Transportation Research Part C 126 (2021) 103090

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

On-road virtual reality autonomous vehicle (VRAV) simulator: An empirical study on user experience

Xin Zou [a], Steve O'Hern [b], Barrett Ens [c], Selby Coxon [d], Pascal Mater [e], Raymond Chow [e], Michael Neylan [f], Hai L. Vu [a,*]

[a] Institute of Transport Studies, Monash University, Clayton, VIC 3800, Australia
[b] Transport Research Centre Verne, Tampere University, Tampere 33014, Finland
[c] Faculty of Information Technology, Monash University, Caulfield East, VIC 3145, Australia
[d] Faculty of Art, Design & Architecture, Monash University, Caufield East, VIC 3145, Australia
[e] Department of Civil Engineering, Monash University, Clayton, VIC 3800, Australia
[f] Department of Human-Centred Computing, Monash University, Caufield East, VIC 3145, Australia

A R T I C L E I N F O

Keywords:
Driving simulator
Virtual reality
Evaluation study
Autonomous vehicle
Wizard-of-Oz

A B S T R A C T

Autonomous-vehicle (AV) technologies are rapidly advancing, but a great deal remains to be learned about their interaction and perception on public roads. Research in this area usually relies on AV trials using naturalistic driving which are expensive with various legal and ethical obstacles designed to keep the general public safe. The emerging concept of Wizard-of-Oz simulation is a promising solution to this problem wherein the driver of a standard vehicle is hidden from the passenger using a physical partition, providing the illusion of riding in an AV. Furthermore, head-mounted display (HMD) virtual reality (VR) has been proposed as a means of providing a Wizard-of-Oz protocol for on-road simulations of AVs. Such systems have potential to support a variety of study conditions at low cost, enabling simulation of a variety of vehicles, driving conditions, and circumstances. However, the feasibility of such systems has yet to be shown. This study makes use of a within-subjects factorial design for examining and evaluating a virtual reality autonomous vehicle (VRAV) system, with the aim of better understanding the differences between stationary and on-road simulations, both with and without HMD VR. More specifically, this study examines the effects on user experience of conditions including presence, arousal, simulator sickness and task workload. Participants indicated a realistic and immersive driving experience as part of subjective evaluation of the VRAV system, indicating the system is a promising tool for human-automation interaction and future AV technology developments.

1. Introduction

Autonomous-vehicle (AV) technologies, which offer the possibility of fundamentally changing transportation, are rapidly advancing, and are soon expected to become commonplace. AVs promise to improve safety, reduce congestion, improve mobility, and deliver more sustainable transportation (Haboucha et al., 2017). However, preparing for their widespread adoption will require a firm understanding of how they are perceived and accepted by both the passengers and the general public (Panagiotopoulos and

* Corresponding author.
E-mail address: hai.vu@monash.edu (H.L. Vu).

https://doi.org/10.1016/j.trc.2021.103090
Received 18 August 2020; Received in revised form 24 February 2021; Accepted 12 March 2021
Available online 25 March 2021
0968-090X/

Dimitrakopoulos, 2018). At its core, AV is human-automation interaction (HAI) research, as it "involves automation in the form of an embodied, situated agent, which is used by non-professional users in a time-sensitive, safety–critical context" (Janssen et al., 2019, p. 101). Research into HAI often relies on the use of driving simulators (Greenlee et al., 2018; Li et al., 2019) or trial studies (i.e., field tests) (Banks and Stanton, 2016; Eriksson et al., 2017). The major challenge of the former is to replicate naturalistic driving conditions in terms of inertial forces and vehicle dynamics, while the latter is expensive, and faces legal obstacles designed to keep the general public safe (Eriksson et al., 2017).

One proposed method of gaining insights into passenger perceptions is to use Wizard-of-Oz simulations, wherein the driver of a standard vehicle is hidden from the passenger using a physical partition, providing the illusion of riding in an AV (Wang et al., 2017). Much simpler and cheaper to maintain compared to existing lab simulators, and offering a greater degree of freedom, the Wizard-of-Oz approach is a realistic alternative to simulation, allowing the rider to experience the physical forces produced by motion in naturalistic driving conditions (senses such as acceleration/deceleration and the rocking sensation of passing over the humps). In addition, this approach is useful in on-road AV research even in the absence of legal constraints (Kim et al., 2020). More recently, head-mounted display (HMD) virtual reality (VR) has been investigated as a means of enabling Wizard-of-Oz studies (Nebeling and Madier, 2019). HMD VR not only can provide a full view of an AV interior, but can also be used to simulate external events involving other vehicles, pedestrians, or other objects that may be unsafe or impractical to produce in a real-world environment. Nevertheless, despite its advantages, little is known about the realism of such a system in comparison to other existing simulation approaches.

The research presented in this paper fills the above knowledge gaps by conducting an empirical study of user experiences in an on-road VR simulation which is capable of a level-3 (conditionally automated; SAE International, 2018) AV. The study compares VR- and non-VR-based simulations in both on-road (Wizard-of-Oz) and off-road (lab-environment) conditions. The factorial design allows us to independently contrast the effects of VR-based and on-road Wizard-of-Oz simulations on participants' subjective and objective perception of the simulator.

To the best of our knowledge, this is the first study offering a comprehensive overview of user experiences with on-road Wizard-of-Oz simulations, and is based on the introduction of a virtual reality autonomous vehicle (VRAV) system. Developing such a system, taking advantage of low-cost VR technology, has a great potential to create experiences that are more realistic, immersive and cost-effective than the existing conventional fixed-base driving simulators, and allows for greater flexibility and safety in the test environments than on-road AV driving. Nevertheless, it is also necessary to examine whether such system will induce greater simulator sickness and increased task workload. By no longer requiring a real AV, researchers can test and develop novel interfaces and interaction patterns inexpensively and safely. The contributions made in this study in terms of a VRAV system and research protocol can allow future researchers to run on-road studies with controlled events, simulate autonomous driving and conduct HAI studies in a high-fidelity environment.

The remainder of the paper is organized as follows: **Section 2** reviews relevant prior research. **Section 3** includes the VRAV system proposal and study design. **Section 4** presents results of the evaluation to reveal participant experiences, as well as comparisons among study conditions. **Section 5** concludes the paper by highlighting areas for future research.

2. Related work

This section summarises research relevant to the use of VR in transportation research and on road AV studies. The review considered the use of VR in driving simulators, AV research in on-road studies, on-road AV simulation and previous simulator validations.

2.1. VR use in driving simulation

VR headsets offer an excellent alternative to conventional simulators, with the advantages of being cheaper, more immersive and having the ability to easily accommodate multiple users (Pai Mangalore et al., 2019). Several studies have investigated the use of VR in driving simulations.

Specifically, Sportillo et al. (2018) looked at whether VR can be an effective tool for training potential AV users in best practices for takeover of partially automated vehicles. They found that participants in a fixed-base simulator and VR systems responded faster to takeover requests than users trained on laptop systems. Agrawal et al. (2018) developed a VR, headset-based latent hazard anticipation and mitigation training program (V-RAPT), finding a significantly greater proportion of latent hazards than drivers who had completed PC-based training programs. Pai Mangalore et al. (2019) evaluated the use of VR headsets for measuring driving performance (latent hazard anticipation behaviours), with results showing that VR headsets can be used to effectively measure driver performance. At least in hazard anticipation scenarios, VR headsets were found to generate minimal simulator sickness (Pai Mangalore et al., 2019).

2.2. AV use in on-road studies

Several studies have made use of AV technology to increase external validity to the study of user perceptions and behaviour. Due to the current state of systems, these studies have tended to focus on levels 2 (partial automation) or 3 (conditional automation), which require the driver to always be aware of the vehicle's surroundings, and be ready to take control (SAE International, 2018).

Eriksson et al. (2017) conducted a study to determine whether observed behaviours, such as non-critical control transitions in driving simulators, correlate with observations in on-road AVs. Although there was a strong positive correlation, they found that drivers were quicker to take control in on-road than in simulated driving.

In a study of level-2 AVs, Shutko et al. (2018) observed that the off-road glancing behaviour was similar to that of drivers of manually operated vehicles. They categorized drivers into two groups: active drivers, who keep their hands on the steering wheel, and supervisors, who mainly keep their hands away from it. The study found that whether participants kept their hands on the steering wheel or not is not related to eyes on/off road glance behaviour.

2.3. On-Road AV simulation

Since AV technologies are still in their infancy, researchers have sought to reduce the cost of user studies by introducing Wizard-of-Oz protocols to simulate AVs. The first such protocol was the real-road autonomous-driving simulator (RRADS) platform (Baltodano et al., 2015), introduced by researchers at Stanford University. This platform uses a partition to shield the driving "wizard" from the participant's view. Rather than deceive participants, the protocol relies on suspension of disbelief, providing the illusion of riding in an AV. To show that such a protocol enables ecologically valid testing, the researchers conducted a study to evaluate a haptic feedback device that warns of the AV's intentions. Results present that the platform may be an effective tool to evaluate prototypes and scenarios specific to open road human-AV interactions.

Wang et al. (2017) later built on the RRADS platform with their Marionette system, which provides the driving wizard cues about participant intentions. Using a false steering wheel and pedal interface, it gives the driver information about when he or she should move the steering wheel or press the accelerator. and aims to increase the ability of researchers to provide a low-cost platform for the study of user reactions in partially and conditionally AVs at levels 2 and 3. The study demonstrated that the Marionette facilitates the illusion of partially autonomous (level 2) driving while maintaining the safety for on-road studies.

