POLITECNICO DI TORINO
Repository ISTITUZIONALE

Comparing state-of-the-art and emerging augmented reality interfaces for autonomous vehicle-to-pedestrian communication

Original
Comparing state-of-the-art and emerging augmented reality interfaces for autonomous vehicle-to-pedestrian communication / Pratticò, F.G., Lamberti, F., Cannavò, A., Morra, L., Montuschi, P.. - In: IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. - ISSN 0018-9545. - STAMPA. - 70:2(2021), pp. 1157-1168. [10.1109/TVT.2021.3054312]

Availability:
This version is available at: 11583/2864934 since: 2021-05-11T23:14:20Z

Publisher:
IEEE

Published
DOI:10.1109/TVT.2021.3054312

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

29 November 2026

Comparing State-of-the-Art and Emerging Augmented Reality Interfaces for Autonomous Vehicle-to-Pedestrian Communication

F. Gabriele Pratticò, *Student Member, IEEE,* Fabrizio Lamberti, *Senior Member, IEEE,* Alberto Cannavò, *Student Member, IEEE,* Lia Morra, *Senior Member, IEEE,* Paolo Montuschi, *Fellow, IEEE*

***Abstract*—In the last few years, a considerable literature has grown around the theme of how to provide pedestrians and other vulnerable road users (VRUs) with a clear indication about a fully autonomous vehicle (FAV)'s status and intentions, which is crucial to make FAVs and VRUs coexist. So far, a variety of external interfaces leveraging different paradigms and technologies have been created. Proposed designs include vehicle-mounted devices (like LED panels), short-range on-road projection, and road infrastructure interfaces (e.g., special asphalts with embedded displays). These designs have been experimented in different settings, using mockups, specially prepared vehicles, or virtual environments, with heterogeneous evaluation metrics. Some promising interfaces based on Augmented Reality (AR) have been proposed too, but their usability and effectiveness have not been tested yet. This paper aims to complement such body of literature by presenting a comparison of state-of-the-art interfaces and new designs under common conditions. To this aim, an immersive Virtual Reality-based simulation was developed, recreating a well-known scenario used in previous works represented by pedestrian crossing in urban environments under non-regulated conditions. A user study was then performed to investigate the various dimensions of vehicle-to-pedestrian interaction in both objective and subjective terms. Results showed that, although an interface clearly standing out over all the considered dimensions does not exists, one of the studied AR designs was able to provide state-of-the-art results in terms of safety and trust, at the cost of a higher cognitive effort and lower intuitiveness compared to LED panels showing anthropomorphic features. Together with rankings on the various dimensions, indications about advantages and drawbacks of the various alternatives that emerged from the study could be an important information source for next developments in the field.**

***Index Terms*—fully autonomous vehicles, human-machine interaction, virtual reality, augmented reality, vehicle-to-pedestrian communication, pedestrian crossing.**

I. Introduction

ADVANCEMENTS in the field of automation are continuous, and promise to revolutionize most of everyone's activities. Autonomous vehicles, in particular, will play a key role in this ongoing revolution. While in the early 2010's autonomous vehicles were still regarded as visionary by almost all car manufacturers, today this sector has changed into a multi-billion dollars business. Huge investments are being made [1], and the expectations are that fully autonomous vehicles (FAVs) [2] will reach the market by the next decade. Given the disruptive potential of FAVs, their acceptance could be hindered by open challenges related not only to technical aspects, but also to societal factors [3], which deserve significant attention as both people with and without technical expertise will be requested to trust machines [4]. This need can be addressed from two viewpoints: that of the drivers (in the future, of the passengers) and of in-vehicle interfaces; and that of vulnerable road users (VRUs), like, e.g., pedestrians, and of interfaces external to the vehicle.

In the last years, significant efforts have been devoted to designing interaction paradigms capable of raising occupants awareness about vehicles status and intentions, with the aim of improving trust in their autonomous decisions [5]. However, tackling the needs and expectations of VRUs is substantially more complicated. Driving is a complex social behavior based on continuous interactions with other drivers and road users in uncertain and ambiguous situations. When adding (or replacing traditional vehicles with) FAVs, the lack of human drivers introduces a communication gap, potentially dangerous for all road users [6]. For this reason, interfaces with VRUs (mainly pedestrians) recently started to be investigated to increase the safety and acceptability of FAVs.

