interaction Wizard sat in the back seat of the car. He was in charge of analyzing the driving environment and controlling the HMI and the state of the car from a laptop connected to the car. Moreover, he talked to the participants and logged valuable information during the test drive.

The HMI was controlled with a software running on a computer placed in the trunk. The computer provided a wireless access point and a server with a webpage displaying a user interface from which the *Interaction Wizard* could update the state of the car could be updated and send TORs.

D. The test drive

After the training, all the participants performed a test drive on a public road. The aim of the test drive was to assess how people interacted with the vehicle in a real life driving scenario.

The participants were informed about the presence of the *Pilot Wizard*, who was required to ensure the correct functioning of the car and about the itinerary of the test drive, which was already stored in the GPS system of the car. The experimenters clarified that the test drive was not part of the training; for this reason, to interact with the car, the participants had to rely on the knowledge learned during the training beforehand.

As instructed, during the automated driving, the participants were free to perform non-driving related tasks (NDRTs). On the on-board computer they could choose between watching a movie or playing some games. Moreover, they were allowed to use their own phone to do what they wished. Sleeping was not allowed. As the vehicle did not perform lane changes, the driver was also free to take-over in order to overtake another vehicle.

All the test drives were performed during daylight. The weather during the test drive varied from clear to cloudy (slight rain in one case); however, this study does not take into account the weather variable.

1) The driving scenario: The participants drove for about 25 kilometers on a public freeway (dual carriageway with central barrier). The stretch of road used for the test drive is known for heavy traffic in particular time slots and, unfortunately, for accidents (mostly collisions with no serious consequences). Moreover, during the weeks in which this test took place, roadwork was scheduled and carried out in a short stretch of road. The road works caused a narrowing of the carriageway that a Level 3 ADS could not handle. All these features were relevant to the study, in particular because the participants could face all the three types of TOR presented during the training.

The itinerary included 3 planned TORs. The first TOR was a 50-second "End of the autonomy zone" TOR launched after about 7 kilometers. The drivers were required to take-over, exit from the freeway, and re-enter in the opposite direction after a roundabout. Afterwards, they could re-activate the Autonomous Driving. After 5 kilometers, a 30-second "Roadwork" TOR was launched. The TOR was justified from the temporary road marking, and the narrowing of the carriageway due to the presence of traffic cones (which however did not require a lane change). After 5.5 kilometers, a 50-second "End of the autonomy zone" TOR was launched. Before definitely exiting the freeway to end the test drive, a 10-second TOR was launched, for no apparent reason to the driver. The itinerary is illustrated in Fig.4c. In addition to these requests to intervene, supplementary urgent 10-second TORs could be sent in case of emergency. This happened, for example, in presence of heavy traffic on the entrance ramps, signaled accidents on the road, or stationary cars in the lane. Moreover, because of too intense traffic or blocked roads, in few cases it was necessary to make small changes to the itinerary.

E. Participants

A panel of sixty volunteers (N = 60) was recruited by a company specialized in hiring consumer tests participants. The participants, 29 men and 31 women, were aged between 25 and 73 (mean age 46.2, SD 14) and they had a valid driving license and no previous experiences with autonomous cars. However, some of them had previously used some automated

(a) (b) (c)

Fig. 4: The test drive: (a) the interior of the car with the participant on the left and the *pilot wizard* on the right; (b) a participant makes a phone call while the car is in autonomous mode; in the background, the *interaction wizard*; (c) the default test drive itinerary on the public freeway

TABLE I: Demographic features distributed across the different systems

System	Gender (F/M)	Age y (SD)	Age Group <36 / [36,56] / > 56	Car with Cruise Control? Yes (no use) / No
Video	10/10	46.3 (15.1)	7/6/7	15(9) / 5
AR	10/10	46.7 (14.4)	6/7/7	13(7) / 7
VR	11/9	45.8 (13.3)	7/6/7	11(2) / 9
Total	31/29	46.2 (14)	20/19/21	39(18) / 21

driving functions such as Cruise Control and Lane Keeping Assistance.

The panel was divided into three homogeneous groups of 20 participants, equally distributed in gender and age groups (less than 36 years old, between 36 and 56, more than 56 years old). Each group was trained with one of the systems previously described. Details of the demographic features are reported in Table I. At the end of the experiment, each participant was rewarded with a 45 euro voucher.