Our VRAV prototype improves on the implementation suggested by Goedicke et al. (2018), who introduced VR for on-road, Wizard-of-Oz simulations. The VR provides a view of an AV interior and the participant's own hands, tracked by a display-mounted sensor. We conducted a pilot validation study with 6 participants. We build on this work by improving the fidelity of the simulation and the user experience, as explained in the following section.

2.4. Validity of driving simulators

Validity is "the extent to which the results of studies with simulation correlate to the results in reality" (Zhao and Sarasua, 2018, p. 384). Currently, a vast amount of research evaluating the different types of validity, such as behavioural validity and ecological validity (Mullen et al., 2011), of driving simulators has been conducted, with most studies assessing their absolute and relative validity (Blaauw, 1982), especially by comparing simulated speeds with those in the real world (Törnros, 1998; Godley et al., 2002; Knapper et al., 2015). Methods of assessing simulator validity have not been standardized, and therefore considerable variation in validation-study methodologies exists (Wynne et al., 2019). Scholars usually validate specific aspects of driving simulators based on their research purposes (e.g., they validate speed in order to conduct speed research). For example, Törnros (1998) used a driving simulator to assess driver behavior in a simulated road tunnel by comparing the speeds and lateral positions of 20 subjects; relative validity was good for both, and it was determined that the simulator could be used to study problems related to these. By using a driving simulator, and comparing the speeds under distracting conditions in real and simulated environments, Knapper et al. (2015) verified the relative validity of the simulator for use in speed under distracting conditions research.

Subjective assessments of research settings are also important for addressing validation; these make use of the perceptions and experiences of participants. For example, Llopis-Castelló et al. (2016) conducted a driving-simulator validation by way of driver perceptions, with most participants assessing the quality of the virtual environment, and the similarity of the simulated driving tasks to those in the real world, as moderate or high. Hussain et al. (2019) validated a fixed-base driving simulator by assessing its subjective validity (the quality and performance of the simulator) using a post-test questionnaire covering comfort, speed perception, graphics and so on to compare it with the real world. The result was that the participants deemed the simulator to have a high level of realism.

Note that the above work focused on the users' experience specifically to validate the use of driving simulators, whereas other research was concentrating on human behaviours and their psychology through the users' experience while using the driving simulators including (Du et al., 2019; Khastgir et al., 2018; Jamson et al., 2008). In particular, Du et al. (2019) conducted a within-subject experiment in a driving simulator with 32 participants to investigate the effects of the timing of AV explanations and the degree of autonomy on drivers' trust, preference for AV, anxiety and mental workload (using NASA-TLX). To explore the effect of knowledge about the automation capability on trust in the automated system, Khastgir et al. (2018) conducted a driving simulator study with two different types of automated systems (low and high capability). In a driving simulator investigation by Jamson et al. (2008), participants experienced two forward collision warning (FCW) systems, and questionnaires were administered to give self-reports for mental effort using and user acceptance of the FCW.

Currently, to our knowledge, no study has offered subjective evaluation of driving simulators from a comprehensive range of aspects together, namely presence, arousal, acceptance, simulator sickness and realism. For this study, VRAV simulation was validated by way of subjective evaluation (participant perceptions) obtained through questionnaires. This also allowed the negative effects of the simulations to be determined, which are always an issue with driving simulators (Llopis-Castelló et al., 2016).

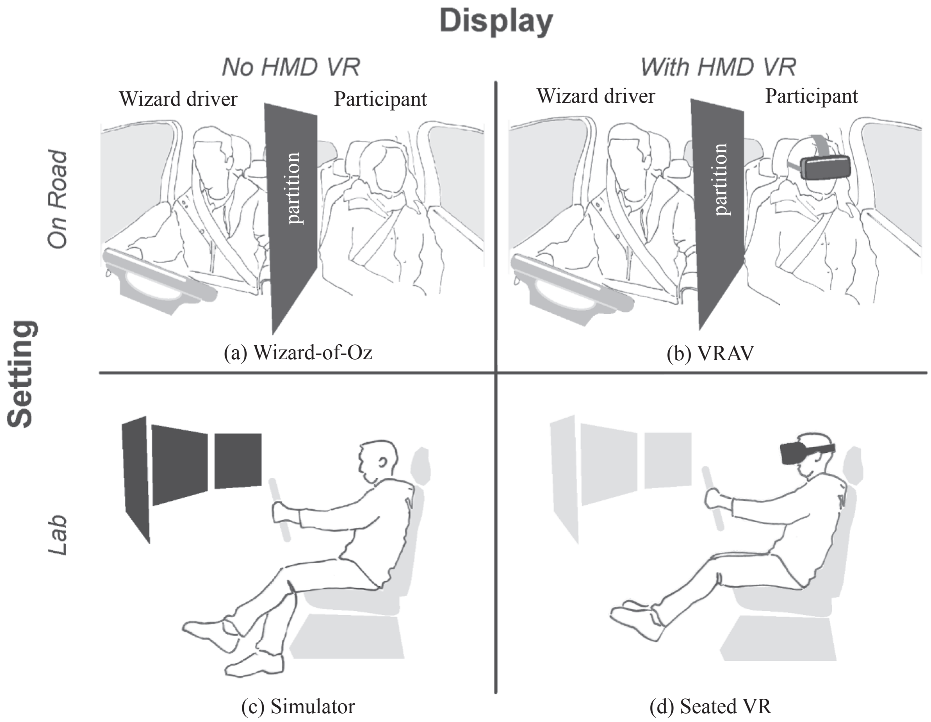


Fig. 1. Multiple variations of our simulation design are created to fulfil each study condition.

3. VRAV system design

This work is the initial step of a longer-term project aimed at understanding the perception of AVs and their responses while interacting with other entities on the public roads. We predict that immersive technologies such as VR, and in later stages augmented reality (AR), will play an important role in providing affordable and flexible tools to enable the study of user perceptions by allowing conditions and events to be simulated at a relatively low cost. This work takes an important step toward this goal by attempting to determine how an on-road VR simulation compares to the original Wizard-of-Oz and lab simulations.

3.1. Initial workshop

Before beginning development of our VRAV simulation, we conducted a workshop to gather insights on its potential applications. This initial workshop involved round-table discussions made up of small groups of experts from both the industry and research community in AVs. Their areas of expertise included AR/VR, HMI, in-vehicle design, driving simulation, and road safety. In the first step of the workshop, common goals were revealed using a Goals Grid (i.e., what to achieve, avoid, preserve and eliminate) (Nickols, 2015). Stated goals included making the experience as visually realistic as possible while eliminating negative aspects such as motion sickness, and scenario safety operation on public roads. For the second step, scenarios were designed and modelled through a process of "object storming," leading to the creation of imagined application scenarios, for instance scenes of streets with the infrastructure and expected impediments to an AV's progress. Feasibility of these modelled scenarios was assessed using a C-Box chart, based on level of innovation and difficulty of implementation. A consensus was reached that, within the available research time frame, a re-creation of an uneventful and unimpeded on-road driving scenario would best fulfil our key aspirations of implementing and validating an improved VRAV proof of concept. Full details of the workshop procedure and included in **Appendix A**.

3.2. Design Goals

Highly automated test vehicles are still rare, and (independent) researchers often have limited access to them. In addition, developing fully functioning system prototypes is time-consuming and costly (Habibovic et al., 2016), and on-road tests involve many legal and safety issues. The goal of the current study is to design a VRAV simulation system that can be used to explore HAI (and

Fig. 2. Monash connected and autonomous vehicle and driving simulator.

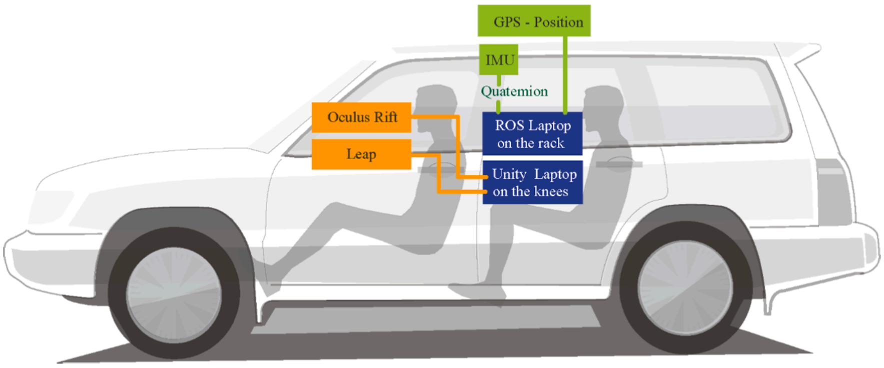


Fig. 3. VRAV system diagram.

ultimately, levels of resistance to AV adoption) in a controlled environment at minimal cost and risk.

To understand the user experience of on-road VR simulations, we need to compare them with a variety of other protocols. To create these, we chose to divide the study design into two factors (see Study Design, below, for a full description):

- Display (with and without HMD VR)
- Setting (on-road and lab)

We developed a simulated AV experience that can be used in each of the four set-ups that result, as described in the following sub-section.

3.3. Study design

We created four different user experiences, corresponding with the conditions created by the Display-Setting matrix in Fig. 1, in order to make an overall comparison of our VRAV with a traditional simulator using screens or HMD VR, and a real car without HMD VR.

- ***Wizard-of-Oz*** (*On Road, No HMD VR*) — Following the Wizard-of-Oz protocol used by Wang et al. (2017), with the participants in the front passenger seat of a real vehicle along with a driver. A partition is placed in the centre of the vehicle to block the driver from the participant's view (Figs. 1 and 7).

