Taylor & Francis
Taylor & Francis Group

🔓 OPEN ACCESS | Check for updates

# Behavioural Gap Assessment of Human-Vehicle Interaction in Real and Virtual Reality-Based Scenarios in Autonomous Driving

Sergio Martín Serrano[a] 📷, Rubén Izquierdo[a] 📷, Iván García Daza[a] 📷, Miguel Ángel Sotelo[a] 📷, and David Fernández-Llorca[a,b] 📷

[a]Computer Engineering Department, Universidad de Alcalá, Alcalá de Henares, Spain; [b]European Commission, Joint Research Centre, Seville, Spain

## ABSTRACT

In the field of autonomous driving research, the use of immersive virtual reality (VR) techniques is widespread to enable a variety of studies under safe and controlled conditions. However, this methodology is only valid and consistent if the conduct of participants in the simulated setting mirrors their actions in an actual environment. In this paper, we present a first and innovative approach to evaluating what we term the *behavioural gap*, a concept that captures the disparity in a participant's conduct when engaging in a VR experiment compared to an equivalent real-world situation. To this end, we developed a digital twin of a pre-existed crosswalk and carried out a field experiment (N = 18) to investigate pedestrian-autonomous vehicle interaction in both real and simulated driving conditions. In the experiment, the pedestrian attempts to cross the road in the presence of different driving styles and an external Human-Machine Interface (eHMI). By combining survey-based and behavioural analysis methodologies, we develop a quantitative approach to empirically assess the *behavioural gap*, as a mechanism to validate data obtained from real subjects interacting in a simulated VR-based environment. Results show that participants are more cautious and curious in VR, affecting their speed and decisions, and that VR interfaces significantly influence their actions.

## 1. Introduction

As autonomous vehicle (AV) technology advances, the need for rapid prototyping and extensive testing is becoming increasingly important, as real driving tests alone are not sufficient to demonstrate safety (Kalra & Paddock, 2016; Llorca & Gómez, 2021). The use of physics-based simulations allows the study of various scenarios and conditions at a fraction of the cost and risk of physical prototype testing, providing valuable insights into the behaviour and performance of AVs in a controlled environment (Schwarz & Wang, 2022).

However, one of the main challenges in the development of autonomous driving digital twins is the lack of realism of simulated sensor data and physical models. The so-called *reality gap* can lead to inaccuracies because the virtual world does not adequately generalise all the variations and complexities of the real world (García Daza et al., 2023; Stocco et al., 2023). Furthermore, despite attempts to generate realistic synthetic behaviours of other road agents (e.g., vehicles, pedestrians, cyclists), simulation lacks empirical knowledge about their behaviour, which negatively affects the gap in behaviour and motion prediction, communication, and human-vehicle interaction (Eady, 2019).

Including behaviours and interactions from real agents in simulators is one way to reduce the *reality gap* of autonomous driving digital twins. This can be addressed by using real-time immersive VR (Serrano et al., 2022; 2023). The immersive integration of real subjects into digital twins allows, on the one hand, human-vehicle interaction studies in fully controlled and safe environments. Various HMI modalities can be included to explore extreme scenarios without risk to people and vehicle prototypes. On the other hand, it makes it possible to obtain synthetic sequences from multiple viewpoints (i.e., simulated sensors of AVs) based on the behaviour of real subjects, which can be used to train and test predictive perception models. However, this approach would only be valid if the behaviour of the subjects in the simulated environment is equivalent to their behaviour in a real environment. We refer to this difference in behaviour as the *behavioural gap*, and in order to model it, it is necessary to empirically assess the behaviour of subjects under equivalent real and simulated conditions.

Meanwhile, the attempt to introduce autonomous driving into daily life makes it crucial to study humans-AVs interactions, as the absence of a driver has an impact on the perception of risk, trust (Li et al., 2019) and the level of acceptance by all users (Detjen et al., 2021), including non-driving passengers and external road agents (i.e., pedestrians, cyclists and other drivers) (Llorca & Gómez, 2021). AVs are

faced with the need to communicate their intentions using all available resources, which translates into the use of HMIs as a form of explicit communication. Nonetheless, some previous studies suggest the primary basis for crossing decisions taking by pedestrians is the implicit interaction, such us perceived vehicle speeds or safety gap sizes (Clamann et al., 2017; Zimmermann & Wettach, 2017). Thus, the first interest of our research is to evaluate together an explicit form of communication (eHMI) and an implicit one, as in this case a different braking manoeuvre of the vehicle. In this paper, we present the results of the first part of our field study on human-AVs interactions, in a real-world crosswalk scenario (Izquierdo et al., 2023) and which answers our first research question:

- **RQ1**: To what extent do the variables "eHMI" and "braking manoeuvre" influence the crossing behaviour of a pedestrian in a real-world crosswalk in terms of (1) vehicle-gazing time, (2) space gap, (3) body-motion, and (4) subjective perception?

On the other hand, we employed a novel framework to insert real agents into the CARLA simulator (Serrano et al., 2022; 2023). Through the CARLA tools and the added motion capture system, we enable an immersive VR interface for a pedestrian and reproduce the same interaction conditions with the vehicle (i.e., eHMI and driving style) (Serrano et al., 2023), allowing us to pose our second research question:

- **RQ2**: To what extent do the variables "eHMI" and "braking manoeuvre" influence the crossing behaviour of a pedestrian in a virtual crosswalk in terms of (1) vehicle-gazing time, (2) space gap, (3) body-motion, and (4) subjective perception?

Furthermore, as our interest is focused on providing a pioneering measure of the *behavioural gap* that exists in the activity of a participant depending on whether s/he acts in a physical-real or virtual environment, we developed a digital twin of the exact same crosswalk of the first part of the study, imitating its visibility conditions and road dimensions. The same experiment setup is repeated in a real-world and an identical virtual scenario to answer our last research question:

- **RQ3**: To what extent does pedestrian crossing behaviour differ between a real and a virtual environment in terms of (1) vehicle-gazing time, (2) space gap, (3) body-motion, and (4) subjective perception?

To our knowledge, this is the first approach that is concerned with evaluating whether human behaviour is realistic within a VR setup for autonomous driving.

## 2. Related work

### 2.1. Understanding pedestrian-AVs interaction

The research of the interactions between pedestrians and AVs is essential to ensure the safety and public acceptance of this emerging technology (Dey et al., 2018; Fernandez-Llorca & Gomez, 2023). To date, numerous studies have been conducted to investigate the role of eHMIs and AV driving styles on the pedestrian crossing experience, in controlled real-world environments (Dey et al., 2021; Izquierdo et al., 2023) and in VR environments (Nascimento et al., 2019; Serrano et al., 2023; Stadler et al., 2019).

Among the eHMI forms commonly explored, we can find several lighting signals designs, textual messages, inclusion of anthropomorphic featuring or trajectory projection on the ground (Bazilinskyy et al., 2019; Furuya et al., 2021; Mason et al., 2022). For instance, an AV equipped with robotic eyes that look at the pedestrian or head-on helps them make more efficient crossing choices (Chang et al., 2017; 2022). Various approaches have studied the effect of light-based communication in Wizard-of-Oz designs in which automated driving is simulated that appears to be driverless (Hensch et al., 2020; 2020). Despite the fact that in many cases visual messages can be displayed on an external surface to indicate the status of the vehicle (e.g., real-time predicted risk levels (Song et al., 2023) or directional information (Bazilinskyy et al., 2022)), some studies note that their participants prefer direct written instructions to cross the road (i.e., "walk" or "stop") (Ackermann et al., 2019; Deb et al., 2020). This could be misleading when the traffic situation involves more than one pedestrian (Song et al., 2023) so road projection-based eHMIs may be an alternative for scalability to communicate vehicle intentions in shared spaces (Dey et al., 2021; Mason et al., 2022; Nguyen et al., 2019). Most of the research on eHMI development in virtual reality focuses on visual components, as commercially available hardware and software are at an early stage of development, which poses difficulties in creating multimodal experiences (Le et al., 2020). Auditory, haptic and interactive elements, such as the movement of participants and the virtual representation of their bodies, are mainly used to increase the sense of presence in the virtual environment. However, these elements could also enhance the authenticity of participants' reactions.

In another sense, it has also been shown that pedestrians use implicit communication signals to estimate the behaviour of the vehicle, and apply it to their decisions (Tian et al., 2023). Moreover, leading works suggest that implicit information (i.e., their movement) may be sufficient (Moore et al., 2019) or that eHMIs only help convince pedestrians to cross the road when the vehicle speed is ambiguous (Dey et al., 2021). Deceleration or the distance to the vehicle are more useful in interpreting the intention to yield than the drivers' presence and apparent attentiveness (Velasco et al., 2021). This type of communication has been found to be even more relevant in unmarked locations (Kalantari et al., 2023; Lee et al., 2022).

Although survey-based studies to assess human behaviour in traffic scenes are prevalent (Fridman et al., 2017; Li et al., 2018; Merat et al., 2018), they fail to collect immediate feedback from experiments (Dijksterhuis et al., 2015). Recording-based studies allow direct measurements and help mitigate potential biases associated with self-reporting

(Tom & Granié, 2011). Metrics extracted from objective data can be treated as dependent variables and analysed using a linear mixed models, including road-crossing decision times, gaze-based times, crossing speed or distances to the vehicle (Feng et al., 2023; Guo et al., 2022).

## 2.2. Bridging the Simulation-to-reality gap

Testing in simulated environments offers some advantages over real-world testing, such as more safety for participants in the experiments and the facility of constructing scenarios (Fratini et al., 2023). This saves a lot of costs in terms of time and effort. However, differences in lighting, textures, vehicle dynamics and agents behaviour between simulated and real environments raise doubts about the validity of the results in this new context (Hu et al., 2024).