IV. Results

All the participants completed the study (no dropout occurred during the training or the test drive). To evaluate how the training and the test drive affected participants' impression of and opinion about autonomous driving, they filled out the same set of questions three times: at the beginning of the study (with no prior knowledge of autonomous driving), after the training and after the test drive.

After the training phase, the participants answered a post-hoc questionnaire to evaluate the training phase, and a Knowledge Test which required them to classify autonomous driving scenarios, identify interfaces in the car, and explain activation and deactivation procedure of the system.

It is known that the exposure to VR systems, and in particular immersive headsets, can produce a feeling of sickness in some users: limiting its occurrence is crucial for any VR application. The Simulator Sickness Questionnaire (SSQ) was used to evaluate the simulator sickness produced by the VR simulator [36].

Drivers' take-over performance was evaluated with the reaction time, defined as the elapsed time from TOR until the driver takes back control. This measure has been used and validated as a performance metric in all the take-over studies. Since the prototype used in the test drive was not equipped with sensors, it was not possible to evaluate other well-known

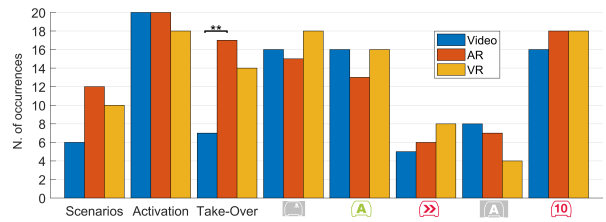


Fig. 5: Correct answers to the Knowledge Test for each training system (**p <.01)

metrics such as the position in the lane and the time to collision from a vehicle ahead.

In addition to these measures, two cameras inside the car recorded the drivers' behavior and a live log annotated by the *Interaction Wizard* during the test drive. All the data was synchronously collected and anonymously stored according to the local privacy policy. If not differently specified, all the variables were tested for group difference using ANOVAs for continuous normally distributed data and the Kruskal-Wallis test for categorical, ordinal and non-normally distributed data. A significant level of 5% was chosen for all the tests.

A. Objective measures

1) Knowledge Test: The maximum score possible of the Knowledge Test (KT) was 13. Summarizing the answers of the KT a significant difference was observed between the scores of the AR and Video group ($Video = 8, AR = 10, VR = 9; H_{(2,57)} = 8, p < 0.05$). This difference was mainly due to the questions concerning driving scenarios understanding ($Video = 6, AR = 12, VR = 10; ns$) and the description of take-over procedure ($Video = 7, AR = 17, VR = 14, H_{(2,57)} = 13.81, p < 0.01$).

Concerning the icons of each state, the hardest to identify were the ones indicating the availability of the autonomous mode and emergency stop (19 correct and 41 wrong answer).

2) Reaction Time: In total, 234 TORs were correctly computed, including 128 TOR of 50 seconds, 51 TORs of 30 seconds and 55 TORs of 10 seconds. Although in the test drive there were some fixed predetermined situations (exit from the highway and roadwork), unplanned circumstances required participants to take-over: this implies that the number

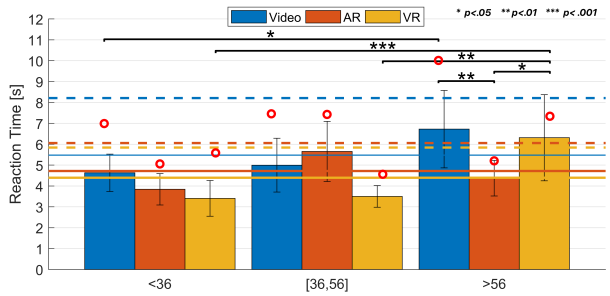


Fig. 6: Take Over reaction time (RT) according to group age and training system. Red dots are the RTs of the first TOR for each group. In dashed lines, the mean of the RT to the first TOR. In solid lines the mean of the RT to all the TORs.