Several alternatives have been proposed already, focusing on the most common vehicle-to-pedestrian interactions, i.e., road crossings. Possible interaction paradigms include showing anthropomorphic features [7], or using LED strips/panels to communicate FAV's intentions [8]–[10]. In other designs, on-vehicle visual hints were replaced by on-road projections, possibly leveraging well-known metaphors like crosswalk or stop signs [11], [12]. Some prototype implementations also introduced changes in the road infrastructure, collecting data from connected vehicles to communicate with pedestrians.

Whenever a new vehicle-to-pedestrian interaction paradigm (or interface) was introduced, it was generally compared with some of the previously proposed ones in qualitative and/or quantitative terms, often working with prototype implementations or mockups. Some works resorted to Virtual Reality (VR) for comparing a number of alternatives at once [11]. Despite the great relevance of these works for next developments in the field, available experimental evaluations suffer from several drawbacks. For instance, when a representative set of interfaces are considered, not all the experimental conditions

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The authors are with the GRAINS – GRAphics And INtelligent Systems group at the Dipartimento di Automatica e Informatica of Politecnico di Torino, 10129, Torino, Italy. e-mail: (see http://grains.polito.it/people.php).

Manuscript received XXXX XX, XXXX; revised XXXX XX, XXXX.

studied in other works are investigated (and viceversa). When the above requirements are met, often the analysis does not fully recreate real-world conditions or does not address the same objective and subjective dimensions which were deemed relevant in other studies. Finally, and most importantly, not all the interaction paradigms devised so far have been tested. This is the case, for instance, of Augmented Reality (AR)-based interfaces, which proved to be particularly effective for in-vehicle interaction [5]. Some AR-based designs for vehicle-to-pedestrian interfaces have been also proposed but, at present, they have not been compared (or even tested) yet with the mentioned alternatives.

In this work, a VR-based simulation system was first designed, by taking into account the above issues and endowing it with the capabilities required to support a fair comparison of multiple interface designs relying on heterogeneous technologies and offering different functionalities. This system was then exploited to run a user study aimed to compare the most relevant interfaces proposed so far from different categories. Specifically, AR-based interfaces were considered, to shed some light on their possible role in next-generation vehicle-to-pedestrian interaction paradigms. A wide set of metrics derived from relevant literature was used, providing interested readers with a comprehensive picture of advantages and drawbacks of available paradigms.

II. Related work

As discussed in the previous section, the implicit communication mechanism provided by vehicle movement alone might not be able to guarantee an efficient interaction between VRUs and FAVs. This is why, in the last years, a number of interfaces have been developed to elicit vehicles' status and/or intentions.

1) Vehicle-mounted interfaces: the first interfaces proposed relied on visual hints provided by equipment mounted on the vehicle exterior. A first example is represented by the *"Eyes on a car"* design [7], which aims to replace eye contact between pedestrians and drivers of conventional vehicles with digital eyes placed on the front lights. As soon as the vehicle's sensors detect a pedestrian intending to cross, the eyes start staring at him or her to signal that it will stop to allow crossing. Otherwise, the eyes gaze remains fixed on the road. Experiments performed in a Virtual Environment (VE) showed that pedestrians were faster in deciding whether to cross or not when the interface was available. However, some users found the eyes artificial, leading to an undesirable Uncanny Valley effect [11], and not sufficiently reliable or safe.

Other interfaces were inspired by the well-known *traffic light* metaphor. For instance, the concept in [8] is based on a LED light placed in the vehicle's front. Two lighting patterns were proposed: green, flashing yellow, red (GYR), and white, flashing red, red (WRR). For each set, the first color indicates when it is safe to cross, the last color when it is not; to refer to intermediate situations, the middle color is used. Studies performed on this design confirmed that users were able to properly associate colors to messages the vehicle intended to communicate. However, the flashing red of WRR tended to be (mis)interpreted as a danger signal rather than a warning one, especially compared to the GYR flashing yellow. In another study dealing with color codes [13], the authors suggested to remove the intermediate warning state (collapsing it to danger), as they realized that pedestrians tended to selfishly cross even when they were putting the vehicle's passengers at risk. Authors also stated that an indication of approaching vehicle could be more effective than a safe/unsafe to cross information (though without any experimental evaluation).