The first approach to assessing whether simulation-based testing can be a reliable substitute for real-world testing is to validate the virtual models of the sensors by determining whether their discrepancy with reality is sufficiently low. We found works that do this in the case of radar (Ngo et al., 2021) and a camera-based object detection algorithm (Reway et al., 2020). Typically, the gap between synthetic and real-world datasets is well-known (Gadipudi et al., 2022), and there are already proposals to alleviate it as methods that obtain realistic images from those recorded in simulation or that bridge the differences in system dynamics (Cruz & Ruiz-del Solar, 2020; Pareigis & Maaß, 2022). We emphasise that the gap worsens in multi-agent systems due to the complexity of transferring agent interactions and the synchronisation of the environment (Candela et al., 2022).

One of the strategies researchers employ to bridge the gap between simulation and reality in autonomous driving are the digital twins (DTs) (Almeaibed et al., 2021; Ge et al., 2019; Hu et al., 2024; Yu et al., 2022; Yun & Park, 2021). Some study utilises a real small-scale physical vehicle and its digital twin to investigate the transferability of behaviour and failure exposure between virtual and real-world environments (Stocco et al., 2023). There have been no previous approaches to assess the gap in the behaviour of real agents (e.g., pedestrians) within a simulation, as we do in this work with a full-scale digital twin of a scenario and immersive VR for real-time interaction with an AV.

# 3 Method

## 3.1. Experiment design

The currently study presents improvements over previous immersive VR experiments with pedestrians, since *(i)* it is conducted in the CARLA simulator (Dosovitskiy et al., 2017) and not in Unity, which allows the use of highly specialised functions for autonomous driving, and *(ii)* a motion capture system is added to accurately collect the participants motion data. On one hand, we can assess interactions by the usual methods, such as eye contact with the vehicle or questionnaires (Rasouli & Tsotsos, 2020). Furthermore, we combine explicit and implicit communication under safe conditions, and capture the behaviour of the participants by video and inertial sensors.

### 3.1.1. Experiment scenario design

An existing crosswalk on the area of the University of Alcalá (Spain), was chosen to perform the real driving tests and also as the baseline to construct the VR environment (see Figure 1). In this scenario, an AV drives on a day with plenty of sunlight along a street in a straight line until it reaches a crosswalk. The pedestrian, who wishes to cross the road perpendicularly, needs to take 2-3 steps to have visibility to their left side (due to other parked vehicles and vegetation).

The map model is downloaded from OpenStreetMap (Steve Coast, 2024) and converted to a Unreal Engine project where the elements are detailed. From the vehicle blueprints offered by CARLA we choose the model and colour of the physical vehicle and attach the sensors to perceive its surroundings (i.e., LiDAR, radar and cameras).

In order to facilitate interaction, the pedestrian waits with their back to the road and is instructed to turn around when the vehicle is at a distance of about 40 meters. As can be seen in Figure 2, two braking manoeuvres were designed. In both cases, the vehicle travels at a speed of 30 km/h and applies a constant deceleration of $-0.9\,\text{m/s}^2$ (smooth) or $-1.8\,\text{m/s}^2$ (aggressive) until it comes to a complete stop in front of the crosswalk and yields the right-of-way. This is done to study whether the pedestrian perceives the situation as more risky when the vehicle brakes with less anticipation and the time-to-collision (TTC) is smaller.

To alert the pedestrian of its intention to yield, the vehicle was equipped with the GRAIL (Green Assistant Interfacing Light) system (Gonzalo et al., 2022). As shown in Figure 3(a,b), the AV uses green to indicate awareness of the pedestrian (which implies that it will stop if necessary), and red to warn that nothing prevents it from continuing on its way. It is also possible that the interface is deactivated so the pedestrian does not have any explicit information about the vehicle intention. This front-end design is sufficient for the specific scenario of this work. However, more complex scenarios with poorer visibility conditions might require enhancements, such as extending the LED lightband to the sides of the vehicle, or even incorporating a 360-degree eHMI approach (Hub et al., 2023).

### 3.1.2. Experiment task design

The combination of eHMI on or off, and the different strategies of deceleration result in the road-crossing tasks listed in Table 1. When activated, the eHMI starts emitting the red light and changes to green when the vehicle has covered a 30% of the braking distance (12 or 6 meters, depending on the type of manoeuvre). All tasks were performed in a random order specific to each participant, except for task 0 (warm-up task), which always started the experiment and in which the vehicle did not stop and the participant only had to turn towards the road and watch the vehicle without initiating the crossing action.
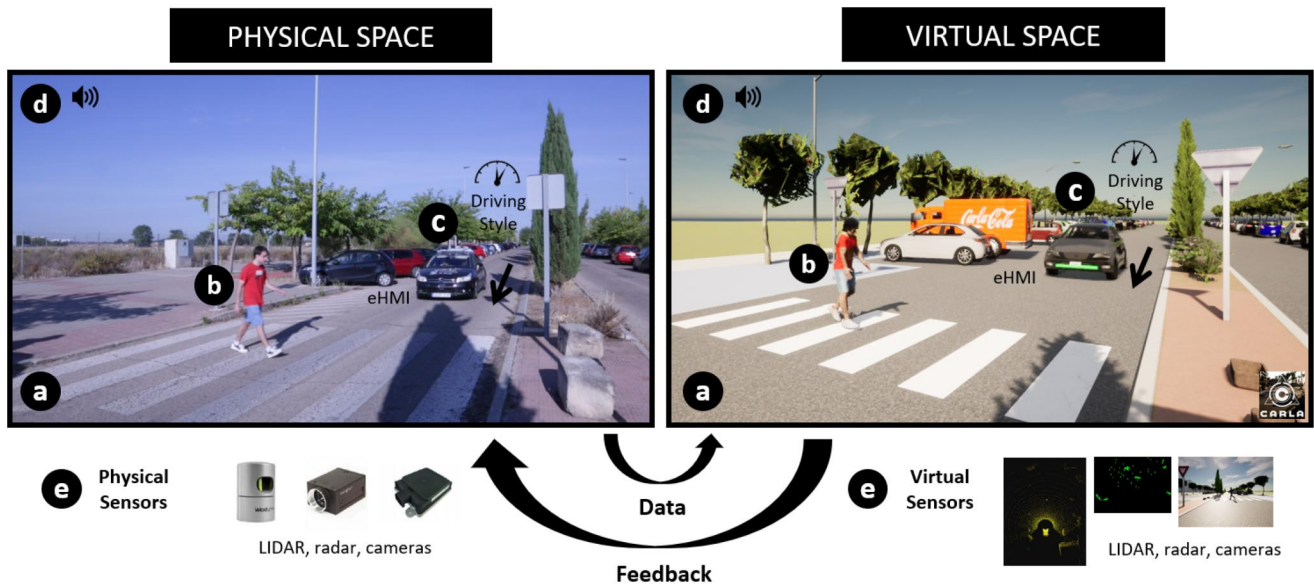
**Figure 1.** Digital twin for human-vehicle interaction in autonomous driving. (a) 3D crosswalk scenario. (b) Pedestrian attempting to cross. (c) Autonomous vehicle (eHMI, driving style). (d) ambient sound, lighting and traffic signals. (e) Physical versus virtual sensors.
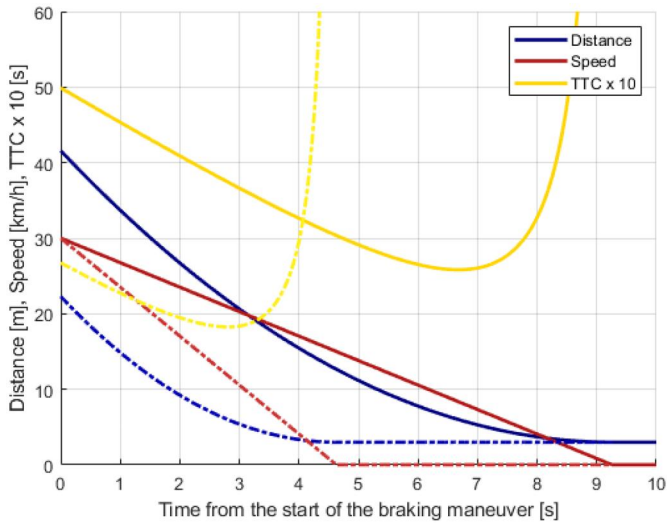


**Figure 2.** Smooth (continuous line) and aggressive (dashed line) deceleration.
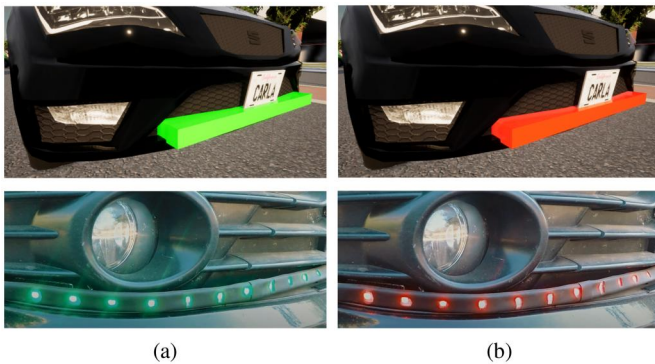


**Figure 3.** AV with eHMI activated communicating green (a) and red (b) status. Virtual (above) and physical (bottom) design.

### 3.2. Virtual reality apparatus

Tests under simulated driving conditions were conducted in a VR space of 8 meters long x 3 meters wide. The virtual

**Table 1.** Experimentation tasks settings.

| Task number | AV strategy deceleration | External HMI | Stop |
|---|---|---|---|
| 0 | – | – | No |
| 1 | Smooth | – | Yes |
| 2 | Aggressive | – | Yes |
| 3 | Smooth | Activated | Yes |
| 4 | Aggressive | Activated | Yes |

environment was constructed under a 1:1 scheme mapped to the real-life environment, so participants adopted the real-walking locomotion style, leading to a more realistic movement and a greater sense of presence.