- ***VRAV*** (*On Road, With HMD VR*) — Here the participant rides in the front passenger seat as in the Wizard-of-Oz setup, but wears a VR display (an Oculus Rift with a resolution of 640 × 800 per eye) which provides a view of the car interior, with an empty seat where the driver should be. The participant sees an avatar representing his or her body and hands, and a virtual replica of the outside environment through the vehicle windows, while feeling and hearing the real engine and outside environment. Our implementation closely follows the proof-of-concept VR-OOM system demonstrated by Goedicke et al. (2018); however, we aim to create a more engaging experience through a high-fidelity simulation.
- ***Simulator*** (*Lab, No HMD VR*) — The participant sits in a driving simulator, *Australia DS*, viewing the same virtual scene as in the VR conditions displayed on three 1280 × 768 resolution monitors at a frequency of 59 Hz, while listening to artificial engine noise through speakers.
- ***Seated VR*** (*Lab, With HMD VR*) — This is similar to the Simulator condition, except that the participants view the simulated scene through a VR display instead of the simulator screens, while listening to artificial engine noise through headphones.

We describe the technical details of each implementation in the following sub-sections.

3.3.1. Vehicle and tracking

For the on-road simulation we use a 2004 Subaru Forester XT (Fig. 2) equipped with a power inverter that provides AC power for use by the laptop and other on-board equipment. Vehicle position and motion are tracked using a Swift Navigation Piksi Multi GPS receiver and an Xsens MTi-3 IMU mounted within the vehicle (Fig. 3). Data from a local mount point 1 km away from the test site is obtained through the AUSCORS correction service, which allows GPS correction to an accuracy of 1 cm. Additional data can be obtained through the vehicle OBD-II port or ABS sensors. However, the data available and update rates vary among models and manufacturers. While modern vehicles have higher minimum standards, the vehicle used was manufactured before these standards were implemented. While we were able to obtain velocity information, we found the roughly 10 Hz data rate too slow for practical use.

The robot-localization package in the Robot Operating System (ROS) is used to input the position and rotation data into a digital model of the car to estimate vehicle position and speed. An unscented Kalman filter (UKF) with the package's default parameters is used. Specifically, position is input using an East-North-Up frame, with the origin of this frame set along the route. Angular velocities from the IMU are also input, but absolute angle relative to north (heading) is not. It was found during testing that the IMU was unable to accurately and consistently provide stable angle information due to magnetic-field interference. Heading is therefore determined by deriving the GPS position and fusing it with angular velocity. This method only works if the car is in motion, and therefore the heading cannot be determined at the beginning using this method. Since the route is the same for all runs, however, an initial heading is specified instead.

3.3.2. Driving route

The simulation replicated a real 1.0 km stretch of road adjacent to our local campus; it takes roughly 2 min to cover under typical driving conditions. We chose this location because of its convenience, and because the low volume of traffic help to ensure safety. The route also has some interesting features, such as a large roundabout, several speed humps, and recognizable campus buildings. The complete route is shown in Fig. 4.

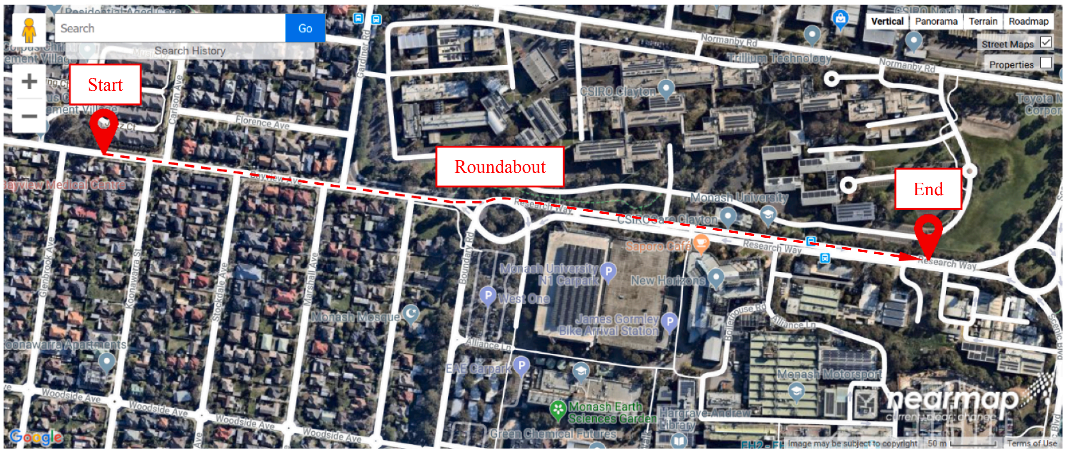


Fig. 4. The route used in our simulation overlaid on an aerial map. **Source:** nearmap.com.

3.3.3. Simulated environment

Our simulated environment was created using Unity 2018.2.16f1, with a frame rate of 90 fps. To avoid a sensory disconnect, as was reported by participants in preliminary evaluations conducted by Goedicke et al. (2018), we recreated the real route with a high level of fidelity, including details such as road surface, speed humps and adjacent buildings. To further enhance fidelity, our virtual car had realistic suspension physics and care was taken to model and place the speed humps accurately in the simulated environment in order to mimic the sensation of driving over the humps (Figs. 5 and 6). Pedestrians and other cars are automatically generated on the map and follow paths that will not interact/interfere with the autonomous vehicle (i.e., they are passive agents). It is not an exact replica of what is occurring in the real world, but is to give a sense of realism to the virtual environment. Furthermore, hypothetical scenarios can be set up in the virtual environment to study dangerous or rare events without compromising the safety of the participant and/or the public.

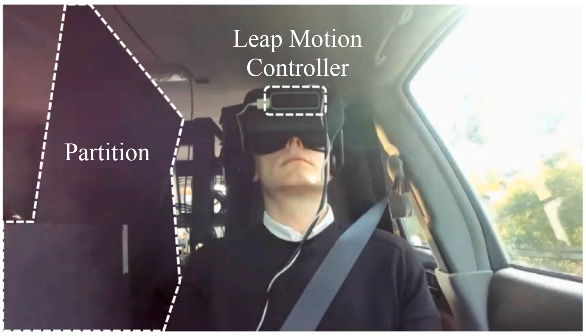
As in the VR-OOM system (Goedicke et al., 2018), we used a Leap Motion controller to track participants' hands, which were replicated within the scene (Figs. 7 and 8). We also included an androgynous avatar to provide a sense of presence if a participant decided to look down (Fig. 8), which was not included in the VR-OOM. Note that although the participant was seated on the passenger side of the vehicle, they actually saw themselves in the driver seat (and in the right side of the vehicle) in the virtual environment. A replica of the steering wheel and brake were developed for the participants, while care was taken to model the 3D car interior to the same dimensions as the real car to ensure if the participant reached out for the wheel, both their hand and the virtual hand would reach it in a similar fashion. To synchronize with the participant's actions, an indicator can be placed next to the wizard driver which informs the rotation and pressure of the imitated wheel and brake. Nevertheless, this might still be difficult for the wizard driver to maintain synchronization and consistencies (Wang et al., 2017). However, note that the purpose of this study is not about emergency and control handover but about the development of the driving simulator itself, as well as to comply with ethics procedures and local laws, we did not provide participants in the On Road experiments with the haptic steering wheel and pedal interface as used in the Marionette (Wang et al., 2017) or VR-OOM (Goedicke et al., 2018) on-road systems, since this would have covered the vehicle's front-passenger airbag, posing a safety hazard. Other options (having some kind of collapsible column to mount the steering wheel onto) to allow safe mounting of the steering wheel while not blocking the airbag were also investigated, but to comply with local laws, these mounting options would have needed to be engineered to collapse.

Fig. 5. Aerial view of the location used to compare the 3D virtual simulation (upper) and the real world (lower). **Note:** Locations of speed humps are marked with red pins. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 6. Views of speed humps in both the virtual (upper) and real worlds (lower) at the same time.

(a) Participant view in the virtual world.

(b) Real view recorded by a camera
(at the same time).



(c) Partition used to hide the driver.

Fig. 7. VRAV (on-road with VR) setting.

Fig. 8. Virtual hands and body avatar.

3.3.4. *Lab simulator*

The laboratory driving simulator *Australia DS* (Fig. 2), manufactured by Innosimulation (Seoul, South Korea), comprises a driver seat, dashboard with steering wheel, brake and accelerator pedals, and a gear-shifter mounted on a fixed platform. To enhance realism a vibration device was mounted under the driver seat. A built-in desktop computer with an NVIDIA graphics card (GeForce GTX 680) is

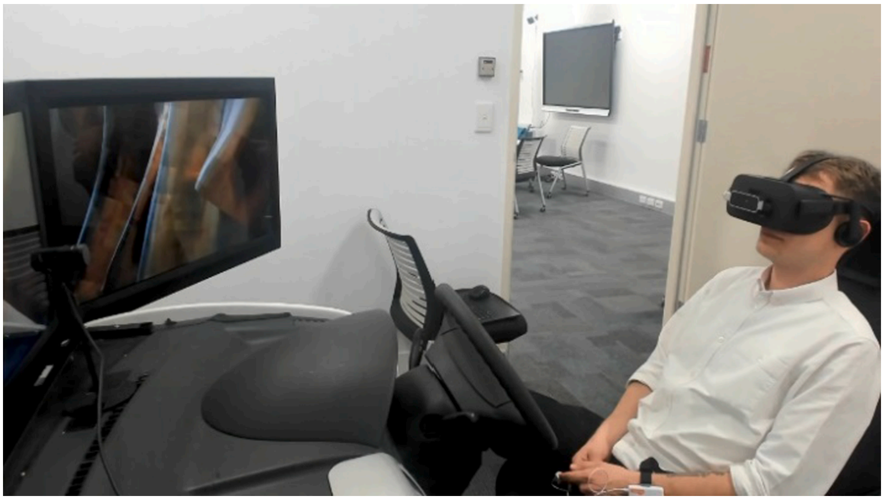


Fig. 9. The “Lab with VR” setting.

used to run the (Unity) simulation. The visual display comprises three 32-inch monitors positioned to provide a 135-degree forward view (used in *Simulator* condition, Fig. 1), while the VR headset provides a 110-degree field of view that also enables participants to look up and down (used in *Seated VR* condition, Figs. 1 and 9). This made it far easier for them to look at the mirrors and take in the full interior of the car. The scene within the headset is not warped, while the three flat screens have obvious “seams” and presented an overall image that is slightly distorted (see Fig. 2).