Another type of interface which proved to be quite intuitive for pedestrians consists in *LED strips* mounted in various positions of the vehicle's exterior. An example is given in [9], [10]. In this implementation, the strip is mounted over the windshield, and only the central LEDs are lighted while driving. When a pedestrian in detected, LEDs start to light from the center to the edges of the strip. During crossing, all the LEDs are lighted, and when the vehicle resumes motion, a reverted animation from edges to center is activated. Experiments carried out with a Wizard of Oz technique on a specially prepared vehicle showed that, after a short training, users were able to properly use the interface. It is worth noting that, in [9], two different experiments were run to gather both direct and indirect measures of the pedestrian perceived safety. In [11], a comparison between the above interface and a different design with two strips on the vehicle's sides was performed in a VE. Users preferred the latter interface showing appreciation for its intuitiveness and unambiguity, though continuous feedback provided was judged as not particularly useful.

Other examples of this interface category are provided in [14]–[16]. Besides the position on the vehicle's exterior, the main differences among the various designs lay in the strip's shape, in the lighting pattern, and in the color(s) used (and their meanings). In some cases, the above features are combined together. For instance, in [17], a flashing yellow light is used to indicate that the vehicle is not yielding, whereas a blue light moving from top to bottom indicates that it is going to stop; a fading blue light shows that vehicle is waiting and, lastly, flashing yellow is used again to indicate vehicle restart.

An alternative design, still based on a LED strip but exploiting again anthropomorphic features like the interface in [7], is the *"Smiling car"* [11]. The interface shows a horizontal yellow line in normal driving conditions, which changes to a smile when a pedestrian is detected to inform him or her that the vehicle will yield. Based on studies performed in a scenario involving one vehicle and one pedestrian [11], [18], users found this interface as very simple to use and able to provide unambiguous information.

LED panels have been used also to provide pedestrians with written information on vehicle's status (e.g., "Braking" [18]) or on what to do (e.g., "Cross now" [19]–[21]). Compared with other interfaces, such a direct communication approach did not prove particularly effective, due to possible readability and language issues [11]. In some configurations, interfaces request the pedestrians to notify their intention to cross with a specific gesture. In this case, LED strips have been used to inform them that gesture was correctly recognized [22].

2) Projection-based interfaces: the main limitation of vehicle-mounted interfaces is that they communicate only by means of visual signs placed on the vehicle, which may be

poorly visible either because of adverse weather or lighting conditions, or simply because the vehicle is still too far.

A way pursued to cope with these limitations consists in *projecting* visual indications on the road. For instance, the prototype interface in [23] projects a pattern of parallel lines in front of the vehicle. Lines are perpendicular to the driving direction, and get closer/farther when the vehicle is decelerating/accelerating. Experiments performed with a real, non-autonomous vehicle showed that pedestrians were faster in deciding whether to cross or not, but they focused their attention on the road rather than on the vehicle itself. In [12], a more sophisticated projection pattern was used. During normal driving, a wave-shaped red pattern is projected. When a pedestrian is detected and vehicle starts to slow down, pattern color changes to yellow. When vehicle has come to a complete stop and crossing is safe, pattern changes into a green crosswalk shape, which turns to red when vehicle is going to restart again. The implementation proposed in [11], characterized by a lower number of states, was rated as very pleasant; it was also judged as futuristic, but this result could be due to the particular vehicle the interface was mounted onto. Despite the increased visibility, these interfaces may not work well with all road conditions. Moreover, even though they rely on the well-known crosswalk concept which should be familiar for pedestrians, experiments indicated that these interfaces induce a high mental workload.

3) Smart road interfaces: a different approach to support vehicle-to-pedestrian communication consists in using the road infrastructure itself. In the so called "Smart roads" concept, visual hints are provided through LED panels embedded in the road pavement to indicate when it is safe to cross (e.g., by showing a white crosswalk) and when it is not (e.g., lighting red bars on the sidewalk edge). For instance, the prototype implemented in London [24] was judged as very trustworthy. The main drawback of such interfaces is the cost associated with updating the road infrastructure [11].

4) Multi-modal interfaces: the above designs rely only on the visual communication channel, which is also the focus of the present study. It is worth noticing, however, that this choice makes these interfaces not suitable for all VRUs, e.g., for visually impaired persons. For this reason, multi-modal interfaces including a combination of visual, audio and haptic-based notifications have been proposed [25]. However, the audio channel proved not particularly effective in noisy traffic.