We use a specific framework for the insertion of real agents in CARLA (Serrano et al., 2022; 2023). An immersive interface is enabled in VR for the incorporation of a pedestrian into the traffic scene. Some of the features added to the simulator were real-time avatar control, positional sound or body tracking. The Meta Quest 2 headset was connected via WiFi to a Windows 10 desktop and an NVIDIA GeForce RTX 3060 graphics card. We chose Perception Neuron Studio (Noitom, 2022) as the motion capture system to record the user's pose and integrate it into the scenario.

### 3.3. Experiment procedure

The experimental procedure differed between the real and virtual contexts, yet it could be distinctly delineated into four phases:

1. *Introduction*: At the beginning, participants were provided with written information about the experiment, such as its purpose, the explanation of the AV and the functionality of the eHMI. They were also assigned a unique anonymous identifier and were assured of their ability to discontinue the experiment at any time if they so desired. Lastly, they were asked to sign the consent to participate as subjects in the study.

2. *Familiarisation (warm-up)*: Participants were aided in donning the inertial sensors and VR headset, following which they were invited to explore the virtual environment void of any vehicular traffic. Subsequently, the Perception Neuron system underwent calibration, and the initial task of the experiment (task 0) was conducted as an illustrative example.
3. *Experimentation*: Throughout this phase, participants completed the four tasks of the experiment (see Table 1) while answering questions posed by an accompanying researcher about their subjective perception.
4. *Filling in the post-questionnaire*: After concluding the experiment, participants removed the VR headset and inertial sensors and, in both the real and virtual context, were asked to fill out a post-questionnaire.

## 3.4. Data collection

During the experiment various types of data were collected to analyse the resulting pedestrian-AV interactions, including objective measurements (i.e., movement path, gaze time) as well as responses to questionnaires.

In the first instance, the AV in real configuration was fitted with a RTK-GPS system that provided its precise position with respect to the crosswalk and served as a reference for applying the braking manoeuvre, while an external camera mounted on the top of the vehicle recorded the environment at 10 Hz. Within Unreal Engine 4 and Axis Studio (Serrano et al., 2022; 2023), all data from the VR experiment were recorded as the (1) timestamp, (2) vehicle's position and parameters (i.e., coordinate x, y, z, rotation, brake, steer, throttle, gear), (3) participant's position and animation (i.e., coordinate x, y, z, rotation, .fbx), and (4) playbacks of the Quest 2 view, the VR setup, and from within the simulator, synchronised at 18.8 Hz.

The questionnaire collected participant's information (e.g., age, gender, familiarity with AVs and with VR) and subjective feedback on the influence of the different types of communication in each interaction through the following questions:

Q1: How safe did you feel at the scene?
Q2: How aggressive did you perceive the braking manoeuvre of the vehicle?
Q3: Did the visual communication interface improve your confidence to cross?

Answers were tabulated on a 7-step Likert scale (Joshi et al., 2015). In addition, the participants completed a 15-item presence scale (depicted in Appendix A) to evaluate the quality of pedestrian immersion in the scene.

## 3.5. Participant's characteristics

A total of 18 participants, aged between 24 and 62 years (M = 40.11, SD = 11.62), with a gender distribution of 33% women and 67% men, were recruited from both inside and outside the university area and engaged in the experiment.

In regard to familiarity with AVs, 38.9% had extensive knowledge of the subject, another 38.9% considered that they had an average knowledge of Advanced Driver Assistance Systems (ADAS), 44.4% had previously interacted with an AV (either as a user or pedestrian) compared to 55.6% who had not, and 22.2% had no prior exposure or understanding of AVs. For VR experience, the majority of participants had either never used VR goggles (50%) or had only tried them once before (33.3%). All participants had normal vision or wore corrective glasses (22.2%) that they kept when fitting the VR headset, and had normal mobility, so they were able to complete the experiment successfully.

## 3.6. Data analysis

Different metrics can be acquired from the objective data (i.e., movement trajectory, gaze point) gathered during the experiments. The metrics chosen for analysis in this research are defined as follows:

- *Vehicle-gazing time while crossing ($T_c$):* it represents the cumulative duration of gazing at the AV while crossing, as inferred from the collected eye gazing data.
- *Crossing initiation time (CIT):* computed as the interval from when the pedestrian visually identifies the AV until s/he decides to cross. If the pedestrian crosses before noticing the AV, then CIT is zero.
- *Vehicle-gazing time ($T_{av}$):* it represents the cumulative duration of gazing at the AV throughout the entire crossing process, as inferred from the collected eye gazing data. That is, $T_{av} = CIT + T_c$.
- *Space gap (L):* the distance between the AV and the pedestrian, measured from the AV to the centre of the crosswalk when the pedestrian decides to cross.
- *Pace cadence ($F_p$):* defined as the dominant step frequency at which the pedestrian crosses the road.
- *Gait cycles (G):* referring to the number of gait cycles when the pedestrian makes the decision to cross, along with the stabilisation times of the two ankles.

To obtain the above indicators, the crossing intention event is defined as the moment the pedestrian decides to cross the crosswalk and is extracted from the video recordings and the reconstructed trajectory in the virtual environment. The rules for identifying the event state are the following:

1. In case the pedestrian is stopped and starts to move into the crosswalk, the decision is made at the first frame in which the movement is discernible.
2. If there is not a stop and the pace is slowed, the decision occurs at the frame the pedestrian starts accelerating.
3. If there is no alteration in the pedestrian's speed, the decision is made upon sighting the vehicle.
4. If the pedestrian does not look at the vehicle, we take the first frame when the pedestrian appears on the vehicle's front camera.
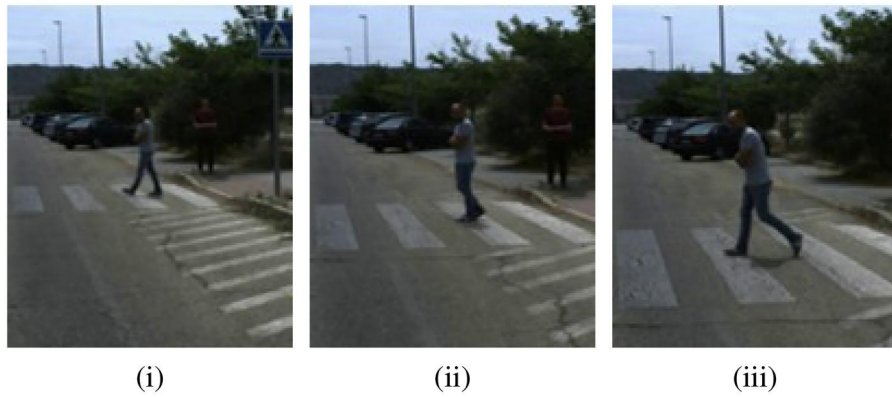
**Figure 4.** Crossing decision event. (i) The pedestrian takes two steps forward to gain visibility. (ii) The vehicle is approaching and the pedestrian slows down without stopping. (iii) The pedestrian makes the decision to cross.

An example of the crossing decision in the real environment can be seen in Figure 4.

Ultimately, to conclusively state that there are differences in crossing decision making in each task of the experiment, we employed the Student's t-test (De Winter, 2013). For the analysis of the questionnaire, we used the Wilcoxon signed-rank test (Woolson, 2007).

## 4. Results

This section presents the results obtained in the experiment with the real and virtual setup. As can be seen in Figure 5, the VR headset projects the crosswalk onto its lenses and allows mobility around the scene. We aim to examine the significant effects of implicit and explicit vehicle communication on pedestrian crossing behaviour.

### 4.1. Vehicle Gazing ($t_{av}$) and crossing initiation times (CIT)

To establish categorical statements about the impact of the braking manoeuvre or eHMI on the crossing decision, we utilise the Student's t-test (De Winter, 2013). The procedure for this test involves calculating the difference between the means of two groups of samples and adjusting this difference for within-group variability and sample size. This adjusted difference is compared to a probability t-distribution to determine if it is large enough to be considered significant. If this happens with the means of the data extracted from the experimental tasks, the null hypothesis ($H_0 : \mu_i \leq \mu_j$) is rejected in favour of the alternative hypothesis ($H_1 : \mu_i > \mu_j$). Figure 6 shows the box-plots of the gaze duration to the vehicle in the tests, considering combinations of two factors: deceleration type and activation of the eHMI (see Table 1 for details).

The Table 2 expresses categorical statements, i.e., a 1 in a cell means rejection of the null hypothesis and acceptance of the alternative hypothesis with a confidence level of 95%, meaning the gaze times in task $i$ (in the row) are significantly larger than those in task $j$ (in the column).

A first aspect to highlight is that the active eHMI decreases the observation times in the two experimental setups ($T_{av}$: t1 > t3 and t2 > t4). This effect cannot be appreciated as directly comparing the two types of deceleration, since the vehicle approaches at different speeds and does not reach the crosswalk at the same time. To analyse the time pedestrians spend observing the vehicle before crossing, we must focus on the CIT, which eHMI shortens by maintaining a smooth deceleration (CIT: t1 > t3). The same impact of eHMI during aggressive deceleration is only seen in the virtual setup (CIT: t2 vs t4).

When comparing the two scenarios, rather than making categorical statements, we show the probability of significant discrepancy between the tasks that will be used at the end of the research to quantify the *behavioural gap*. It is evident from the data provided in Table 3 that the gaze duration is greater in the virtual environment. Upon separately examining the time preceding and following the decision to cross, we see that the disparity is less pronounced within the CIT. In the absence of eHMI, pedestrians observe more of the vehicle before crossing in the virtual setup (CIT: $t1_{virtual} > t1_{real}$ and $t2_{virtual}$ vs $t2_{real}$) while, if eHMI is activated, the CIT resembles more closely. The notable differences in vehicle gazing times between both setups and across all experiment variations occur while walking on the road ($T_c$: $t_{i,virtual} > t_{i,real}$). This suggests that pedestrians pay significantly more attention to the vehicle after making the decision to cross when they are interacting in the virtual environment.