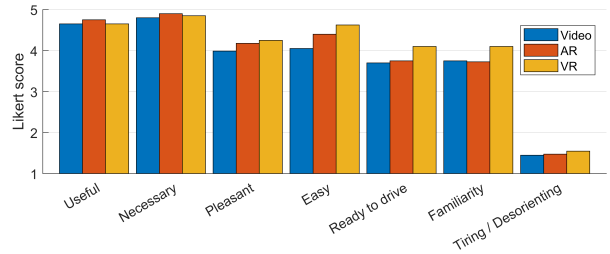


Fig. 7: The results of the questionnaire for the training phase evaluation per age group (up) and total (down).

and the order of TORs were not the same for all the subjects. Note that since reaction time data was not normally distributed we conducted non-parametric statistical tests, expect from the 2-way ANOVA.

VR training produced the lowest reaction times in all the three types of TOR, but only for what concerns the urgent 10-second TORs, this difference is significative ($rt_{Video10} = 3.07s$, $rt_{AR10} = 3.12s$, $rt_{VR10} = 2.08s$; $H_{(2,52)} = 9.04$, $p < .05$). From the analysis of reaction time it can be observed that even when the available time budget triples (TOR 30) or quintuples (TOR 50), drivers reacted very quickly anyway. If the difference between the 3 types of TOR is considered negligible ($rt_{TOR10} = 2.50s$, $rt_{TOR30} = 3.83$, $rt_{TOR50} = 4.46s$; $H_{(2,231)} = 28.8$, $p < .001$), it is possible to average all the reaction times to have a more general view (Fig.6, solid lines). In this case, a significant difference can be observed, but only between VR and Video ($rt_{Video} = 4.50s$, $rt_{AR} = 4.04$, $rt_{VR} = 3.47s$, $H_{(2,231)} = 8.31$, $p < .05$). Since the training groups were also equilibrated in terms of age, a 2-way ANOVA was performed with the age groups and the training conditions: it can be noted that the oldest group benefits from the AR training in a statistically significant way ($rt_{AG3Video} = 6.72s$, $rt_{AG3AR} = 4.37s$, $rt_{AG3VR} = 6.31s$, $F_{(4,225)} = 3.18, p < .05$). On the other hand, VR training seems to be more effective for the youngest groups. If only the first TOR is taken into consideration (Fig.6, dashed lines), the participants trained with the Video tutorial reacted around 2 seconds slower than the VR and the AR groups, but not in a significant way ($rt_{1stVideo} = 8.2s$, $rt_{1stAR} = 6.1s$, $rt_{1stVR} = 5.8s$, ns). In particular, the highest reaction time to the first TOR was observed for the third group of age trained with the Video tutorial.

B. Self-reported measures

1) Training evaluation: The training part was evaluated by the participants with a 5-point Likert scale survey about perceived usefulness, easiness of understanding and familiarity. The results reported in Fig.7 show that there are no significant differences for questions related to the training in general (usefulness and necessity of training). However, when it comes to questions specific to the training system, VR seems the preferred in terms of familiarity with the vehicle, easiness of understanding, and readiness to drive, in particular for the third age group. In order to validate the results for the VR group, the SSQ was filled out by the involved participants. 50% of them reported 0 for all the symptoms. On a maximum possible total score of 78.54, the mean was 5.0, which represents the limit for negligible symptoms in the categorization of score proposed by Kennedy [36].

2) Perceived trust and usefulness of autonomous driving: The participants were asked to indicate their level of agreement (on a 5-point Likert scale) with a set of sentences about the concept of autonomous driving. They filled out the same questionnaire three times: before the training, after the training and after the test drive. The questions were grouped in three categories (Fig.8): (i) trust in automation, (ii) perceived usefulness of the autonomous driving, (iii) willingness to perform a NDRT. Friedman's test used as repeated measure test, showed a statistically significant increment in the three sets of questions for all the training conditions (Trust: $H_{(2,38)} = \{$Video: 14.7, AR: 20.07, VR: 17.84$\}$, Usefulness: $H_{(2,38)} = \{11.11, 17.76, 13.21\}$, NRDT: $H_{(2,38)} = \{12.12, 20.38, 11.48\}$). For no question there was a decrease while the higher increment was noticed in the questions "I can imagine myself doing other tasks than driving" and "I trust the driving decisions made by the vehicle" after the test drive.