5) AR interfaces: the last category in this review is represented by AR interfaces. Thanks to technological advancements and the ever-larger availability of consumer-grade devices [26], it is possible to imagine a not too distant future in which pedestrians will wear their own AR glasses, enabling the development of sophisticated vehicle-to-pedestrian and, more generally, human-robot interaction paradigms [27]. However, experiments in this field are still rare. In a recent study [28], three concepts were explored. In the first one, AR visual hints are exploited to show to the pedestrian, through a blue tape overlapped to his or her field of view, the safer path to follow for crossing the road; moreover, AR-based yellow arrows are drawn in front of the vehicle to show where it could stop, if needed. The second design, referred to as "safe zones", uses AR to draw large green regions indicating where it is safe to cross, and red regions closer to the vehicle indicating where it would be dangerous to cross. The third design is a combination of the previous ones. A user study was performed, in which participants were shown static representations of the three designs. Despite the amount of visual indications provided, participants preferred the third design.

The above review, which showed that various interface categories have indeed been proposed but only a few of them have been tested under representative and/or comparable conditions, motivated the design of the simulation system and the experimental analysis that are reported in this work.

III. Simulation system

In this section, the VR-based system created to support the study of the various interfaces is illustrated. In particular, the simulation environment is presented first. Afterwards, interfaces implementation and scenario configuration are discussed.

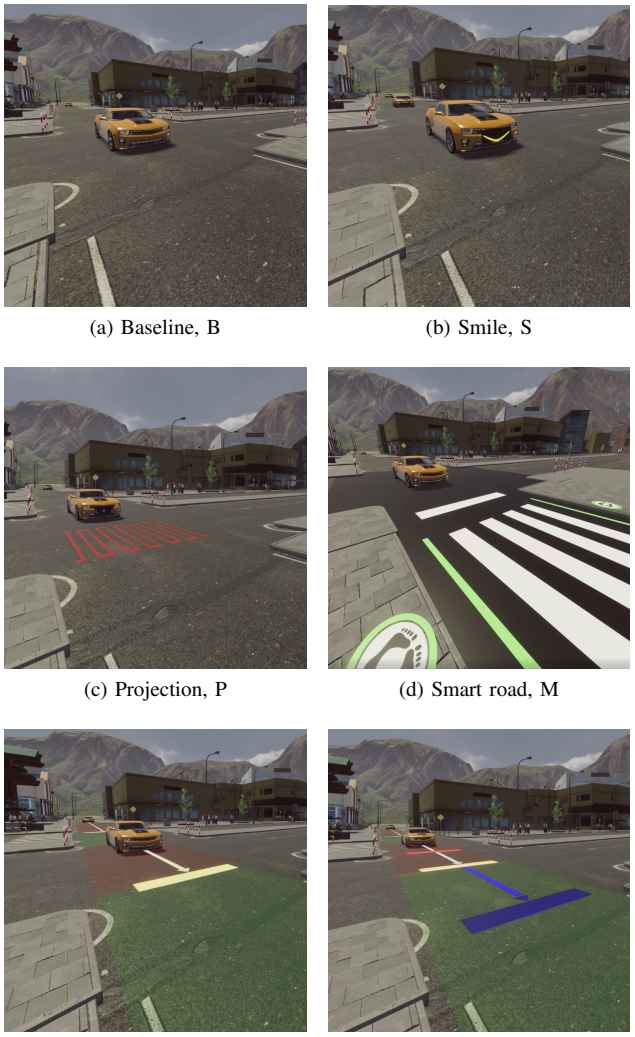
A. Simulation environment

The simulation environment has been built on top of the AirSim open-source software [29]. Originally developed by Microsoft as a simulator for collecting data needed to train computer vision algorithms for unmanned vehicles, AirSim supports hardware-in-the loop simulations and is characterized by an extremely high visual fidelity. Also, a "Windridge City" urban scenario is provided free of charge.

However, the "as-is" performance of AirSim is not suitable for an effective fruition via immersive VR. Hence, data collection from vehicle's simulated sensors was disabled (as not required in this work), and a number of optimizations were implemented in order to target the minimum framerate required for immersive experiences (90 frames per seconds or more to prevent motion sickness). To this purpose, among the graphics platforms supported by AirSim, Unity was selected; optimizations leveraged ad-hoc functionalities available in this graphics engine to improve the performance of the application logic and of graphics processing.