### 4.2. Space gap (L)

Box-plots of the space gap in each task (i.e., the distance between the pedestrian and the AV at the crossing decision) are depicted in Figure 7 for both real and virtual environments. In addition, Table 4 shows the results of the Student's t-test to evaluate the impact of eHMI and the type of manoeuvre.

With a confidence level of 95%, we assert that the smooth braking manoeuvre increases the distance to the vehicle when pedestrians decide to cross (Space gap: t1 > t2 and t3 > t4). The same applies to eHMI activation while maintaining the smooth braking manoeuvre (Space gap: t3 > t1). However, although the impact of the eHMI persists

**Figure 5.** Pedestrian-AV interaction in VR setup. (upper row) The pedestrian waits while eHMI displays a red status. (lower row) The eHMI switches to green status and the pedestrian decides to cross. From left to right: VR experimentation environment; overview of the simulated virtual scenario; pedestrian perspective; AV perspective (simulated camera).
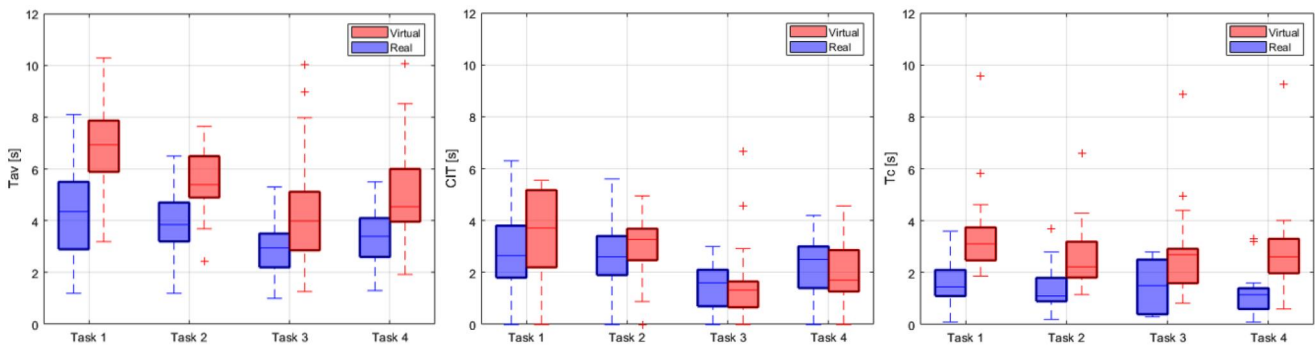


**Figure 6.** Gazing times in the virtual and real environment.

**Table 2.** Gazing times, student t-test, $\alpha = 0.05$.

| | | | | Task number $j$ | | | |
|---|---|---|---|---|---|---|---|
| $H_1 : \mu_i > \mu_j$ | | | | 1 | 2 | 3 | 4 |
| Task number $i$ | Tav | Real setup testing | 1 | – | 0 | 1 | 1 |
| | | | 2 | 0 | – | 1 | 1 |
| | | | 3 | 0 | 0 | – | 0 |
| | | | 4 | 0 | 0 | 1 | – |
| | | Virtual setup testing | 1 | – | 1 | 1 | 1 |
| | | | 2 | 0 | – | 1 | 1 |
| | | | 3 | 0 | 0 | – | 0 |
| | | | 4 | 0 | 0 | 0 | – |
| | CIT | Real setup testing | 1 | – | 0 | 1 | 1 |
| | | | 2 | 0 | – | 1 | 0 |
| | | | 3 | 0 | 0 | – | 0 |
| | | | 4 | 0 | 0 | 1 | – |
| | | Virtual setup testing | 1 | – | 0 | 1 | 1 |
| | | | 2 | 0 | – | 1 | 1 |
| | | | 3 | 0 | 0 | – | 0 |
| | | | 4 | 0 | 0 | 0 | – |
| | Tc | Real setup testing | 1 | – | 0 | 0 | 1 |
| | | | 2 | 0 | – | 0 | 0 |
| | | | 3 | 0 | 0 | – | 0 |
| | | | 4 | 0 | 0 | 0 | – |
| | | Virtual setup testing | 1 | – | 1 | 1 | 1 |
| | | | 2 | 0 | – | 0 | 0 |
| | | | 3 | 0 | 0 | – | 0 |
| | | | 4 | 0 | 0 | 0 | – |

**Table 3.** Certainty of the discrepancy, student t-test.

| Alternative hypothesis: | | Task 1 | Task 2 | Task 3 | Task 4 |
|---|---|---|---|---|---|
| Tav | $\mu_{virtual} > \mu_{real}$ | 99.9 % | 99.9 % | 99.4 % | 99.7 % |
| | $\mu_{real} > \mu_{virtual}$ | 0.0 % | 0.0 % | 0.6 % | 0.3 % |
| CIT | $\mu_{virtual} > \mu_{real}$ | 95.9 % | 88.4 % | 67.5 % | 33.9 % |
| | $\mu_{real} > \mu_{virtual}$ | 4.1 % | 11.6 % | 32.5 % | 66.1 % |
| Tc | $\mu_{virtual} > \mu_{real}$ | 99.9 % | 99.9 % | 99.7 % | 99.9 % |
| | $\mu_{real} > \mu_{virtual}$ | 0.0 % | 0.0 % | 0.3 % | 0.0 % |
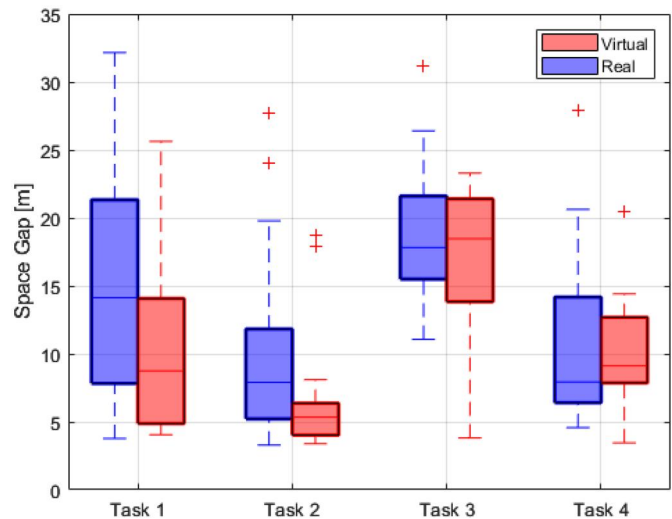


**Figure 7.** Box-plots Of the pedestrian-AV distances at the crossing decision. Virtual and real testing.

in the virtual setup by maintaining aggressive braking, this is not the case in the real setup (Space gap: t4 vs t2).

The Table 5 provides a direct comparison of space gaps across both setups. The findings indicate that the

**Table 4.** Space gap, student t-test, $\alpha = 0.05$.

| $H_1 : \mu_i > \mu_j$ | | | Task number $j$ | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| Task number $i$ | Real setup testing | 1 | – | 1 | 0 | 1 |
| | | 2 | 0 | – | 0 | 0 |
| | | 3 | 1 | 1 | – | 1 |
| | | 4 | 0 | 0 | 0 | – |
| | Virtual setup testing | 1 | – | 1 | 0 | 0 |
| | | 2 | 0 | – | 0 | 0 |
| | | 3 | 1 | 1 | – | 1 |
| | | 4 | 0 | 1 | 0 | – |



**Figure 8.** Keypoints in the image - DWPose.

**Table 5.** Certainty of the discrepancy in Space gap, student t-test.

| Alternative hypothesis: | Task 1 (%) | Task 2 (%) | Task 3 (%) | Task 4 (%) |
|---|---|---|---|---|
| $\mu_{virtual} > \mu_{real}$ | 0.1 | 0.6 | 12.0 | 22.4 |
| $\mu_{real} > \mu_{virtual}$ | 99.8 | 99.3 | 88.0 | 77.6 |

participants cross significantly earlier in the real setup (i.e., with a larger space gap) when the eHMI is deactivated (Space Gap: $t1_{real} > t1_{virtual}$ and $t2_{real} > t2_{virtual}$). Concerning experimental tasks which employ explicit communication (t3 and t4), the values of space gap exhibit greater similarity, leading to the non-rejection of the null hypothesis and, thus, precluding any definitive statement.

### 4.3. Body motion

Among the advantages of inserting real agents into a simulation environment (Serrano et al., 2022; 2023) is the possibility of generating synthetic sequences from various perspectives and configurations. To accomplish this, it is necessary to reconstruct the trajectory and 3D pose of the participant within the scenario, for which Perception Neuron's sensors and software provide an .fbx file over time (Noitom, 2022). This approach allows an accurate analysis of the participant's motion style throughout the experiments.

Within the scope of this research, we aimed to establish a methodology for acquiring motion metrics that could be standardised between both real-world and virtual environments. Employing a whole-pose estimator (Yang et al., 2023), we identify the keypoints of the pedestrian's body

**Table 6.** Body proportions. DWPose vs Perception Neuron sensors.

| Right - Left | DWPose (%) | PNs (%) |
|---|---|---|
| Trunk | 100.0 | 100.0 |
| Neck | 33.9 | 24.9 |
| Shoulder | 31.1–31.1 | 38.8–38.6 |
| Arm | 51.2–52.7 | 54.2–54.7 |
| Forearm | 44.4–43.6 | 47.2–48.6 |
| Hip | 18.7–18.7 | 18.1–17.9 |
| Leg | 65.8–64.9 | 76.8–76.9 |
| Foreleg | 62.1–62.6 | 72.4–72.0 |

(see Figure 8) in images captured by the front camera of the AV in the real environment (recall Figure 4). Subsequently, the 3D keypoints localised in the virtual environment are projected onto the plane parallel to the crosswalk, aligning with the format of the 2D estimator output. Table 6 outlines the body proportions derived from both procedures for constructing the pedestrian avatar.