In addition, the participants evaluated on a 5-score Likert scale to what extent the training helped them interact with the vehicle in the various situations (activation, take-over, recognizing alerts) during the test drive ($Video = 4.6, AR = 4.7, VR = 4.6; ns$).

V. Discussion

A first important outcome of this study is that all the sixty participants in this study had the feeling of being in an actual autonomous car and their illusion was total: in other words, no subject realized that the car was actually driven by a human pilot. Although further work is necessary to validate the Wizard-of-Oz protocol for the autonomous driving research, we observed that it represents a robust and effective research methodology that allows for an assessment of the general public's interaction with autonomous vehicles in real driving

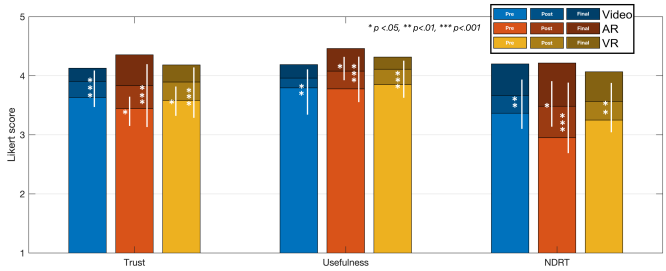


Fig. 8: Likert responses to the pre, post and final questionnaire about perceived trust and usefulness of autonomous driving and willingness to perform a NDRT.

scenarios. The three training programs proposed in this paper were designed to make drivers ready to operate their vehicle by providing general understanding of the autonomous system, the rules of use, and by supporting the know-how to interact with it.

All the trained participants transferred the training to the real scenario: during the test drive they were able to correctly activate the autonomous driving and safely take over in all the situations in the given time, without the need to perform an "emergency stop". The AR and the Video training both took place in the real vehicle and, in addition, the AR and the VR training included the presentation of simulated driving scenarios in which the drivers could practice the interaction with the system and Take-Over Requests. We suppose that this difference may have influenced the results of the Knowledge Test (KT) and the performance in the TORs.

Although not in a statistically significant way, the participants trained with AR and VR tended to react faster to the TORs in general, and to the first one in particular, than those trained with the Video. However, it is not clear, and we do not have sufficient elements to make a strong claim, if these results are due to the immersion level or to the practice of a TOR during the training. The role of the first TOR is crucial for the purpose of this study since a first bad takeover may be already dangerous or compromise future uses of the system. Taking into account also the age groups, it can be observed that while for the first two age groups the reaction time was comparable among all the training systems, older participants trained with the Video tutorial reacted slower than the others; the group trained with Augmented Reality, instead, was the fastest inside the age group and reacted in a time comparable to younger groups. The outcomes of this study about take over time are in line with related work in the field in which it is proven that subjects who execute a take over during the training performed better in the test drive [12], [13]. It has to be said that the participants were not obliged, only invited, to perform a secondary task: this might not guarantee the same level of distraction for all the subjects at the moment of the TOR. However, video analysis showed that older drivers were usually more unwilling to perform secondary activities. The results about objective measures suggest that the different age groups would benefit from different training systems. In terms of reaction time, if VR seems to be more convenient for young and middle-aged drivers, older people would take more advantage from the AR training.

In general, the participants judged the training programs useful and necessary for the purpose of using the automated system. The VR simulator however produced better results in terms of easiness of understanding, readiness to drive and familiarity with the vehicle; this can be explained by the fact that this training provided a higher sense of immersion and isolation, which may have allowed the participants to better familiarize themselves with the car and the driving situations they would face. In addition, the simplified cockpit of the virtual car and the bare virtual environment may have helped users focus their attention on the interfaces relevant to the training. The VR system received good scores in particular from the third age group; although this sounds in contrast with the objective results, it may underline a difficult for older drivers to transfer the training skills from the virtual environment to the real scenario.

Self-reported measures pointed out also the importance of the test drive for what concerns trust in automation, perceived usefulness and willingness to perform a NDRT. This result was expected since, thanks to the implementation of the WoZ protocol, the functioning of the automated system was ideal. Also, as some participants admitted, the presence of the experimenters in the vehicle during the test drive reassured and helped them in having a pleasant driving experience.