In particular, concerning the former aspect, the novel Data-Oriented Technology Stack (DOTS) paradigm was used, which allows developers to exploit the parallel processing capabilities and large cache availability of modern processors. The benefit in performance is especially relevant for applications that need to handle multiple instances of the same simulated object, in this case the logic (and physics) of vehicles (and their parts, like wheels) and of pedestrians. The above paradigm was implemented by relying on the Unity's architectural pattern named Entity Component System (ECS) together with the C# Job System, which enables an effective handling of frame-synchronized multi-threading, and the Burst Compiler, which allows for the generation of high-performance native code that can exploit parallel hardware acceleration.

For the second aspect, the Unity's Scriptable Render Pipeline (SRP) was exploited, since it allows to better customize graphics quality based on hardware capabilities and application requirements. In particular, the High Definition Render Pipeline (HDRP), which has been recently made



(a) Baseline, B (b) Smile, S

(c) Projection, P (d) Smart road, M

(e) Safe roads, F (f) Safe roads extended, E

Fig. 1. Simulation environment and interfaces included in the evaluation depicted while the vehicle is braking.

available for VR applications, was used. The HDRP allowed to effectively manage the rendering of 3D elements with different levels of detail (LODs) based on actual distance from the user, to use dynamic resolution for reducing the workload on the GPU, to enable efficient calculation of reflections, transparencies and shadows, and to activate volumetric lighting for realistic effects. Furthermore, the SRP supports the Shader Graph (SG), which is a visual shader programming tool that was of paramount importance to effectively backing the implementation of the vehicle's interfaces.

Finally, several minor changes were applied to the original urban scenario like, among others, simplifying the colliders' shape, activating GPU instancing on materials, and enabling asynchronous map loading.

B. Interface selection and implementation

In order to provide the users with a reference implementation, a "Baseline" behavior was defined for the virtual FAVs, in which no specific interface is available (Fig. 1a).

Braking was implemented as follows: as soon as the FAV detects a pedestrian intending to cross, because close enough to the sidewalk edge, looking either at the vehicle or at the road, and within the detection range, it starts braking with a constant deceleration. Deceleration is calculated by considering the current speed and aiming to reach a full stop at a certain distance from the pedestrian; in preliminary experiments, other deceleration strategies were found to largely alter the users' perception of the interface behavior, as confirmed also by [30], and hence were excluded from the evaluation. If the FAV is not able to stop in due time (too fast and/or too close when crossing starts), the horn is activated to signal the danger.

By default, AirSim's built-in vehicles are provided with a controller logic for driving them with keyboard or joystick. However, in this work FAVs' behavior described above was scripted using the ECS, and speed was managed using a PID controller (with output shaping) since, as it will be shown later, only deceleration and acceleration on a straight path had to be handled in the designed experiments.

By building on top of the above "Baseline" (abbreviated B), one interface was implemented for each of the categories discussed in detail in Section 2, focusing specifically on visual interfaces. To this aim, we chose the interfaces that scored better in experiments reported in previous literature. Neither the FAV's behavior nor the vehicle shape were altered to avoid biases in the interface comparison [11].

For vehicle-mounted interfaces [11], [18], the "Smiling car" concept was selected [11], later referred to as "Smile" (abbreviated S). In normal driving conditions, an horizontal straight yellow line is shown on the vehicle's front side; when a pedestrian is detected and the FAV starts braking, the line turns smoothly into a smile (Fig. 1b) to signal that it will yield.

Among projection-based interfaces [11], [12], [23], the concept originally presented in [12] was considered (later named "Projection", P). To foster comparability of results, the implementation in [11] was used (Fig. 1c), but the sound played at vehicle restart was removed in order to make the interface rely only on visual indications, like the other ones. The interface integrates a LED panel on the vehicle's front side, whose pattern changes based on actual projection: all LEDs lighted in normal driving conditions, LEDs lighting from edges to center while decelerating, LEDs lighting in the crossing direction while stopped, then transition to the original state at vehicle restart.

For the "Smart road" (M) category, the implementation studied in [24] was selected (Fig. 1d). In the original work, pedestrian detection is performed by the infrastructure (thus, for instance, crosswalk marking appears even when there is no vehicle approaching); in this study, detection was moved onto the FAVs side. When there is no vehicle approaching, the interface does not provide any feedback (like with the other designs). When a vehicle is approaching, the interface's state is controlled by the vehicle itself, simulating a connection established between the two entities; if the vehicle is able to stop, a crosswalk appears together with the predicted stop position, and white lines on the sidewalk edge are turned green (otherwise, they are turned red and no crosswalk is shown).