Despite the non-correspondence of the keypoints given by Perception Neuron and the 2D estimator, this strategy allows us to conduct an equivalent analysis of the pedestrian's gait from the vehicle's perspective in the two environments. To calculate the pedestrians' pace while crossing, we apply the Fast Fourier Transform (FFT) (Oberst, 2007) on the lateral position on their ankles, and extract the highest peaks of the frequency spectrum. From the results of the Table 7 (Student's t-test cannot be used since it does not involve comparing a series of frequency magnitudes), it can be deduced that the eHMI activation makes pedestrians walk faster ($F_P$: t3 > t1 and t4 > t2), while in the VR setup they walk slightly slower in all experimental tasks ($F_P$: t $_{i,real}$ > t $_{i,virtual}$).

The previous inferences are also supported by the count of strides within the 4-second time window defined for the crossing decision (see Table 8). The stabilisation times of both ankles increase when the braking is aggressive or when eHMI is non-activated, indicating that pedestrians halt more their movement in such instances to evaluate the situation, as also shown in Table 8.

### 4.4. Subjective measures

To make categorical statements regarding the influence of the braking manoeuvre or eHMI on participants' questionnaire responses, we use the Wilcoxon signed-rank test (Woolson, 2007) which is an alternative to Student's t-test when working with ordinal or interval scales. The procedure for this non-parametric statistical test utilised to compare two related samples involves arranging the values of the absolute differences between the two samples and subsequently calculating a sum of ranks to determine whether the difference between the samples is statistically significant. The null hypothesis of the Wilcoxon test is that there is no difference between the two samples ($H_0 : \mu_i \leq \mu_j$), while the alternative hypothesis is that there is a significant difference ($H_1 : \mu_i > \mu_j$).

Table 9 provides categorical statements, where a 1 in a cell implies rejection of the null hypothesis and acceptance of the alternative hypothesis meaning that the responses to a

**Table 7.** Magnitudes of the three highest peaks, pace (FFT).

| Crossing pace rate | | | Frequencies, Steps per Second | | | |
|---|---|---|---|---|---|---|
| | | | 0.488 Hz | 0.732 Hz | 0.976 Hz | 1.220 Hz |
| Task number $i$ | Real setup testing | 1 | 2.304 ($f_3$) | 4.748 ($f_1$) | 3.589 ($f_2$) | |
| | | 2 | 2.570 ($f_3$) | 3.933 ($f_1$) | 3.454 ($f_2$) | |
| | | 3 | | 3.378 ($f_2$) | 4.973 ($f_1$) | 2.461 ($f_3$) |
| | | 4 | | 3.188 ($f_2$) | 4.372 ($f_1$) | 2.817 ($f_3$) |
| | Virtual setup testing | 1 | 2.788 ($f_2$) | 3.363 ($f_1$) | 2.087 ($f_3$) | |
| | | 2 | 2.828 ($f_2$) | 3.827 ($f_1$) | 2.195 ($f_3$) | |
| | | 3 | 1.683 ($f_3$) | 4.782 ($f_1$) | 2.825 ($f_2$) | |
| | | 4 | 2.157 ($f_3$) | 4.149 ($f_1$) | 2.676 ($f_2$) | |

**Table 8.** Stride count, swing and stabilisation times.

| | Task number $j$ | | | |
|---|---|---|---|---|
| 4-second window of crossing decision | 1 | 2 | 3 | 4 |
| Strides (real testing) | 2.54 | 2.28 | 3.29 | 2.94 |
| SD | 0.69 | 0.73 | 0.57 | 0.62 |
| Strides (VR testing) | 2.22 | 2.28 | 2.53 | 2.72 |
| SD | 0.63 | 0.45 | 0.50 | 0.65 |
| Left Swing Phase | 1.23 s | 1.11 s | 1.36 s | 1.26 s |
| Left Stance Phase | 2.77 s | 2.89 s | 2.64 s | 2.74 s |
| SD | 0.36 | 0.30 | 0.25 | 0.31 |
| Right Swing Phase | 1.31 s | 1.22 s | 1.38 s | 1.24 s |
| Right Stance Phase | 2.69 s | 2.78 s | 2.62 s | 2.76 s |
| SD | 0.33 | 0.30 | 0.22 | 0.28 |

**Table 9.** Wilcoxon signed rank test, Q1-3, $\alpha = 0.05$.

| | | | | Task number $j$ | | | |
|---|---|---|---|---|---|---|---|
| $H_1 : \mu_i > \mu_j$ | | | | 1 | 2 | 3 | 4 |
| Task number $i$ | Q1 | Real setup testing | 1 | – | 1 | 0 | 0 |
| | | | 2 | 0 | – | 0 | 0 |
| | | | 3 | 1 | 1 | – | 1 |
| | | | 4 | 0 | 1 | 0 | – |
| | | Virtual setup testing | 1 | – | 0 | 0 | 0 |
| | | | 2 | 0 | – | 0 | 0 |
| | | | 3 | 1 | 1 | – | 1 |
| | | | 4 | 1 | 1 | 0 | – |
| | Q2 | Real setup testing | 1 | – | 0 | 0 | 0 |
| | | | 2 | 1 | – | 1 | 0 |
| | | | 3 | 0 | 0 | – | 0 |
| | | | 4 | 1 | 0 | 1 | – |
| | | Virtual setup testing | 1 | – | 0 | 0 | 0 |
| | | | 2 | 1 | – | 1 | 1 |
| | | | 3 | 0 | 0 | – | 0 |
| | | | 4 | 1 | 0 | 1 | – |
| | Q3 | Real setup testing | 1 | – | 0 | 0 | 0 |
| | | | 2 | 0 | – | 0 | 0 |
| | | | 3 | 1 | 1 | – | 0 |
| | | | 4 | 1 | 1 | 0 | – |
| | | Virtual setup testing | 1 | – | 0 | 0 | 0 |
| | | | 2 | 0 | – | 0 | 0 |
| | | | 3 | 1 | 1 | – | 0 |
| | | | 4 | 1 | 1 | 0 | – |

**Table 10.** Certainty of the discrepancy, wilcoxon signed rank test.

| Alternative hypothesis: | | Task 1 (%) | Task 2 (%) | Task 3 (%) | Task 4 (%) |
|---|---|---|---|---|---|
| Q1 | $\mu_{virtual} > \mu_{real}$ | 0.6 | 7.8 | 44.4 | 71.9 |
| | $\mu_{real} > \mu_{virtual}$ | 99.4 | 92.2 | 55.6 | 28.1 |
| Q2 | $\mu_{virtual} > \mu_{real}$ | 67.7 | 87.9 | 58.3 | 30.1 |
| | $\mu_{real} > \mu_{virtual}$ | 32.3 | 12.1 | 41.7 | 69.8 |
| Q3 | $\mu_{virtual} > \mu_{real}$ | 0.0 | 0.0 | 95.0 | 77.6 |
| | $\mu_{real} > \mu_{virtual}$ | 0.0 | 0.0 | 4.9 | 22.4 |

question in task $i$ (in the row) are significantly greater than those in task $j$ (in the column).

With a confidence level of 95%, we assert that activating the eHMI enhances the pedestrian's perception of safety (Q1: t3 > t1 and t4 > t2). On the other hand, the smooth braking manoeuvre also increases the feeling of safety, although it is an effect that is not perceived within the virtual setup when the eHMI is disabled (Q1: t3 > t4 and t1 vs t2). Participants appreciate the difference between the smooth and aggressive type of maneuver (Q2: t2 > t1, t3 and t4 > t1, t3). It is worth noting that in the virtual setup the non-activation of the eHMI makes the same manoeuvre feel even more aggressive (Q2: t2 vs t4). Lastly, eHMI is considered to be useful (Q3: t3 > t1, t2 and t4 > t1, t2).

Table 10 presents direct comparisons between the questionnaire responses collected from the two setups. Pedestrians feel less safe in the virtual setup when the AV does not communicate its intentions explicitly (Q1: t1$_{real}$ > t1$_{virtual}$ and t2$_{real}$ > t2$_{virtual}$). In addition, they suggest the virtual eHMI has a more positive impact on their decision-making process (Q3: t3$_{virtual}$ > t3$_{real}$).

Assessing the sense of presence during the VR experiment can help to uncover the reasons of discrepancies in pedestrian crossing behaviour between the real and virtual testing setup. *Self-presence* measures how much users project their identity into a virtual world through an avatar, while *autonomous vehicle* and *environmental presence* examine how users interact with mediated entities and environments as if they were real. Most of the participants perceived the avatar as an extension of their body (M = 4.04, SD = 0.95), including when moving their hands or walking on the road. The vehicle presence was well rated (M = 3.94, SD = 0.97), although not all participants heard the sound of the engine or felt any braking manoeuvre threatening. Environmental-presence (M = 4.34, SD = 0.63) was the most satisfactory, as they claimed to have the feeling of actually being at a crosswalk.

## 5. Discussion

### 5.1. Variables influence in a real environment (RQ1)

Quantitative data shows that participants in the real-world crosswalk experiment notably extended the Space Gap when making their crossing decision if the AV performed a smooth braking manoeuvre. On the contrary, the impact of the "eHMI" factor seemed evident solely when activated alongside gentle braking. Activation of the visual interface did not accelerate pedestrian crossing decisions in instances of aggressive braking manoeuvres. Nevertheless, in the questionnaires, they indicated that both a braking manoeuvre signalling the vehicle's intention to yield and the activation of the eHMI conveyed a greater sense of safety compared to the opposite scenario. The FFT also notes that explicit communication encouraged them to cross the road faster after

**Table 11.** Certainty of the behavioural gap, fisher's method.

| $H_1 : L \downarrow, Q1 \downarrow, Q2 \uparrow, CIT \uparrow$ | Task 1 (%) | Task 2 (%) | Task 3 (%) | Task 4 (%) |
|---|---|---|---|---|
| More caution in virtual world | 99.9 | 99.7 | 72.5 | 26.4 |
| More caution in real world | 0.1 | 0.0 | 8.5 | 52.7 |

entering in the lane, while leading to a decrease in eye contact with the AV.