4. VRAV evaluation

The objective of the VRAV evaluation was to conduct an empirical study of user experience with an on-road VRAV simulator. By comparing user experience across four simulator conditions, this study highlights the differences between on-road VR simulations and laboratory and Wizard-of-Oz AV simulations. The aim is to explore whether the VRAV simulator is an effective alternative to traditional lab-based simulators, and whether it is able to replicate a real AV for simulating autonomous driving in a high-fidelity environment.

Table 1
Summary of demographic information.

	Frequency	Percentage
Gender		
Male	15	53.57%
Female	12	42.86%
Non-binary	1	3.57%
How frequently do you drive a motor vehicle?		
Daily	10	35.71%
4–6 times a week	3	10.71%
2–3 times a week	2	7.14%
Once a week	5	17.86%
Less than once a week	8	28.57%
Please estimate the number of kilometres you have driven in a vehicle over the past year?		
More than 25,001 km	1	3.57%
15,001–20,000 km	4	14.29%
10,001–15,000 km	4	14.29%
5,001–10,000 km	6	21.43%
1,001–5,000 km	4	14.29%
<1,000 km	9	32.14%
What is your previous experience with virtual reality?		
Have used a lot	2	7.14%
Have used several times	4	14.29%
Have used a few times	4	14.29%
Have tried once or twice	10	35.71%
Have never used	8	28.57%

4.1. Implementation

The study utilized a within-subject design to investigate user experiences in the four simulator conditions, and control for variances among the participants. To control for carryover effects, they undertook the study over two sessions, one week apart, completing the two on-road and the two laboratory conditions during separate visits. For the sake of counterbalancing, half of the participants completed the on-road session first, the remaining half the laboratory session. Within each of these groups, the participants were split in half and assigned to either the VR or non-VR condition first.

4.2. Participants

Twenty-eight participants (15 male, 12 female and 1 non-binary) aged 19 to 57 (mean = 28, SD = 8.35) were recruited for the study, and provided informed consent. Table 1 shows that 13 (46.42%) of participants drove a motor vehicle at least four times a week, while another 13 participants only drove once a week or not at all. Most of the participants (67.86%) had travelled<10,000 km over the prior year, which is less than the average distance travelled annually by Victorians (14,100 km) (ABS, 2019). Regarding their previous experience with VR, nearly a third of participants had never used VR, and for the majority of those who had, experience was very limited. The study excluded participants with medical conditions that might be aggravated by use of the simulators, including epilepsy, high blood pressure, and a history of heart attacks, motion sickness, or simulator-induced sickness. Recruitment was undertaken using a combination of methods. We contacted research participants registered in a database, and the study was also advertised to undergraduate and postgraduate students. Participants received a small gift voucher as a compensation for their time and travel expenses. The University Human Research Ethics Committee approved the study.

4.3. Measures

The participants provided information on their demographic characteristics, driving patterns, and experiences using VR (Table 1). In order to compare simulator conditions, they also completed a range of previously validated questionnaires assessing sense of presence, arousal, acceptance, workload, and feelings of sickness within the simulator. Specific details of the questionnaires are as follows.

4.3.1. Presence

A greater sense of presence might evoke more realistic driving performance and behaviours, and thus should increase the validity of results obtained within a virtual environment (Burnett et al., 2017). In this study, presence was assessed using the ITC-Sense of Presence Inventory (ITC-SOPI; Lessiter et al., 2001), which assesses user experiences of immersive environments across four factors: sense of physical space, engagement, ecological validity, and negative effects. The “negative effects” aspect of the ITC-SOPI also indicates the degrees of nausea the simulator induces, making it useful for monitoring the user experience (Burnett et al., 2017). This test was chosen because it was designed for comparison between different media, while other common presence questionnaires are not. With users responding to 44 questions on a five-point Likert scale (strongly disagree to strongly agree), the ITC-SOPI has shown good reliability, with Cronbach alphas ranging from 0.76 to 0.94.

4.3.2. Arousal

Arousal is an important indicator of ecological validity, and the self-assessment manikin (SAM; Bradley and Lang, 1994) has been found effective for evaluating the arousal of a VR system, as well as for verifying it (Liao et al., 2019). In this study, assessments of personal arousal levels also used the SAM, measuring the pleasure, arousal, and dominance levels associated with each simulator condition on a ten-point pictorial scale.

4.3.3. Acceptance

A prerequisite for the introduction of new in-vehicle technology is public acceptance. System acceptance was assessed using the Van Der Laan acceptance scale (Van Der Laan et al., 1997), which assesses acceptance of new technology, and consists of nine 5-point rating-scale items, scored from −2 to + 2, of two factors: usefulness of the system, and the satisfaction it provides (Cronbach alphas 0.91 to 0.94).

4.3.4. Simulator sickness questionnaire

When a participant experiences nausea while using a simulator, this has been shown to make the experience sharply negative, as well as affecting the fidelity of the simulation being perceived, and the validity of the results (Burnett et al., 2017). This makes it essential to assess the degree to which participants feel sick while using the simulator. Participants completed the Simulator Sickness Questionnaire (SSQ) to assess simulator discomfort (Kennedy et al., 1993). The SSQ is a self-reported-symptoms checklist that covers sixteen symptoms associated with simulator discomfort, and asks participants to rate them on a four-point scale (none to severe). Simulator discomfort was measured both in terms of total severity and on three sub-scales: *Nausea*, *Oculomotor* and *Disorientation*.

4.3.5. Task workload

Although task workload is not a measure of validity, this measure was included because it is important to examine whether participants experience a significantly greater task workload when using the VRAV (in other words, to determine whether the VRAV is significantly worse than conventional simulation in this respect). We assessed subjective workload using NASA-TLX (Hart and Staveland, 1988), a workload-rating scale with six dimensions: mental demand, physical demand, temporal demand, effort, performance, and frustration. Participants rated the workload associated with a task for each dimension, on a 100-point scale with 5-point increments; the results are presented as unweighted sub-score scales, with total workload reported as the mean of the six sub-scale scores.

4.3.6. Simulation realism

The degree of realism that a simulator presents to a user is also pertinent to its validity, and important because it not only influences a participant's motivation and perception within the simulator, but also affects wider stakeholder acceptance, such as acceptance of the results of a simulator study (Burnett et al., 2017). To compare the overall realism of the VRAV and traditional screen-based simulators, these were assessed on a 7-point Likert scale (strongly disagree to strongly agree) and scored out of 10.

4.4. Procedure

The participants attended two sessions as part of the experiment. At the first session they provided consent, then were administered a demographic questionnaire following which (in both sessions) participants completed a baseline SSQ. They were then given an explanation of the study tasks, after which they completed the two simulator conditions for the session; each lasted 15–20 min, and was followed by an evaluation questionnaire.

4.5. Analysis

Statistical analyses were performed using IBM SPSS Statistics (version 25) and STATA version 13.1 (StataCorp, 2013). Generalised estimating equations (GEE) with identity links, and an unstructured correlation matrix, were used to test the main effects of and interactions between the independent variables, representing the study conditions, and the dependent variables from the survey instruments. Pairwise comparisons were utilized to assess differences between conditions, with 95th-percentile confidence intervals (CI) reported. A divergent stacked bar chart was used to visualize simulation realism according to the statistics of Likert responses, while average realism scores were calculated and compared.

4.6. Results

Table 2 below provides a summary of the sub-scale scores for the survey instruments.

Table 2
Sub-scale scores [Mean (SD)]

Survey instrument	Sub-scale*	On road		Lab	
		with VR	no VR	with VR	no VR
		(VRAV)	(Wizard-of-Oz)	(Seated VR)	(Simulator)
ITC-SOPI	Engagement (5)	3.76 (0.69)	3.51 (0.56)	3.79 (0.69)	3.17 (0.99)
	Sense of Physical Space (5)	3.63 (0.80)	3.90 (0.75)	3.49 (0.76)	2.76 (1.07)
	Ecological Validity (5)	4.01 (0.98)	4.55 (0.46)	3.85 (0.66)	3.35 (1.12)
	Negative Effects (5)	2.00 (0.84)	1.75 (0.54)	2.06 (0.80)	1.65 (0.65)
SAM	Valence (10)	7.42 (1.93)	7.42 (1.50)	7.17 (1.36)	6.92 (2.14)
	Arousal (10)	4.32 (2.74)	3.60 (2.67)	4.71 (2.43)	3.03 (2.16)
	Dominance (10)	4.39 (2.87)	4.60 (2.62)	4.35 (2.62)	3.92 (2.34)
Acceptance sale	Usefulness (2)	1.08 (0.60)	1.02 (0.63)	1.10 (0.56)	0.79 (0.73)
	Satisfying (2)	1.31 (0.76)	1.27 (0.68)	1.25 (0.62)	1.04 (0.76)
SSQ	Nausea (66.78)	6.47 (10.0)	3.40 (5.92)	10.5 (23.7)	4.42 (9.53)
	Oculomotor (53.06)	9.74 (15.5)	4.60 (7.53)	12.7 (20.0)	5.41 (9.42)
	Disorientation (97.44)	8.94 (17.4)	1.98 (6.24)	15.4 (39.6)	5.46 (15.7)
	Overall (78.54)	9.75 (15.4)	4.14 (6.92)	14.5 (29.1)	5.87 (11.8)
NASA-TLX	Mental Demand (100)	21.0 (23.7)	15.8 (21.7)	24.6 (29.5)	17.6 (23.5)
	Physical Demand (100)	13.5 (17.8)	12.8 (18.9)	14.8 (20.1)	12.6 (21.1)
	Temporal Demand (100)	18.2 (21.7)	18.2 (23.6)	20.8 (24.1)	18.5 (23.1)
	Performance (100)	8.04 (12.4)	4.82 (8.97)	7.68 (13.2)	6.43 (8.37)
	Effort (100)	18.3 (21.9)	11.6 (17.1)	18.0 (23.8)	25.0 (33.6)
	Frustration (100)	16.4 (22.5)	10.0 (15.7)	13.5 (22.2)	11.2 (19.7)
	Overall Workload (100)	16.0 (16.6)	12.2 (15.0)	16.6 (17.1)	15.3 (16.9)

* Maximum values for each scale are shown in parentheses.