Lastly, for the AR-based category, the third design in [28]

leveraging AR hints for showing zones suitable/unsuitable for crossing was chosen, later referred to as "Safe roads" (F). The interface shows in front of the vehicle a yellow arrow, terminated by a yellow bar whose length is equal to the current estimated stopping distance (Fig. 1e). The arrow length is determined only by the vehicle's dynamics, and it is not related in any way to pedestrian detection. When accelerating or moving at cruise speed, the stopping distance is calculated by applying a reference deceleration (deemed comfortable for the passengers) of 3 m/s^2 [31]. When braking, the stopping distance is obtained by sampling a pre-calculated braking curve at the current velocity: the curve is selected among two alternatives (comfortable 3 m/s^2, and maximum 6 m/s^2), picking the one with the lowest deceleration value still greater than current deceleration. To reinforce the feedback provided by the interface, the region of the road between the vehicle and the arrowhead is colored red, the rest green. It is worth observing that, since the arrowhead (and the colored region) reflects solely the dynamics of the FAV, the pedestrian is not provided with a clear, immediate indication of the fact that the vehicle actually detected him or her, and will stop accordingly.

Since the goal of the present work was to study, in particular, the effectiveness of AR-based vehicle-to-pedestrian interaction, a further interface was added to the analysis. This choice was motivated by the fact that, unfortunately, the interface in [28] is one of the few concepts proposed so far that has still to be tested by letting users actually experiment road crossing with it. Hence, a new interface was designed, named "Safe roads extended", by combining key communication abilities that were found in some of the best interface designs, but were lacking in F. The behavior of the original yellow arrow is maintained. However, a red tick is added on the arrow body (Fig. 1f) to show the pedestrian where the vehicle would stop in case of an emergency brake (6 m/s^2): the distance between the vehicle and the tick represents the minimum stopping distance. Moreover, an additional blue arrow, with a blue bar on the head, is drawn only when a pedestrian is detected (thus avoiding unneeded visual clutter). The blue arrowhead position is fixed as the vehicle's speed decreases, showing the (estimated) point where it will stop; thus, the pedestrian is implicitly informed that he or she has been detected, and can use this information to decide whether to cross or not.

C. Virtual scenario

In order to test the interfaces, a representative case study within the selected VE had to be defined. Based on previous literature, we decided to focus on an unregulated crossing scenario [11], [20], which could become commonplace in urban contexts populated only by FAVs. Moreover, in the absence of other external signals (e.g., traffic lights), VRUs have to rely on information provided by the FAVs' interfaces only, leveling out possible environmental contributions. In particular, a one-way, 5m wide road was chosen, as done, e.g., in [11], [12]. In this way, the users could ground their decisions to cross or not on the observation of a limited number of vehicles moving on a straight path rather than, e.g., on vehicles changing lanes or stopping at different locations; evaluation metrics can also be made independent of the crossing direction.

Following the approach adopted, e.g., in [20], [32], vehicles were organized in a pattern whose characteristics were controlled in order to ensure realism while at the same time presenting the users a number of different situations stimulating the various interfaces in many ways (Fig. 2). Cruise speed was set to 50 Km/h (14 m/s) for all the vehicles, since previous work indicates that above 40 Km/h the demographics have low to no impact on the pedestrians' behavior [33]. Since the experience was expected to strongly depend on space (hence, time) separating the users from the approaching vehicle, it was decided to vary the distances between consecutive vehicles, setting them to either 45m, 60m or 100m; within the pattern, a pseudo-random distribution guaranteeing a uniform distribution of the inter-vehicle distance was used. So called "not-yielding" vehicles were also included in the pattern: this characteristic has been used rarely in the literature, but authors of works where it was not exploited lamented the fact that its lack negatively influenced the realism of the simulation [11]. These vehicles are essential to explore trust in human-to-vehicle interactions in the presence of faulty FAVs. Detection range of vehicles' sensors was set to 60m, and yielding vehicles were programmed to stop 5m ahead of the crossing pedestrian.

IV. Experimental protocol

In this section, the experimental protocol devised to carry out the evaluation will be illustrated in detail.

A. Experiment design and preparation

For each interface, the user is instructed to cross as quickly and as many times as possible, whenever he or she feels it safe to cross. The testing of each experience is concluded when the following two conditions are met: the user has completed at least 15 crossings in total and at least one per distance. This choice ensures that the user experiences the full spectrum of situations within a reasonable amount of time. Additionally, as done in [20] a maximum limit to the duration of the simulation was introduced, in our case an upper bound of 300 generated vehicles was set. FAVs are set to become visible at 140m (Fig. 2) and are allowed to queue, though queues are prevented from growing too much in order to avoid long waiting times that could be physically and/or mentally demanding in VR. The user was informed that crossings in front of queued vehicles would not be considered (they will be later referred to as invalid crossings).