## 5.2. Variables influence in a virtual environment (RQ2)

In the virtual crosswalk experiment, the results reveal that both the smooth braking manoeuvre and the eHMI activation widen the Space Gap when pedestrians decide to cross. In the questionnaires, they report feeling safer when the eHMI is active compared to when it is not, and express a preference for smooth over aggressive braking, but only when the eHMI is operational. Explicit communication results in participants spending less time making eye contact with the AV to assess hazards. Additionally, according to FFT, it prompts them to walk slightly faster once they have entered the lane.

## 5.3. Measuring the behavioural gap (RQ3)

A first point to note is that the Student's t-test shows that the space gap L is significantly higher in the real environment than in the virtual environment when the visual interface (i.e., eHMI) is disabled. This finding is supported by the CITs, as pedestrians who spend more time observing the approaching vehicle encounter a smaller space gap L when they eventually decide to cross. Still, we must mention that this discrepancy in the crossing behaviour between the real and virtual testing setup disappears when the eHMI starts working. The CITs and the distance separating the pedestrian from the AV when deciding to cross then are not noticeably different.

The responses to the questionnaire follow the same line of argument. Participants perceive a greater sense of safety in the real environment compared to the virtual environment when the eHMI is deactivated, and feel equally safe when it is activated. This leads us to think the eHMI contributes to increased confidence in the experiment and prompts participants to make the decision to cross earlier. Furthermore, this effect is particularly pronounced in the virtual environment, where the eHMI is most prominently visible, as reported by Q3. Not activating the eHMI heightens the perception of the virtual AV's aggressive braking as even more aggressive (Q2: $t2_{virtual} > t4_{virtual}$).

To gather the evidence on the existence of the *behavioural gap* we employ the Fisher's method (Yoon et al., 2021), a statistical technique utilised to combine the results of independent significance tests performed on the same data set. The Fisher's method is particularly useful when multiple hypothesis tests are performed and it is desired to combine the evidence from all of these tests to reach an overall conclusion. In Table 11 the general alternative hypothesis ($H_1$) is defined as follows: pedestrians adopt a more cautious crossing behaviour in the virtual world than in the real world (i.e., less Space Gap, less trust Q1, more

perceived aggressiveness Q2, more CIT). It is shown that participants demonstrate increased caution in the simulated scenario when the eHMI is inactive, while no conclusive findings can be drawn in the opposite direction.

Comparing the impact of each variable, in the real-world environment an implicit communication is obeyed before an explicit one (Space gap: $t1_{real} > t4_{real}$), while in the virtual environment more trust is placed in explicit communication (Q1: $t4_{virtual} > t1_{virtual}$). FFTs indicate that participants walked more slowly on the road in the virtual environment, probably because they were more curious and entertained by observing the AV, as shown by the eye gazing data.

## 5.4. Limitations

This research was conducted in a simple traffic scenario, featuring only one approaching vehicle, devoid of any social activities in the background. This could have led to collecting information only on individual decision-making and crossing behaviours without the influence of other co-located pedestrians and vehicles. The lighting and weather conditions were also specific (clear sunny day), and results in different contexts may vary.

The immersive VR system for real agents currently employed (Serrano et al., 2022; 2023) relies on Unreal Engine 4 and Windows operating system due to the CARLA build and dependencies unique to Meta Quest 2 for Windows. Due to sensors simulation entails a high computational cost, the scene rendering is limited to 15-20 frames per second, which could affect the participants' sense of presence. Moreover, since most of participants had little to no prior VR experience before the experiment, it remains unclear whether the *behavioural gap* results would have differed had the participants been regular users of virtual reality. Increasing the sample size (N = 18) in future studies would allow for a more comprehensive exploration of potential effects, including gender and age disparities in response to the variables investigated. Nonetheless, despite this limitation, we believe the results and conclusions outlined herein provide valuable insights and lay a foundation for further research in the field of real agent simulation.

As demonstrated, the investigation of the *behavioural gap* is intricately tied to specific contextual factors such as the type of scenario, traffic and environmental conditions, etc. Therefore, results cannot be readily generalised across other contexts. However, the methodology presented in terms of combined analysis based on self-reporting and direct measures of behaviour in equivalent real-world and virtual settings, is transferable to other types of scenarios, including different application domains (e.g., robotics). Studying the *behavioural gap* is essential for validating any behavioural data from real subjects interacting with autonomous systems obtained in virtual environments.

## 6. Conclusions and future work

This study advances our understanding of the gap between simulation and reality in contexts that incorporate the

activity of real agents for autonomous driving research. The digital twin of a crosswalk and an AV was crafted by replicating its driving style and the design of the eHMI it featured, within the CARLA simulator. The participants, who had no previous experience in VR, acted more cautiously in their role as a pedestrian in the simulation by delaying lane entry, slowing their movements and paying more attention to all elements of the environment. This did not prevent us from corroborating the impact of implicit and explicit vehicle communication on the crossing behaviour of pedestrians introduced into the virtual environment. Based on our findings, participants prioritised implicit communication over explicit communication in the real-world scenario, whereas in the VR tests, their decisions were more influenced by explicit communication.

For future work in this field, we emphasise the importance of familiarising the participants with the VR environment, not only by proposing them to explore the virtual world for a few minutes before starting the tests, but also by involving them in simulated examples with vehicular traffic and street crossings that do not count for the drawing of conclusions. In order to resemble the effects of the braking manoeuvre and eHMI in the simulator to those in the real-world, techniques could be implemented to enhance the *AV presence* rating through more realistic motion dynamics and an engine sound that commensurate with its revolutions. In addition, the brightness of the virtual eHMI could be adjusted to match its showiness in the real environment. If sufficient data were available, a more automatic approach to assessing *behavioural gap* could be achieved, e.g., by learning behavioural differences within a particular scenario and subsequently generating corresponding scores or distances.

## Acknowledgment

## Disclaimer

The views expressed in this article are purely those of the authors and may not, under any circumstances, be regarded as an official position of the European Commission.

## Disclosure statement

The authors report there are no competing interests to declare.

## Funding

## Safety and ethical considerations

The fundamental pillar guiding the design of the experiments has been the safety and comfort of all participants above any other consideration. On one hand, we chose to implement Level 3 automation in our real testing conditions, despite the fact that Level 4 automation could have been possible. This decision necessitated the presence of a backup driver ready to resume control when needed. In addition, a human supervisor in the rear seats was monitoring the status of all perception and control systems, including access to an emergency stop function. Therefore, human intervention was always possible, both by the backup driver and the supervisor. On the other hand, the braking profile was designed to be extremely conservative, maintaining a substantial margin for reaction, prioritising safety above all else. Furthermore, we rigorously followed internal and institutional ethical assessment and validation procedures, which included informing the participants and obtaining their written consent, ensuring data privacy, allowing subjects to withdraw from the experiments at any time, and implementing data anonymisation, among other protocols.

## ORCID

Sergio Martín Serrano 🆔 http://orcid.org/0000-0002-8029-1973
Rubén Izquierdo 🆔 http://orcid.org/0000-0002-6722-3036
Iván García Daza 🆔 http://orcid.org/0000-0001-8940-6434
Miguel Ángel Sotelo 🆔 http://orcid.org/0000-0001-8809-2103
David Fernández-Llorca 🆔 http://orcid.org/0000-0003-2433-7110

## References

Ackermann, C., Beggiato, M., Schubert, S., & Krems, J. F. (2019). An experimental study to investigate design and assessment criteria: What is important for communication between pedestrians and automated vehicles? *Applied Ergonomics*, 75, 272–282. https://doi.org/10.1016/j.apergo.2018.11.002

Almeaibed, S., Al-Rubaye, S., Tsourdos, A., & Avdelidis, N. P. (2021). Digital twin analysis to promote safety and security in autonomous vehicles. *IEEE Communications Standards Magazine*, 5(1), 40–46. https://doi.org/10.1109/MCOMSTD.011.2100004

Bazilinskyy, P., Dodou, D., & De Winter, J. (2019). Survey on ehmi concepts: The effect of text, color, and perspective. *Transportation Research Part F: Traffic Psychology and Behaviour*, 67, 175–194. https://doi.org/10.1016/j.trf.2019.10.013

Bazilinskyy, P., Kooijman, L., Dodou, D., Mallant, K., Roosens, V., Middelweerd, M., Overbeek, L., & de Winter, J. (2022). *Get out of the way! examining ehmis in critical driver-pedestrian encounters in a coupled simulator* [Paper presentation]. Proceedings of the. 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 360–371). https://doi.org/10.1145/3543174.3546849

Candela, E., Parada, L., Marques, L., Georgescu, T.-A., Demiris, Y., & Angeloudis, P. (2022). *Transferring multi-agent reinforcement learning policies for autonomous driving using sim-to-real* [Paper presentation]. 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 8814–8820). https://doi.org/10.1109/IROS47612.2022.9981319

Chang, C.-M., Toda, K., Gui, X., Seo, S. H., & Igarashi, T. (2022). *Can eyes on a car reduce traffic accidents?* [Paper presentation]. Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ser. AutomotiveUI '22, New York, NY, USA (pp. 349–359). https://doi.org/10.1145/3543174.3546841

Chang, C.-M., Toda, K., Sakamoto, D., & Igarashi, T. (2017). *Eyes on a car: An interface design for communication between an autonomous car and a pedestrian* [Paper presentation]. Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications (pp. 65–73). https://doi.org/10.1145/3122986.3122989