VI. Conclusion

Human interaction with automated vehicles is fundamental as it is required to respond to system limitation in potentially dangerous situations. Familiarizing drivers with the vehicle beforehand is crucial for the correct understanding and use of the automated system, and for road safety. We conducted the first study on a public road aimed at comparing three programs for the drivers' training of conditionally automated cars (SAE Level-3). The application of the Wizard of Oz protocol played a central role in this study; it allowed us to assess transfer of training to the real circumstances and to evaluate driver behavior during an authentic driving experience, satisfying current safety and liability requirements.

Results show that the training is necessary to have a better understanding of the system capabilities and limitations and to increase people perception of trust and usefulness in the automated vehicle; even simple and non-interactive training programs (on-board video tutorials) help drivers in localizing the interfaces and recognizing the alerts. However, participants trained with Virtual Reality and Augmented Reality had generally a better understanding of the take over procedure and better performance in term of reaction time during the test drive. This study offers the insight that customized training programs should be considered according to the age groups. For this reason, further within-subject studies are necessary to explore user preferences about the training programs. Longer test drives should be conducted in order to validate the current results.

REFERENCES

[1] L. Bainbridge, "Ironies of automation," in *Analysis, Design and Evaluation of Man–Machine Systems 1982*. Elsevier, 1983, pp. 129–135.

[2] S. International, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," 2018.

[3] C. Directive, "Council directive 85/374/eec of 25 july 1985 on the approximation of the laws, regulations and administrative provisions of the member states concerning liability for defective products," *Official Journal L*, vol. 210, no. 07/08, pp. 0029–0033, 1985.

[4] M. R. Endsley and E. O. Kiris, "The out-of-the-loop performance problem and level of control in automation," *Human factors*, vol. 37, no. 2, pp. 381–394, 1995.

[5] N. Merat, B. Seppelt, T. Louw, J. Engström, J. D. Lee, E. Johansson, C. A. Green, S. Katazaki, C. Monk, M. Itoh *et al.*, "The "out-of-the-loop" concept in automated driving: Proposed definition, measures and implications," *Cognition, Technology & Work*, pp. 1–12, 2018.

[6] N. Merat, A. H. Jamson, F. C. Lai, M. Daly, and O. M. Carsten, "Transition to manual: Driver behaviour when resuming control from a highly automated vehicle," *Transportation research part F: traffic psychology and behaviour*, vol. 27, pp. 274–282, 2014.

[7] C. Gold, D. Damböck, L. Lorenz, and K. Bengler, ""take over!" how long does it take to get the driver back into the loop?" in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2013, pp. 1938–1942.

[8] A. Eriksson and N. A. Stanton, "Takeover time in highly automated vehicles: noncritical transitions to and from manual control," *Human factors*, vol. 59, no. 4, pp. 689–705, 2017.

[9] K. Zeeb, A. Buchner, and M. Schrauf, "Is take-over time all that matters? the impact of visual-cognitive load on driver take-over quality after conditionally automated driving," *Accident Analysis & Prevention*, vol. 92, pp. 230–239, 2016.

[10] S. S. Borojeni, L. Weber, W. Heuten, and S. Boll, "From reading to driving: priming mobile users for take-over situations in highly automated driving," in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 2018, p. 14.

[11] W. Payre, J. Cestac, N.-T. Dang, F. Vienne, and P. Delhomme, "Impact of training and in-vehicle task performance on manual control recovery in an automated car," *Transportation research part F: traffic psychology and behaviour*, vol. 46, pp. 216–227, 2017.

[12] D. Sportillo, A. Paljic, and L. Ojeda, "Get ready for automated driving using virtual reality," *Accident Analysis & Prevention*, vol. 118, pp. 102 – 113, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0001457518302197

[13] S. Hergeth, L. Lorenz, and J. F. Krems, "Prior familiarization with takeover requests affects drivers' takeover performance and automation trust," *Human factors*, vol. 59, no. 3, pp. 457–470, 2017.

[14] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators & Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.

[15] L. Jensen and F. Konradsen, "A review of the use of virtual reality head-mounted displays in education and training," *Education and Information Technologies*, pp. 1–15, 2017.