As done in almost all the works carried out so far, the user is controlling the only pedestrian that can interact with vehicles in the above scenario: in this way, it is possible to isolate the contribution of his or her interactions with the approaching vehicles from other possibly disturbing factors.

It is worth observing that interactions in more complex scenarios could be investigated in the future using the devised software, which already integrates a traffic simulation system encompassing both vehicles and pedestrians (not used in the experiments).

Experiments were designed according to a within-subjects logic: all the subjects tested all the interfaces, starting with

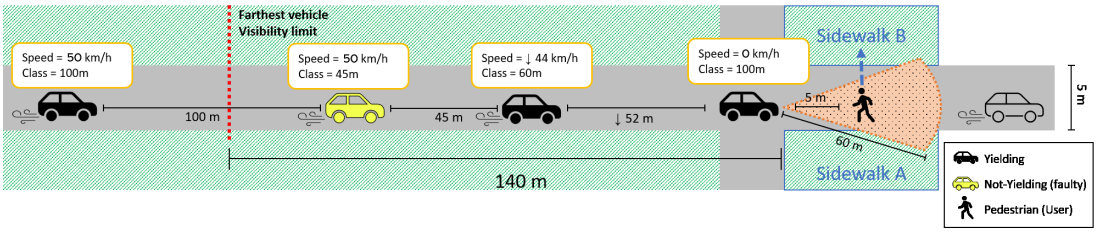


Fig. 2. Scenario devised for the experiment: pattern used for generating vehicles, visibility and detection ranges, safety stop distance, and crosswalk length.

the Baseline. Latin Square order was then used to define the sequence of the remaining tests. In the future, it is reasonable to expect that pedestrians will be accustomed with FAVs' interfaces: we sought to replicate a similar condition in our experimental setting by showing each subject a video to familiarize with each interface prior to the virtual experience. In each video, a pedestrian attempts to cross the road under three different conditions: the vehicle is either too close to stop (hence, it will pass by horning at the pedestrian), it is forced to an emergency brake, or it has sufficient time to stop smoothly. Videos shown to participants are made available[1].

B. Hardware configuration and physical setup

The VR headset device selected to let the users immerse themselves in the created VE was the Samsung Odyssey. It is an HMD equipped with an AMOLED display with a resolution of 1440×1600 pixels per eye, a refresh rate of 90Hz and a horizontal field of view of 110°. The headset supports 3D audio and relies on inside-out tracking technology, meaning that it does not requires external equipment to determine the user's (better, his or her head's) and the controllers' pose in the real environment. An untethered setup was defined using the MSI VR One Backpack PC to run the simulation. The backpack integrates an Intel Core i7-7820HK and a NVidia GeForce GTX 1070 in less than 3.5 Kg. Users could thus walk in a physical space mapping one-to-one with the selected portion of the VE where they were expected to cross. This setup was expected to lead to a more realistic experience, boosting the sense of immersion and presence [34].

C. Objective evaluation metrics

During the simulation, user's behavior is logged by collecting several quantitative measures. For each interaction with an approaching vehicle, the system records the time at which the pedestrian is detected (and the vehicle starts braking, thus initiating the negotiation) and the time at which the pedestrian reaches the opposite sidewalk, i.e., the crossing ends (Fig. 2). *Crossing time* (*CT*), calculated as the difference between the above times, is logged; previous works speculated that it may be associated with user's uncertainty [11]. When the user enters the road, the distance of the approaching vehicle (*distance at crossing*, or *DAC*) and its speed (*speed at crossing*, or *SAC*) are recorded. The interval between the time the user was detected and he or she left the sidewalk (and negotiation ends) is defined as *decision time* (*DT*). The system also keeps track of possible collisions with the vehicles. Aborted crossings, occurring when the user gets off and on the same sidewalk (without reaching the opposite one), are also recorded. An *efficiency* metric was defined as the total number of valid crossings (i.e., non-invalid and non-aborted) divided by the time elapsed from the first to the last crossing.

D. Subjective evaluation metrics

To further evaluate users' experience, a questionnaire[2] was developed based on previous literature on the subject.