Clamann, M., Aubert, M., & Cummings, M. (2017). *Evaluation of vehicle-to-pedestrian communication displays for autonomous vehicles* [Paper presentation]. Proceedings of the 96th Annual Transportation Research Board Meeting. https://trid.trb.org/View/1437891

Cruz, N., & Ruiz-del Solar, J. (2020). *Closing the simulation-to-reality gap using generative neural networks: Training object detectors for soccer robotics in simulation as a case study* [Paper presentation]. 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1–8). https://doi.org/10.1109/IJCNN48605.2020.9207173

De Winter, J. C. (2013). Using the student's t-test with extremely small sample sizes. *Practical Assessment, Research, and Evaluation*, *18*(1), 10. https://doi.org/10.7275/e4r6-dj05

Deb, S., Carruth, D. W., & Hudson, C. R. (2020). How communicating features can help pedestrian safety in the presence of self-driving vehicles: Virtual reality experiment. *IEEE Transactions on Human-Machine Systems*, *50*(2), 176–186. https://doi.org/10.1109/THMS.2019.2960517

Detjen, H., Faltaous, S., Pfleging, B., Geisler, S., & Schneegaß, S. (2021). How to increase automated vehicles' acceptance through in-vehicle interaction design: A review. *International Journal of Human-Computer Interaction*, *37*, 1–23. https://doi.org/10.1080/10447318.2020.1860517

Dey, D., Habibovic, A., Klingegård, M., Fabricius, V., Andersson, J., & Schieben, A. (2018). *Workshop on methodology: Evaluating interactions between automated vehicles and other road users—what works in practice*? 09 (pp. 17–22). https://doi.org/10.1145/3239092.3239095

Dey, D., Matviienko, A., Berger, M., Pfleging, B., Martens, M., & Terken, J. (2021). Communicating the intention of an automated vehicle to pedestrians: The contributions of ehmi and vehicle behavior. *it - Information Technology*, *63*(2), 123–141. https://doi.org/10.1515/itit-2020-0025

Dey, D., van Vastenhoven, A., Cuijpers, R. H., Martens, M., & Pfleging, B. (2021). *Towards scalable ehmis: Designing for av-vru communication beyond one pedestrian* [Paper presentation]. 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ser. AutomotiveUI '21. New York, NY, USA: Association for Computing Machinery (pp. 274–286). https://doi.org/10.1145/3409118.3475129

Dijksterhuis, C., Lewis-Evans, B., Jelijs, B., de Waard, D., Brookhuis, K., & Tucha, O. (2015). The impact of immediate or delayed feedback on driving behaviour in a simulated pay-as-you-drive system. *75*, 93–104. https://doi.org/10.1016/j.aap.2014.11.017

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). *CARLA: An open urban driving simulator* [Paper presentation]. Proceedings of the 1st Annual Conference on Robot Learning (pp. 1–16). https://proceedings.mlr.press/v78/dosovitskiy17a.html

Eady, T. (2019). *Simulations can't solve autonomous driving because they lack important knowledge about the real world – large-scale real world data is the only way*. [Online]. https://medium.com/@strangecosmos/simulation-cant-solve-autonomous-driving-because-it-lacks-necessary/empirical-knowledge-403feeec15e0

Feng, Y., Xu, Z., Farah, H., & Arem, B. (2023). *Does another pedestrian matter? -a virtual reality study on the interaction between multiple pedestrians and autonomous vehicles in shared space*, 09. https://doi.org/10.31219/osf.io/r3udx

Fernandez-Llorca, D., & Gomez, E. (2023). Trustworthy artificial intelligence requirements in the autonomous driving domain. *Computer Magazine*, *56*(2), 29–39. https://doi.org/10.1109/MC.2022.3212091

Fratini, E., Welsh, R., & Thomas, P. (2023). Ranking crossing scenario complexity for ehmis testing: A virtual reality study. *Multimodal Technologies and Interaction*, *7*(2), 16. https://doi.org/10.3390/mti7020016

Fridman, A., Mehler, B., Xia, L., Yang, Y., Facusse, L. Y., & Reimer, B. (2017). *To walk or not to walk: Crowdsourced assessment of external vehicle-to-pedestrian displays*. ArXiv, vol. abs/1707.02698. https://arxiv.org/abs/1707.02698

Furuya, H., Kim, K., Bruder, G., Wisniewski, P. J., & Welch, G. F. (2021). *Autonomous vehicle visual embodiment for pedestrian interactions in crossing scenarios: Virtual drivers in avs for pedestrian crossing* [Paper presentation]. Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3411763.3451626

Gadipudi, N., Elamvazuthi, I., Sanmugam, M., Izhar, L. I., Prasetyo, T., Jegadeeshwaran, R., & Ali, S. S. A. (2022). *Synthetic to real gap estimation of autonomous driving datasets using feature embedding* [Paper presentation]. 2022 IEEE 5th International Symposium in Robotics and Manufacturing Automation (ROMA) (pp. 1–5). https://doi.org/10.1109/ROMA55875.2022.9915679

García Daza, I., Izquierdo, R., Martínez, L. M., Benderius, O., & Fernández Llorca, D. (2023). Sim-to-real transfer and reality gap modeling in model predictive control for autonomous driving. *Applied Intelligence*, *53*(10), 12719–12735. https://doi.org/10.1007/s10489-022-04148-1

Ge, Y., Wang, Y., Yu, R., Han, Q., & Chen, Y. (2019). *Demo: Research on test method of autonomous driving based on digital twin* [Paper presentation]. 2019 IEEE Vehicular Networking Conference (VNC) (pp. 1–2). https://doi.org/10.1109/VNC48660.2019.9062813

Gonzalo, R. I., Maldonado, C. S., Ruiz, J. A., Alonso, I. P., Llorca, D. F., & Sotelo, M. A. (2022). Testing predictive automated driving systems: Lessons learned and future recommendations. *IEEE Intelligent Transportation Systems Magazine*, *14*(6), 77–93. https://doi.org/10.1109/MITS.2022.3170649

Guo, F., Lyu, W., Ren, Z., Li, M., & Liu, Z. (2022). A video-based, eye-tracking study to investigate the effect of ehmi modalities and locations on pedestrian–automated vehicle interaction. *Sustainability*, *14*(9), 5633. https://doi.org/10.3390/su14095633

Hensch, A.-C., Neumann, I., Beggiato, M., Halama, J., & Krems, J. F. (2020). Effects of a light-based communication approach as an external hmi for automated vehicles - a wizard-of-oz study. *Transactions on Transport Sciences*, *10*(2), 18–32. https://doi.org/10.5507/tots.2019.012

Hensch, A.-C., Neumann, I., Beggiato, M., Halama, J., & Krems, J. F. (2020). How should automated vehicles communicate? – effects of a light-based communication approach in a wizard-of-oz study. In *Advances in Human Factors of Transportation*, edited by N. Stanton, 79–91. Springer International Publishing. https://doi.org/10.1007/978-3-030-20503-4_8

Hu, X., Li, S., Huang, T., Tang, B., Huai, R., & Chen, L. (2024). How simulation helps autonomous driving: A survey of sim2real, digital twins, and parallel intelligence. *IEEE Transactions on Intelligent Vehicles*, *9*(1), 593–612. https://doi.org/10.1109/TIV.2023.3312777

Hub, F., Hess, S., Lau, M., Wilbrink, M., & Oehl, M. (2023). Promoting trust in HAVs of following manual drivers through implicit and explicit communication during minimal risk maneuvers. *Frontiers in Computer Science*, *5*(1154476) https://doi.org/10.3389/fcomp.2023.1154476

Izquierdo, R., Martín, S., Alonso, J., Parra, I., Sotelo, M. A., & Fernaádez–Llorca, D. (2023). *Human-vehicle interaction for autonomous vehicles in crosswalk scenarios: Field experiments with pedestrians and passengers* [Paper presentation]. 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC) (pp. 2473–2478). https://doi.org/10.1109/ITSC57777.2023.10422195

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, *7*(4), 396–403. https://doi.org/10.9734/BJAST/2015/14975

Kalantari, A. H., Yang, Y., de Pedro, J. G., Lee, Y. M., Horrobin, A., Solernou, A., Holmes, C., Merat, N., & Markkula, G. (2023). Who goes first? a distributed simulator study of vehicle-pedestrian interaction. *Accident; Analysis and Prevention*, *186*, 107050. https://doi.org/10.1016/j.aap.2023.107050

Kalra, N., & Paddock, S. M. (2016). *Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? RAND Corporation*. Research Report.

Le, D. H., Temme, G., & Oehl, M. (2020). Automotive eHMI Development in Virtual Reality: Lessons Learned from Current Studies. *11*, 593–600. https://doi.org/10.1007/978-3-030-60703-6_76

Lee, Y. M., Madigan, R., Uzondu, C., Garcia, J., Romano, R., Markkula, G., & Merat, N. (2022). Learning to interpret novel ehmi: The effect of vehicle kinematics and ehmi familiarity on pedestrian' crossing behavior. *Journal of Safety Research*, 80, 270–280. https://doi.org/10.1016/j.jsr.2021.12.010

Li, M., Holthausen, B. E., Stuck, R. E., & Walker, B. N. (2019). *No risk no trust: Investigating perceived risk in highly automated driving* [Paper presentation]. Proceedings of the 11th AutomotiveUI Conference. Association for Computing Machinery (pp. 177–185).