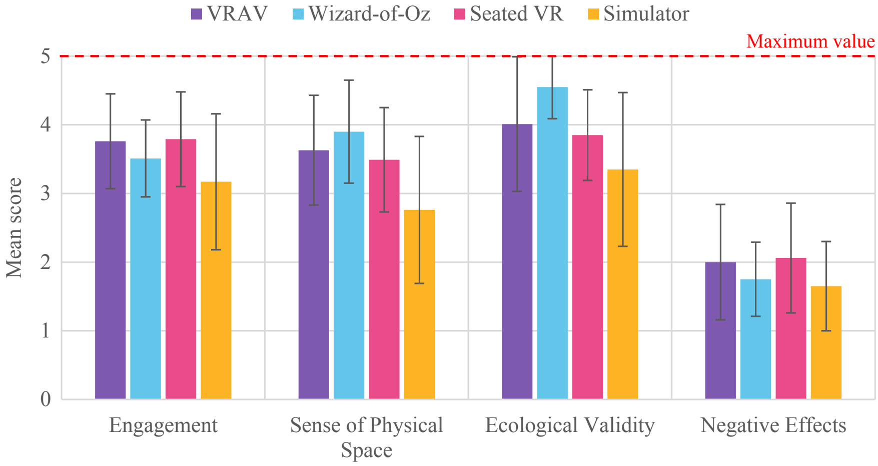


Fig. 10. Sub-scale scores for ITC-SOPI. **Note:** error bars represent standard deviations.

4.6.1. Presence

The sense of presence of participants when using the simulators was assessed on four sub-scales: engagement, sense of physical space, ecological validity, and negative effects (see Fig. 10). The two VR-simulator conditions (VRAV and Seated VR) had the highest mean engagement scores, indicating high levels of engagement ($p < 0.001$).

Not surprisingly, the Wizard-of-Oz condition had the highest mean score for sense of physical space, followed VRAV. A significant main effect identified was that the sense of physical space was higher for the on-road than for the laboratory conditions ($p < 0.001$). The laboratory simulator without VR gave the lowest sense of physical space. Similarly, ecological validity was significantly higher for the on-road conditions ($p < 0.001$) than for the laboratory simulators, with the laboratory condition without VR again receiving the lowest score. The highest level of negative effects was experienced with the seated VR simulator, and there was a significant main effect for the VR study conditions ($p < 0.001$), albeit with the average score for negative effects being generally low for all conditions.

4.6.2. Arousal

Subjective measurement of arousal was undertaken using the SAM (Fig. 11). Across the four conditions, participants reported high levels of valence for all simulator conditions, indicating that they were generally pleased and enjoyed them. The on-road conditions received higher mean scores, but the differences were not significant ($p = 0.513$). Across the four conditions, participants generally experienced low levels of arousal; however, the self-reported levels were higher for the VR-based conditions ($p < 0.001$), indicating that the VR increased excitement and engagement. There were no significant main effects for the laboratory and on-road simulator conditions. Participant scores for the dominance sub-scale tended toward the mean for all conditions, with no significant differences identified among them, indicating that feelings of being in control of the experiment and controlled by it were more or less in balance.

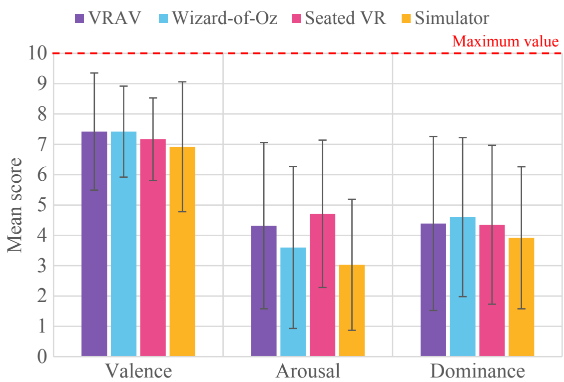


Fig. 11. Sub-scale scores for SAM.

4.6.3. Acceptance

Using the Van Der Laan acceptance scale, the usefulness of and satisfaction provided by the four simulator conditions were assessed (Fig. 12). Overall, the participants found all the simulator conditions useful, with mean positive scores, and the traditional screen-based laboratory simulator rated least useful. Participants reported significantly higher levels of usefulness for the VR conditions ($p = 0.029$). All the simulator conditions also received positive mean scores for satisfaction, with the VRAV simulator condition receiving the highest mean score.

4.6.4. Simulator sickness

Low levels of simulator sickness were recorded on the three sub-scales and overall across the four simulator conditions (Fig. 13). The Seated VR condition consistently induced the highest levels, followed by the VRAV. The VR was found to have a significant main effect in relation to Nausea ($p = 0.018$), Oculomotor ($p < 0.0010$), Disorientation ($p = 0.012$), and overall simulator sickness ($p = 0.003$). However, these results were typically low, and no participants experienced levels of discomfort great enough to adversely impact participation in the study.

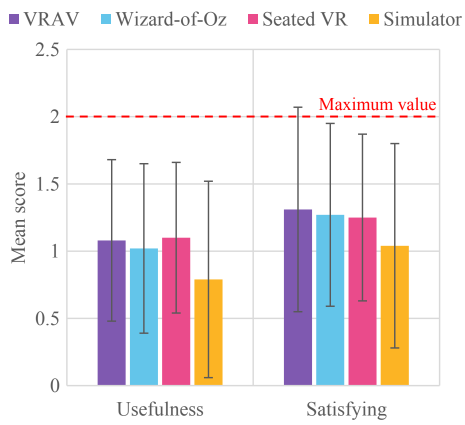


Fig. 12. Sub-scale scores for acceptance scale.

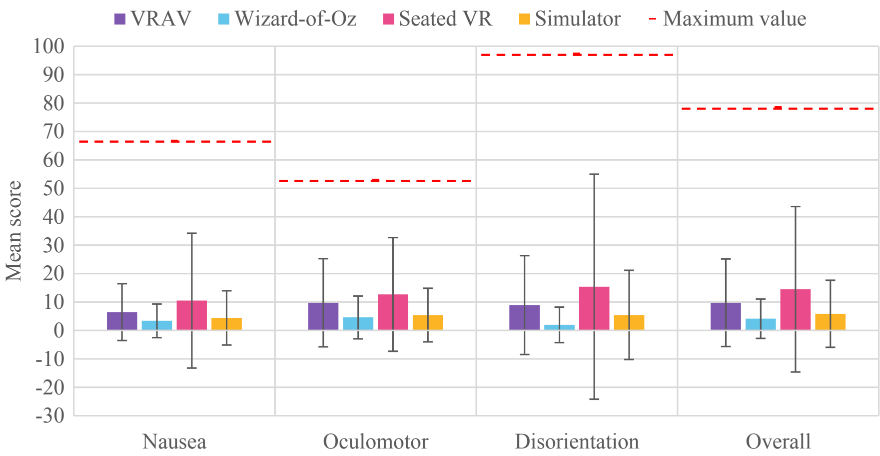


Fig. 13. Sub-scale scores for SSQ.

4.6.5. Task workload

The NASA-TLX was utilized to assess workload (Fig. 14). Participants reported the highest levels of mental demand to be associated with the VR conditions, with a significant main effect identified ($p = 0.01$). Of the four conditions, the highest mental demand was found to be made by the seated VR condition, indicating that there are benefits to using the VRAV over a traditional fixed-base VR simulator. Physical demand was very low of all conditions, and there were no significant differences, which was expected, since the simulator conditions did not require any physical actions to be taken by the participants. The self-reported temporal demand indicated that participants did not feel time pressure in the simulations, with low scores for each, and no significant differences among them.

Across the four conditions, self-reported performance was highest for the VRAV condition, while the differences in mean scores were not significant, and these scores were all very low, which indicates that the participants found the conditions somewhat difficult. Given the low levels of performance reported by participants, it is not surprising that levels of self-reported effort were low for all four conditions, with the lowest reported for the Wizard-of-Oz condition. However, again, the differences in effort were not statistically significant, and indicate that the participants found the four study conditions relatively easy. Interestingly, the VRAV condition received the highest scores for frustration. This indicates that participants found this condition the least relaxing, and indicates the need for improvements to the fidelity of the VRAV simulator. This may also reflect the need for improvements in VR technology, since both the VR conditions scored significantly higher on the frustration sub-scale ($p = 0.047$). Nonetheless, the mean scores for frustration were all towards the lower end of the scale.