[16] T. Williams, D. Szafir, T. Chakraborti, and H. Ben Amor, "Virtual, augmented, and mixed reality for human-robot interaction," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2018, pp. 403–404.

[17] S. Makris, P. Karagiannis, S. Koukas, and A.-S. Matthaiakis, "Augmented reality system for operator support in human–robot collaborative assembly," *CIRP Annals-Manufacturing Technology*, vol. 65, no. 1, pp. 61–64, 2016.

[18] D. Whitney, E. Rosen, E. Phillips, G. Konidaris, and S. Tellex, "Comparing robot grasping teleoperation across desktop and virtual reality with ros reality," in *Proceedings of the International Symposium on Robotics Research*, 2017.

[19] M. A. King and G. L. Paderewski, "Virtual reality teleoperated remote control vehicle," Aug. 22 2000, uS Patent 6,108,031.

[20] T. Zhang, Z. McCarthy, O. Jowl, D. Lee, X. Chen, K. Goldberg, and P. Abbeel, "Deep imitation learning for complex manipulation tasks from virtual reality teleoperation," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–8.

[21] L. Freina and M. Ott, "A literature review on immersive virtual reality in education: State of the art and perspectives." *eLearning & Software for Education*, no. 1, 2015.

[22] J. Bacca, S. Baldiris, R. Fabregat, S. Graf *et al.*, "Augmented reality trends in education: a systematic review of research and applications," 2014.

[23] N. Gavish, T. Gutiérrez, S. Webel, J. Rodríguez, M. Peveri, U. Bockholt, and F. Tecchia, "Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks," *Interactive Learning Environments*, vol. 23, no. 6, pp. 778–798, 2015.

[24] A. C. Boud, D. J. Haniff, C. Baber, and S. Steiner, "Virtual reality and augmented reality as a training tool for assembly tasks," in *iv*. IEEE, 1999, p. 32.

[25] J. Vince, "Virtual reality techniques in flight simulation," in *Virtual Reality Systems*. Elsevier, 1993, pp. 135–141.

[26] S. Kuntz, R. Kulpa, and J. Royan, "The democratization of vr-ar," *Virtual Reality and Augmented Reality: Myths and Realities*, pp. 73–122, 2018.

[27] A. Eriksson, V. Banks, and N. Stanton, "Transition to manual: Comparing simulator with on-road control transitions," *Accident Analysis & Prevention*, vol. 102, pp. 227–234, 2017.

[28] F. Walker, A. Boelhouwer, T. Alkim, W. B. Verwey, and M. H. Martens, "Changes in trust after driving level 2 automated cars," *Journal of advanced transportation*, vol. 2018, 2018.

[29] I. Milleville-Pennel and C. Charron, "Driving for real or on a fixed-base simulator: is it so different? an explorative study," *Presence: Teleoperators and Virtual Environments*, vol. 24, no. 1, pp. 74–91, 2015.

[30] L. D. Riek, "Wizard of oz studies in hri: a systematic review and new reporting guidelines," *Journal of Human-Robot Interaction*, vol. 1, no. 1, pp. 119–136, 2012.

[31] S. Baltodano, S. Sibi, N. Martelaro, N. Gowda, and W. Ju, "The rrads platform: a real road autonomous driving simulator," in *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 2015, pp. 281–288.

[32] P. Wang, S. Sibi, B. Mok, and W. Ju, "Marionette: Enabling on-road wizard-of-oz autonomous driving studies," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017, pp. 234–243.

[33] D. Rothenbücher, J. Li, D. Sirkin, B. Mok, and W. Ju, "Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*. IEEE, 2016, pp. 795–802.

[34] G. Schmidt, M. Kiss, E. Babbel, and A. Galla, "The wizard on wheels: Rapid prototyping and user testing of future driver assistance using wizard of oz technique in a vehicle," in *Proceedings of the FISITA 2008 World Automotive Congress, Munich*, 2008.

[35] A. Habibovic, J. Andersson, M. Nilsson, V. M. Lundgren, and J. Nilsson, "Evaluating interactions with non-existing automated vehicles: three wizard of oz approaches," in *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2016, pp. 32–37.

[36] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The international journal of aviation psychology*, vol. 3, no. 3, pp. 203–220, 1993.