Li, Y., Dikmen, M., Hussein, T. G., Wang, Y., & Burns, C. (2018). *To cross or not to cross: Urgency-based external warning displays on autonomous vehicles to improve pedestrian crossing safety* [Paper presentation]. Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications., ser. AutomotiveUI '18. Association for Computing Machinery (pp. 188–197). https://doi.org/10.1145/3239060.3239082

Llorca, D. F., & Gómez, E. (2021). Trustworthy autonomous vehicles. EUR 30942 EN, Publications Office of the European Union, Luxembourg vol. JRC127051. https://publications.jrc.ec.europa.eu/repository/handle/JRC127051

Mason, B., Lakshmanan, S., McAuslan, P., Waung, M., & Jia, B. (2022). Lighting a path for autonomous vehicle communication: The effect of light projection on the detection of reversing vehicles by older adult pedestrians. *International Journal of Environmental Research and Public Health*, 19(22), 14700. no https://doi.org/10.3390/ijerph192214700

Merat, N., Louw, T., Madigan, R., Wilbrink, M., & Schieben, A. (2018). What externally presented information do vrus require when interacting with fully automated road transport systems in shared space? *Accident; Analysis and Prevention*, 118, 244–252. https://doi.org/10.1016/j.aap.2018.03.018

Moore, D., Currano, R., Strack, G. E., & Sirkin, D. (2019). *The case for implicit external human-machine interfaces for autonomous vehicles* [Paper presentation]. Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ser. AutomotiveUI '19. New York, NY, USA: Association for Computing Machinery (pp. 295–307). https://doi.org/10.1145/3342197.3345320

Nascimento, A., Queiroz, A. C., Vismari, L., Bailenson, J., Cugnasca, P., Junior, J., & Almeida, J. (2019). The Role of Virtual Reality in Autonomous Vehicles' Safety. 12, 50–507. https://doi.org/10.1109/AIVR46125.2019.00017

Ngo, A., Bauer, M. P., & Resch, M. (2021). *A multi-layered approach for measuring the simulation-to-reality gap of radar perception for autonomous driving* [Paper presentation]. 2021 IEEE International Intelligent Transportation Systems Conference (ITSC) (pp. 4008–4014). https://doi.org/10.1109/ITSC48978.2021.9564521

Nguyen, T. T., Holländer, K., Hoggenmueller, M., Parker, C., & Tomitsch, M. (2019). *Designing for projection-based communication between autonomous vehicles and pedestrians* [Paper presentation]. Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ser. AutomotiveUI '19. New York, NY, USA: Association for Computing Machinery (pp. 284–294). https://doi.org/10.1145/3342197.3344543

Noitom (2022). *Perception neuron studio system*. [Online]. Available: https://neuronmocap.com/perception-neuron-studio-system

Oberst, U. (2007). The fast fourier transform. *SIAM Journal on Control and Optimization*, 46(2), 496–540. https://doi.org/10.1137/060658242

Pareigis, S., & Maaß, F. L. (2022). *Robust neural network for sim-to-real gap in end-to-end autonomous driving* [Paper presentation]. Proceedings of the. 19th International Conference on Informatics in Control, Automation and Robotics - ICINCO, INSTICC. SciTePress (pp. 113–119). https://doi.org/10.5220/0011140800003271

Rasouli, A., & Tsotsos, J. K. (2020). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE Transactions on Intelligent Transportation Systems*, 21(3), 900–918. https://doi.org/10.1109/TITS.2019.2901817

Reway, F., Hoffmann, A., Wachtel, D., Huber, W., Knoll, A., & Ribeiro, E. (2020). *Test method for measuring the simulation-to-reality gap of camera-based object detection algorithms for autonomous driving* [Paper presentation]. 2020 IEEE Intelligent Vehicles Symposium (IV) (pp. 1249–1256). https://doi.org/10.1109/IV47402.2020.9304567

Schwarz, C., & Wang, Z. (2022). The role of digital twins in connected and automated vehicles. *IEEE Intelligent Transportation Systems Magazine*, 14(6), 41–51. https://doi.org/10.1109/MITS.2021.3129524

Serrano, S. M., Izquierdo, R., Daza, I. G., Sotelo, M. A., & Fernández Llorca, D. (2023). *Digital twin in virtual reality for human-vehicle interactions in the context of autonomous driving* [Paper presentation]. 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC) (pp. 590–595). https://doi.org/10.1109/ITSC57777.2023.10421914

Serrano, S. M., Llorca, D. F., Daza, I. G., & Sotelo, M. Á. (2022). *Insertion of real agents behaviors in CARLA autonomous driving simulator* [Paper presentation]. Proceedings of the 6th international conference on computer-human interaction research and applications, CHIRA (pp. 23–31). https://doi.org/10.5220/0011352400003323

Serrano, S. M., Llorca, D. F., Daza, I. G., & Sotelo, M. Á. (2023). Realistic pedestrian behaviour in the carla simulator using vr and mocap. In *Computer-Human Interaction Research and Applications, Communications in Computer and Information Science (CCIS)* (Vol. 1882). Springer.

Song, Y., Jiang, Q., Chen, W., Zhuang, X., & Ma, G. (2023). Pedestrians' road-crossing behavior towards ehmi-equipped autonomous vehicles driving in segregated and mixed traffic conditions. *Accident; Analysis and Prevention*, 188, 107115. https://doi.org/10.1016/j.aap.2023.107115

Stadler, S., Cornet, H., Theoto, T. N., & Frenkler, F. (2019). *A tool, not a toy: Using virtual reality to evaluate the communication between autonomous vehicles and pedestrians* (pp. 203–216). Springer.

Steve Coast (2024). https://www.openstreetmap.org/

Stocco, A., Pulfer, B., & Tonella, P. (2023). Mind the gap! a study on the transferability of virtual vs physical-world testing of autonomous driving systems. *IEEE Transactions on Software Engineering*, 49(4), 1928–1940. https://doi.org/10.1109/TSE.2022.3202311

Tian, K., Tzigieras, A., Wei, C., Lee, Y. M., Holmes, C., Leonetti, M., Merat, N., Romano, R., & Markkula, G. (2023). Deceleration parameters as implicit communication signals for pedestrians' crossing decisions and estimations of automated vehicle behaviour. *Accident; Analysis and Prevention*, 190, 107173. https://doi.org/10.1016/j.aap.2023.107173

Tom, A., & Granié, M.-A. (2011). Gender differences in pedestrian rule compliance and visual search at signalized and unsignalized crossroads. *Accident Analysis & Prevention*, 43(5), 1794–1801. https://doi.org/10.1016/j.aap.2011.04.012

Velasco, J. P. N., Lee, Y. M., Uttley, J., Solernou, A., Farah, H., van Arem, B., Hagenzieker, M., & Merat, N. (2021). Will pedestrians cross the road before an automated vehicle? the effect of drivers' attentiveness and presence on pedestrians' road crossing behavior. *Transportation Research Interdisciplinary Perspectives*, 12, 100466. https://doi.org/10.1016/j.trip.2021.100466

Woolson, R. F. (2007). Wilcoxon signed-rank test. In *Wiley encyclopedia of clinical trials* (pp. 1–3).Wiley.

Yang, Z., Zeng, A., Yuan, C., & Li, Y. (2023). Effective whole-body pose estimation with two-stages distillation. In *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (pp. 4212–4222). https://doi.org/10.1109/ICCVW60793.2023.00455

Yoon, S., Baik, B., Park, T., & Nam, D. (2021). Powerful p-value combination methods to detect incomplete association. *Scientific Reports*, 11(1), 6980. https://doi.org/10.1038/s41598-021-86465-y

Yu, B., Chen, C., Tang, J., Liu, S., & Gaudiot, J.-L. (2022). Autonomous vehicles digital twin: A practical paradigm for autonomous driving system development. *Computer Magazine*, 55(9), 26–34. https://doi.org/10.1109/MC.2022.3159500

Yun, H., & Park, D. (2021). *Simulation of self-driving system by implementing digital twin with gta5* [Paper presentation]. 2021 International Conference on Electronics, Information, and Communication (ICEIC) (pp. 1–2). https://doi.org/10.1109/ICEIC51217.2021.9369807

Zimmermann, R., & Wettach, R. (2017). *First step into visceral interaction with autonomous vehicles* [Paper presentation]. Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 58–64). https://doi.org/10.1145/3122986.3122988

## About the authors

**Sergio Martín Serrano** is a PhD student at the University of Alcalá. His research interests are focused on the analysis of Vulnerable Road Users (VRUs) behaviours using Virtual Reality (VR) and autonomous driving simulators, covering predictive perception and human-vehicle interaction.

**Rubén Izquierdo** is Assitant Professor at the University of Alcalá. His research interest focuses on prediction of vehicles' behaviour, human-vehicle interaction, and control algorithms for highly automated and cooperative vehicles.

**Iván García Daza** is Associate Professor at the University of Alcalá. He specialises in Deep Reinforcement Learning for decision-making in autonomous vehicles and in the application of Explainable AI to decision-making systems.

**Miguel Ángel Sotelo** is Full Professor at the University of Alcalá. His research interests focus on road users' behaviour understanding and prediction, human-vehicle interaction, and Explainable AI for decision-making in autonomous vehicles.

**David Fernández Llorca** is Scientific Officer at the European Commission - Joint Research Centre, and Full Professor at the University of Alcalá. His research interests include trustworthy AI for transportation, human-centred autonomous vehicles, predictive perception, and human-vehicle interaction.

## Appendix A

### Self-presence scale items

To what extent did you feel that … (1= not at all − 5 very strongly)

1. You could move the avatar's hands.
2. The avatar's displacement was your own displacement.
3. The avatar's body was your own body.
4. If something happened to the avatar, it was happening to you.
5. The avatar was you.

### Autonomous vehicle presence scale items

To what extent did you feel that … (1= not at all − 5 very strongly)

1. The vehicle was present.
2. The vehicle dynamics and its movement were natural.
3. The sound of the vehicle helped you to locate it.
4. The vehicle was aware of your presence.
5. The vehicle was real.

### Environmental presence scale items

To what extent did you feel that … (1= not at all − 5 very strongly)

1. You were really in front of a pedestrian crossing.
2. The road signs and traffic lights were real.
3. You really crossed the pedestrian crossing.
4. The urban environment seemed like the real world.
5. It could reach out and touch the objects in the urban environment.