In terms of overall workload, the highest scores were for the laboratory VR simulator, and there was a significant main effect for the VR simulators ($p = 0.032$), indicating that participants experienced a higher overall workload when using the VR. This suggests that VR systems do place an added demand on participants compared to the screen-based and Wizard-of-Oz conditions. Nonetheless, overall workload was low for all conditions, indicating that VRAV is a good alternative to conventional simulators.

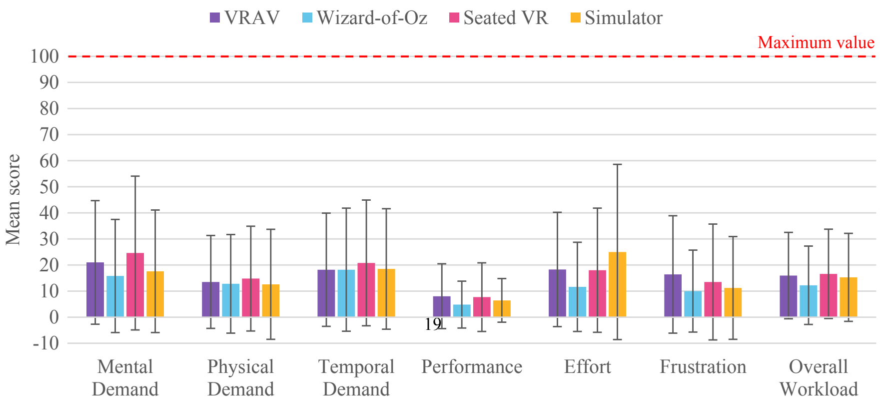


Fig. 14. Sub-scale scores for NASA-TLX.

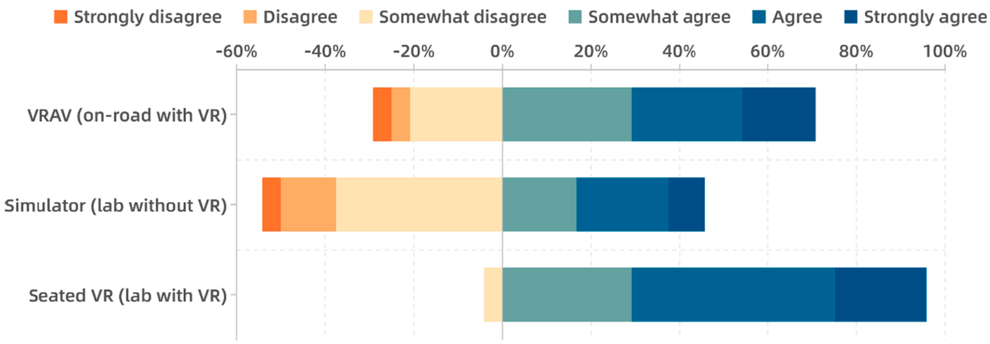


Fig. 15. Participants' responses to "the realism of the simulation condition is almost the same as the real world". Note: As the *Wizard-of-Oz (on-road without VR)* condition is not virtual, it was not included in the figure.

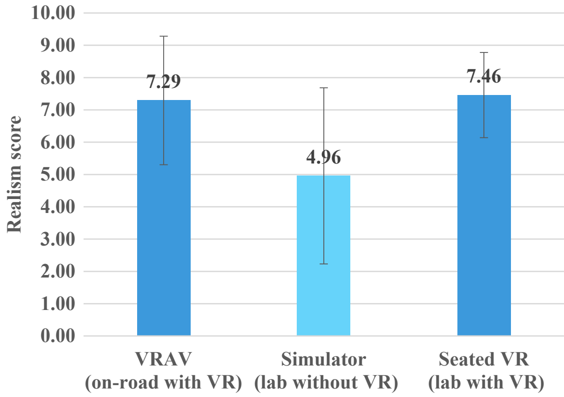


Fig. 16. Average simulation realism scores.

4.6.6. Simulation realism

Because this question (about the realism of the simulation condition when compared to the real world) is focused on the perceived realism of VR displays versus traditional simulator technology, it is not applicable to the real-world Wizard-of-Oz condition. Please note that four participants did not complete this question, so the total number of responses is 24. It can be seen from Fig. 15 that about half (46%) of the participants agreed that the realism of the traditional screen-based laboratory simulator was almost at the same level of the real world with its realism score of 4.96/10, as seen in Fig. 16. As for the VRAV simulator, 71% of the participants agreed (29% somewhat agreed, 25% agreed and 17% strongly agreed) that the simulator was as real as the real world, with a realism score of 7.29/10. Although 96% of the participants agreed that the realism of the Seated VR was nearly at real-world levels, its average realism score (7.46/10) is very near to that of the VRAV, indicating that VR capabilities for these conditions are similar, while both perform significantly better than screen-based simulators. However, the fact that VRAV realism score fall slightly below Seated VR is an indication that further work is needed.

In addition, for the three groups of speed humps (Figs. 5 and 6) distributed along the route, following on-road experiments with the VR, participants reported seeing the humps in the virtual world and feeling the rocking sensation of passing over them in the real world. This indicates that the simulation has realistic graphics and an accurate representation of the physical world. The results are improvements compared with the VR-OOM paper (Goedicke et al., 2018), with participants reporting a disconnect between the virtual and real worlds when the car drove over bumps in their driving area, as these were not simulated in the virtual environment. In assessing the experience of using the lab driving simulator, participants reported that they had not felt the effects of either hitting bumps or of accelerating, and therefore did not feel as if they were riding in an AV.

5. Conclusion and future work

In this paper, we present the development and evaluation of a VRAV simulation system incorporating both the Wizard-of-Oz and VR techniques for on-road simulated autonomous driving studies on commercial vehicles. To evaluate the VRAV simulator based on the driving experiences of participants, surveys were conducted afterward which indicated that they experienced strong feelings of presence and engagement, as well as pleasure and enjoyment. They also found the simulation useful and satisfying, and experienced low levels of simulator-induced sickness or discomfort. Compared to the traditional screen-based driving simulator, VRAV offers a more realistic and immersive experience, and even though the participants experienced a higher overall workload using it, this was still at a low level. All in all, the proposed VRAV simulator has proven to be an effective alternative to conventional simulators, while providing a relatively safe, affordable and low-effort solution for HAI design and research compared with on-road testing using real AVs.

VRAV is intended to be a flexible and economical set up that enables a wide variety of on-road testing to support assessment of human-AV interactions. Potential applications of this system are as follows.

- VRAV can be used to test and evaluate the graphical user interfaces (GUI) of both AV in-car dashboard and head-up displays (e.g. for vehicle control and monitoring) and audio alerts (e.g. for travel-information reminders and system failure/takeover/handover alerts or requests). Above all, the VRAV system allows sounds to be assessed under the influence of other environmental noise in real-world driving contexts, enabling a good understanding of the effects of overlapping voices, and making for enhanced sound design.
- AVs are expected to free drivers to study, work or engage in leisure activities with better comfort than they now enjoy. VRAV can be used to study how people may use their time to conduct in-vehicle activities, and it supports the testing and evaluation of new forms of interior design such as office interiors with conferencing equipment and displays, or leisure interiors with recreational equipment, as well as to investigate how the value people place on their time changes with the purposes or times (morning, afternoon, and so on) of their trips.

- The VRAV simulator is also a promising tool for a variety of driving safety related studies. In future work we would like pursue some of the potential applications that emerged from our initial workshop, such as studying the effect of various traffic environmental or vehicular factors on driver perception and behaviour (e.g., driver situational awareness, re-engagement), testing critical and dangerous scenarios (e.g., a pedestrian who suddenly steps in the road, a car making a sudden stop in front of the test vehicle), developing in-vehicle acquisition systems for driver take-over performance recording, and others. However, further appropriate validation, especially through dynamic interactive experiences, is essential for determining the suitability of the simulator for each of these applications.

Inevitably, the study has several important limitations as noted here.

- One practical limitation was not being able to mount our steering wheel and pedals, meaning that we were not picking up on when the user wanted to take control of the vehicle or additional changes in behaviour that may result with the inclusion of these elements. The location selected for the study was a short and safe section of road, which may have biased the results. However, this was only the first step, and we plan to develop more complicated scenarios for further exploration.
- A limitation in our study design was the use of a convenience sample, which meant that there was self-selection bias. With an age and gender bias, it is not possible to generalize findings for the wider population. We controlled for some of this by performing a within-group study design and counterbalancing the study order and that this was a pilot study and that further research is needed. Another limitation was that participants were not presented the four scenarios in a randomised order, so that there are likely to be order effects in the data.
- According to participant feedback, one important limitation negatively affecting VRAV realism is the fact that sounds outside of the VR headset were not synchronously displayed in the virtual simulation (e.g., the sound of an approaching vehicle, turn indicator or rain). Creative solutions to this limitation would be of great value for exploring ways to correlate actual objects and events with the VR in real time.
- Although the VRAV is a valuable intermediary tool that offers benefits of both conventional driving simulator and naturalistic driving studies, it is not suitable for all experiments. The hybrid naturalistic and simulation-based environment offered by the VRAV needs to be considered prior to undertaking a research study and selection of experimental techniques should consider the aim of the study, research questions and required data outputs.

One potential improvement, which is present in the RRADS paper (Baltodano et al., 2015) in the form of a tablet device on the steering wheel, is some form of interface between our "autonomous car" and the user. The ability for the user to press a "go" button and have some sort of visual feedback as to the car's route would help better sell the illusion of an AV. Additionally, a more robust tracking system would allow for more complicated actions to be tested, such as turns and parking. However, limitations exist in terms of tracking in a moving vehicle. One possibility could be developing an integrated VR headset with a car, which would be a whole new research process in itself.

CRediT authorship contribution statement

Xin Zou: Investigation, Formal analysis, Writing - original draft, Writing - review & editing. **Steve O'Hern:** Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Barrett Ens:** Methodology, Writing - original draft, Writing - review & editing. **Selby Coxon:** Methodology, Writing - review & editing. **Pascal Mater:** Software, Investigation. **Raymond Chow:** Software, Investigation, Writing - review & editing. **Michael Neylan:** Software, Investigation, Writing - review & editing. **Hai L. Vu:** Conceptualization, Methodology, Project administration, Supervision, Writing - review & editing.

Acknowledgement

This research, which won the 2019 Intelligent Transport Systems (ITS) National Research Award, is supported by the Monash Infrastructure (MI) seed funding grant involving researchers from the Monash Institute of Transport Studies, in collaboration with the Faculties of Engineering, Information Technology, Monash University Accident Research Centre (MUARC) and Monash Art Design & Architecture (MADA).

Appendix A. VRAV discovery workshop

Researchers have numerous well-established methods for recording and collecting qualitative data during the early stages of project development. For this study, the strategy was to bring together key experts from both the commercial and academic sectors to take part in a dialogue framed around a workshop model. The purpose was to share our current work examining autonomous vehicles with key influencers and solicit feedback from participants to help focus our direction of the VRAV research. Discussions centred around the role that virtual reality could play in augmenting an authentic passenger experience of riding in an autonomous car, as well as what sort of specific scenarios should be simulated.

The workshop was split into two parts. The first required participants to populate a Goals Grid, which is an established method intended to clarify aims (Nickols, 2015). It provides a visual structure for collecting and categorising objectives, and for detecting conflict between different people's objectives. The premise is that statements are written on post-it notes and placed in one of the four quadrants that make up the grid (Fig. A1). The position of each statement captures what is aspired to and what is not, as well as what is known and what is not.

The participants' use of the Goals Grid revealed common aspirations, such as making the experience as visually realistic as possible while seeking to eliminate negative experiences such as motion sickness.

The second activity focused upon modelling scenarios in which the augmented reality experience could be assessed. This was done through a process of "object storming". Similar to brainstorming, this method used scale models and props to aid participants in expressing ways in which VRAV scenarios could unfold. Cardboard cut-outs and road plans enable communication of ideas and insights at a level that media like interview transcripts may not be able to.

The "object storming" enabled participants to create scenes in which they populated streets with the infrastructure and expected impediments to an AV's progress. Examples include pedestrian crossings, blind corners with buildings, busy intersections, cyclists and pedestrians (Fig. A2).

Following the workshop, outcomes for each of the modelled scenarios were mapped into a C-Box chart where level of idea innovation is mapped against level of implementation difficulty (Fig. A3). This process allowed the research team to sort the relative potential challenges with each scenario against the merits of its potential contribution to VRAV experiment. In the collation of qualitative workshop data for a pilot study like this one, it became clear that the recreation of multiple "events" (e.g., pedestrian impediments and complex intersections) would not be realistically actionable within the given research timeframe. As useful as the mapping of future complex scenarios would be, a proof of concept through the recreation of an uneventful and unimpeded on-road driving scenario would respond to the key aspirations of VRAV proof of concept.

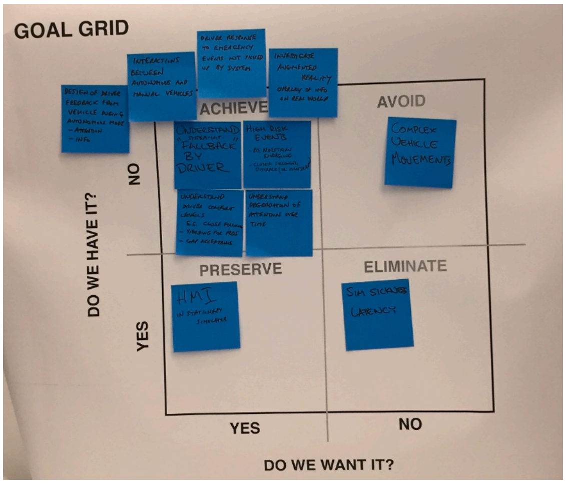


Fig. A1. The Goals Grid.

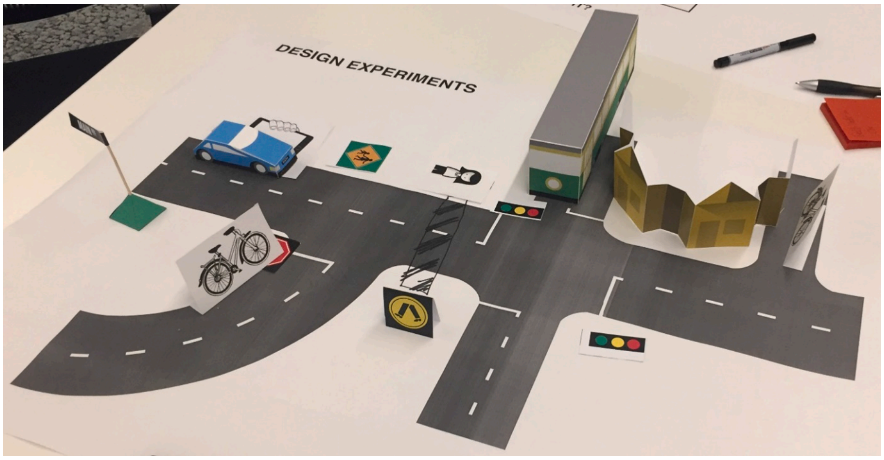


Fig. A2. Scenario design through "object storming".

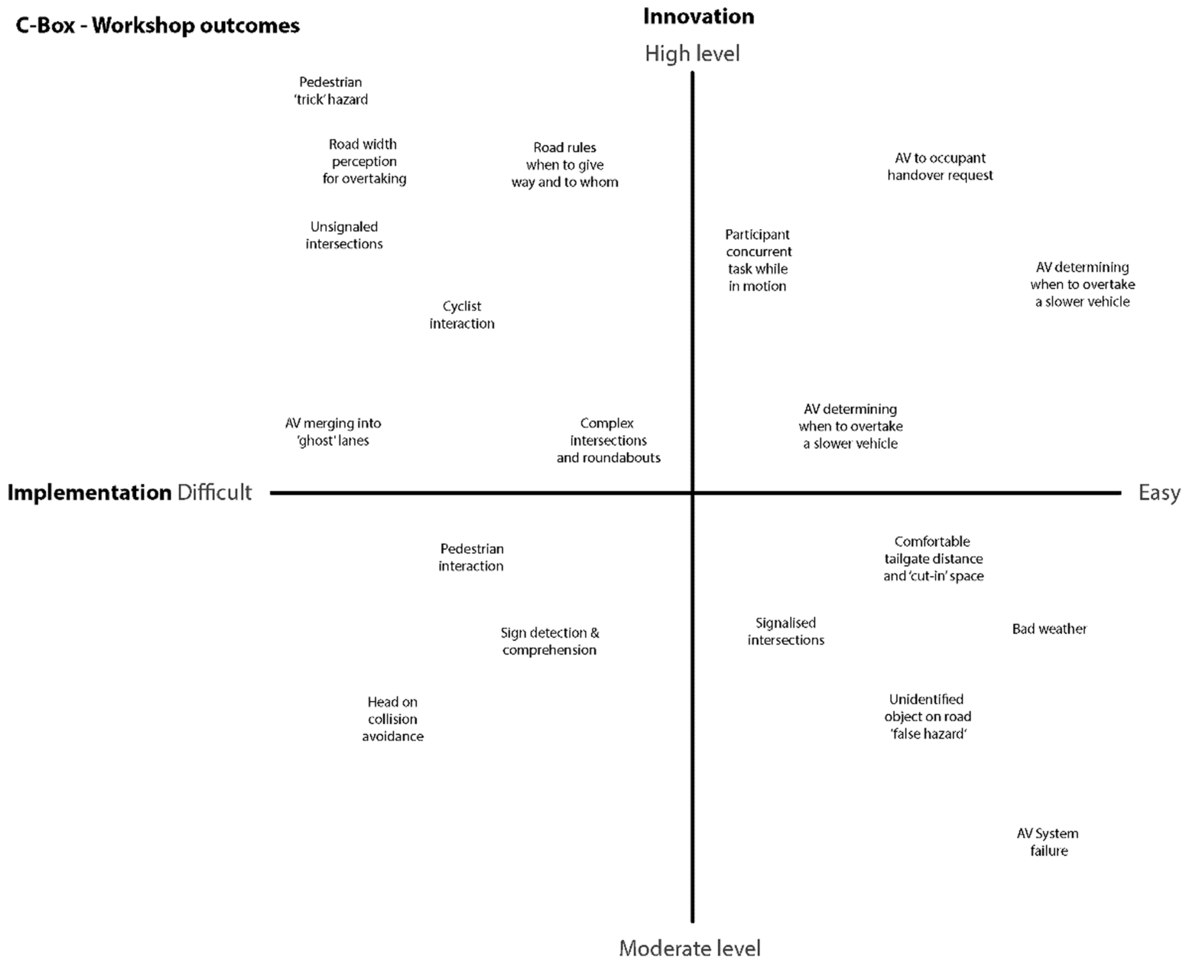


Fig. A3. The C-Box – workshop outcomes.

References

Agrawal, R., Knodler, M., Fisher, D.L., Samuel, S., 2018. Virtual Reality Headset Training: Can It Be Used to Improve Young Drivers' Latent Hazard Anticipation and Mitigation Skills. Transp. Res. Rec. 2672 (33), 20–30. https://doi.org/10.1177/0361198118758311.

Australian Bureau of Statistics (ABS). (2019). Survey of Motor Vehicle Use, Australia, 12 months ended 30 June 2018. https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/9208.0Main+Features112%20months%20ended%2030%20June%202018?OpenDocument.

Baltodano, S., Sibi, S., Martelaro, N., Gowda, N., & Ju, W. (2015). The RRADS platform: a real road autonomous driving simulator. Paper presented at the Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Nottingham, United Kingdom. https://doi.org/10.1145/2799250.2799288.

Banks, V.A., Stanton, N.A., 2016. Keep the driver in control: Automating automobiles of the future. Appl. Ergon. 53, 389–395. https://doi.org/10.1016/j.apergo.2015.06.020.

Blaauw, G.J., 1982. Driving Experience and Task Demands in Simulator and Instrumented Car: A Validation Study. Hum. Factors 24 (4), 473–486. https://doi.org/10.1177/001872088202400408.

Bradley, M.M., Lang, P.J., 1994. Measuring emotion: The self-assessment manikin and the semantic differential. J. Behav. Ther. Exp. Psychiatry 25 (1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9.

Burnett, G., Harvey, C., & Donkor, R. (2017). Driving Simulators for Research. In M. S. Young & M. G. Lenné (Eds.), Simulators for Transportation Human Factors: Research and Practice. Boca Raton, FL: CRC Press. https://www.routledge.com/Simulators-for-Transportation-Human-Factors-Research-and-Practice/Young-Lenne/p/book/9780367879204.

Du, N., Haspiel, J., Zhang, Q., Tilbury, D., Pradhan, A.K., Yang, X.J., Robert, L.P., 2019. Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload. Transport. Res. Part C: Emerg. Technolog. 104, 428–442. https://doi.org/10.1016/j.trc.2019.05.025.

Eriksson, A., Banks, V.A., Stanton, N.A., 2017. Transition to manual: Comparing simulator with on-road control transitions. Accid. Anal. Prev. 102, 227–234. https://doi.org/10.1016/j.aap.2017.03.011.

Godley, S.T., Triggs, T.J., Fildes, B.N., 2002. Driving simulator validation for speed research. Accid. Anal. Prev. 34 (5), 589–600. https://doi.org/10.1016/S0001-4575(01)00056-2.

Goedicke, D., Li, J., Evers, V., Ju, W., 2018. VR-OOM: Virtual Reality On-rOad driving siMulation. Paper presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada. https://doi.org/10.1145/3173574.3173739.

Greenlee, E.T., DeLucia, P.R., Newton, D.C., 2018. Driver Vigilance in Automated Vehicles: Hazard Detection Failures Are a Matter of Time. Hum. Factors 60 (4), 465–476. https://doi.org/10.1177/0018720818761711.

Habibovic, A., Andersson, J., Nilsson, M., Lundgren, V.M., Nilsson, J., 2016. Evaluating interactions with non-existing automated vehicles: three Wizard of Oz approaches. In: Paper presented at the 2016 IEEE Intelligent Vehicles Symposium (IV). https://doi.org/10.1109/IVS.2016.7535360.

Haboucha, C.J., Ishaq, R., Shiftan, Y., 2017. User preferences regarding autonomous vehicles. Transport. Res. Part C: Emerg. Technolog. 78, 37–49. https://doi.org/10.1016/j.trc.2017.01.010.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P.A., Meshkati, N. (Eds.), Advances in Psychology: North-Holland, Vol. 52, pp. 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.

Hussain, Q., Alhajyaseen, W.K.M., Pirdavani, A., Reinolsmann, N., Brijs, K., Brijs, T., 2019. Speed perception and actual speed in a driving simulator and real-world: A validation study. Transport. Res. Part F: Traffic Psychol. Behav. 62, 637–650. https://doi.org/10.1016/j.trf.2019.02.019.

Jamson, A.H., Lai, F.C.H., Carsten, O.M.J., 2008. Potential benefits of an adaptive forward collision warning system. Transport. Res. Part C: Emerg. Technolog. 16 (4), 471–484. https://doi.org/10.1016/j.trc.2007.09.003.

Janssen, C.P., Donker, S.F., Brumby, D.P., Kun, A.L., 2019. History and future of human-automation interaction. Int. J. Hum Comput Stud. 131, 99–107. https://doi.org/10.1016/j.ijhcs.2019.05.006.

Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G., 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. Int. J. Aviation Psychol. 3 (3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3.

Khastgir, S., Birrell, S., Dhadyalla, G., Jennings, P., 2018. Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles. Transport. Res. Part C: Emerg. Technolog. 96, 290–303. https://doi.org/10.1016/j.trc.2018.07.001.

Kim, S., Chang, J.J.E., Park, H.H., Song, S.U., Cha, C.B., Kim, J.W., Kang, N., 2020. Autonomous Taxi Service Design and User Experience. Int. J. Hum.-Comput. Interaction 36 (5), 429–448. https://doi.org/10.1080/10447318.2019.1653556.

Knapper, A., Christoph, M., Hagenzieker, M., Brookhuis, K., 2015. Comparing a driving simulator to the real road regarding distracted driving speed. Eur. J. Transp. Infrastruct. Res. 15 (2).

Lessiter, J., Freeman, J., Keogh, E., Davidoff, J., 2001. A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. Presence: Teleoperators Virtual Environ. 10 (3), 282–297. https://doi.org/10.1162/105474601300343612.

Li, S., Blythe, P., Guo, W., Namdeo, A., 2019. Investigation of older drivers' requirements of the human-machine interaction in highly automated vehicles. Transport. Res. Part F: Traffic Psychol. Behav. 62, 546–563. https://doi.org/10.1016/j.trf.2019.02.009.

Liao, D., Shu, L., Liang, G., Li, Y., Zhang, Y., Zhang, W., Xu, X., 2019. Design and Evaluation of Affective Virtual Reality System Based on Multimodal Physiological Signals and Self-Assessment Manikin. IEEE J. Electromagnet., RF Microwaves Med. Biol. 1–1 https://doi.org/10.1109/JERM.2019.2948767.

Llopis-Castelló, D., Camacho-Torregrosa, F.J., Marín-Morales, J., Pérez-Zuriaga, A.M., García, A., Dols, J.F., 2016. Validation of a Low-Cost Driving Simulator Based on Continuous Speed Profiles. Transp. Res. Rec. 2602 (1), 104–114. https://doi.org/10.3141/2602-13.

Mullen, N., Charlton, J., Devlin, A., Bédard, M., 2011. Simulator Validity: Behaviors Observed on the Simulator and on the Road. In: Fisher, D.L., Rizzo, M., Caird, J.K., Lee, J.D. (Eds.), Handbook of Driving Simulation for Engineering, Medicine, and Psychology. Boca Raton, FL: CRC Press. https://www.routledge.com/Handbook-of-Driving-Simulation-for-Engineering-Medicine-and-Psychology/Fisher-Rizzo-Caird-Lee/p/book/9781138074583.

Nebeling, M., Madier, K., 2019. 360proto: Making Interactive Virtual Reality & Augmented Reality Prototypes from Paper. Paper presented at the Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland UK. https://doi.org/10.1145/3290605.3300826.

Nickols, F., 2015. The Goals Grid: A Tool for Clarifying Goals and Objectives. https://nickols.us/goals_grid.pdf.

Pai Mangalore, G., Ebadi, Y., Samuel, S., Knodler, M.A., Fisher, D.L., 2019. The Promise of Virtual Reality Headsets: Can They be Used to Measure Accurately Drivers' Hazard Anticipation Performance? Transp. Res. Rec. 2673 (10), 455–464. https://doi.org/10.1177/0361198119847612.

Panagiotopoulos, I., Dimitrakopoulos, G., 2018. An empirical investigation on consumers' intentions towards autonomous driving. Transport. Res. Part C: Emerg. Technolog. 95, 773–784. https://doi.org/10.1016/j.trc.2018.08.013.

SAE International (2018). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016_201806). https://www.sae.org/standards/content/j3016_201806/.

Shutko, J., Osafo-Yeboah, B., Rockwell, C., Palmer, M., 2018. Driver Behavior While Operating Partially Automated Systems: Tesla Autopilot Case Study. SAE Technical Paper. https://doi.org/10.4271/2018-01-0497.

Sportillo, D., Paljic, A., Ojeda, L., 2018. Get ready for automated driving using Virtual Reality. Accid. Anal. Prev. 118, 102–113. https://doi.org/10.1016/j.aap.2018.06.003.

StataCorp, 2013. *Stata: Release 13*. Statistical Software. StataCorp LP, College Station, TX.

Törnros, J., 1998. Driving behaviour in a real and a simulated road tunnel—a validation study. Accid. Anal. Prev. 30 (4), 497–503. https://doi.org/10.1016/S0001-4575(97)00099-7.

Van Der Laan, J.D., Heino, A., De Waard, D., 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. Transport. Res. Part C: Emerg. Technolog. 5 (1), 1–10. https://doi.org/10.1016/S0968-090X(96)00025-3.

Wang, P., Sibi, S., Mok, B., Ju, W., 2017. Marionette: Enabling On-Road Wizard-of-Oz Autonomous Driving Studies. In: Paper presented at the Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. https://doi.org/10.1145/2909824.3020256.

Wynne, R.A., Beanland, V., Salmon, P.M., 2019. Systematic review of driving simulator validation studies. Saf. Sci. 117, 138–151. https://doi.org/10.1016/j.ssci.2019.04.004.

Zhao, X., Sarasua, W.A., 2018. How to Use Driving Simulators Properly: Impacts of Human Sensory and Perceptual Capabilities on Visual Fidelity. Transport. Res. Part C: Emerg. Technolog. 93, 381–395. https://doi.org/10.1016/j.trc.2018.06.010.