A before-experience section (*BEQ*) collects demographic information about the participant, as well as his or her opinion about FAVs, crossing habits, and experience with technologies used in the experiments. Possible symptoms associated with motion sickness are recorded using the Simulation Sickness Questionnaire (SSQ) [35].

Afterwards, the testing of the individual interfaces begins. The after-video tutorial section (*ATQ*) aims at ensuring that the interface functioning has been fully understood prior to the actual VR experience: the user is required to answer some questions, describe the interface behavior in words and sketch a graphical representation on paper. Then, the after-interface questionnaire (*AIQ*) is used to investigate various dimensions of vehicle-to-user interaction after the VR experience (lasting approximately 10 minutes). The AIQ combines questions from the Trust Scale (TS) [36], the System Usability Scale (SUS) [37], the NASA Task Load Index (NASA-TLX) [38] and the Short User Experience Questionnaire (UEQ-S) [39]. Additional, custom questions are used to, e.g., investigate the level of perceived safety associated with a given interface, including whether the participant has felt the need to wait for the vehicle to come to a full stop before attempting to cross. Finally, the participant is requested to rate a number of features of the tested interface (perceived safety, familiarity, workload, etc.) with respect to the Baseline.

After testing all the interfaces, a post-experience questionnaire (PEQ) is administered, which includes questions from the SSQ to evaluate possible discomforts due to the VE, from the VRUse tool [40] to determine the usability of the VR simulation, and from the iGroup Presence Questionnaire (IPQ) [41] to measure immersion and presence. Additional questions verify that the proper level of realism is reached in the simulation. Finally, the participant is asked to rank (without ties) all the interfaces based on individual features and overall. A final open interview concludes the experience.

[1]Videos: https://youtu.be/RoPURY1dlZE

[2]Questionnaire: http://tiny.cc/6xoksz

E. Data analysis

Collected data were analyzed using MS Excel with the Real-Statistics add-on (v7.1). Comparative analyses were performed using Friedman's test (pass condition was set at $p \leq 0.05$). Post-hoc comparison on pairwise groups was further applied using the Conover's test. For non-comparative questionnaire items, Cronbach's Alpha (α) on clusters of related items was computed to test internal consistency. Interface rankings in the PEQ were aggregated using a Bucklin voting system; in particular, the relative placement scoring system (RPSS) [42] was applied considering a majority of half of the sample size.

V. Results

Experimental evaluation involved 12 volunteers (8 males, 4 females) aged between 24 and 46 ($M = 28.00$, $SD = 4.68$). Participants were selected among Politecnico di Torino students and from the authors' social networks. Based on demographic information collected from the BEQ, 17% of the participants were unfamiliar with immersive VR (never used, or used once), whereas 83% said they use this technology very often or on a daily basis. Participants reported crossing the roads under non-regulated conditions either very often/daily (50%) or occasionally (50%), and were on average trustworthy about FAVs ($M = 3.42$, $SD = 0.78$, $\alpha = 0.77$). In the following, experimental results will be presented, focusing first on the participants' perception of the VE and of simulation quality, since possible discomforts associated to the use of VR could introduce biases in the evaluation. Afterwards, their experience with the various interfaces will be compared by leveraging subjective feedback and objective measurements.

A. Virtual experience and simulation quality

No significant effect was registered on pre-post experience conditions (p-value computed using the two-tailed Mann-Whitney U-test) for the three SSQ clusters, i.e., oculomotor (O), disorientation (D), and nausea (N) symptoms, as well as overall (T): $\Delta\overline{O} = 13.3 \pm 18.6$ ($p = 0.08$), $\Delta\overline{D} = 8.1 \pm 15.5$ ($p = 0.48$), $\Delta\overline{N} = 14.3 \pm 37.4$ ($p = 0.51$), and $\Delta\overline{T} = 14.3 \pm 22.4$ ($p = 0.07$). Hence, psychophysiological alterations related to motion sickness that could have affected participants' attention level, reaction time, etc. apparently had no significant influence on the experiments.

Results from the IPQ shows that participants experienced a fairly good general sense of presence in the VE ($M = 4.0$, $SD = 0.58$), and assigned medium-high scores for the remaining indicators, i.e., spatial presence ($M = 3.56, SD = 0.46, \alpha = 0.80$), involvement ($M = 3.10$, $SD = 0.66$, $\alpha = 0.76$), and realism ($M = 3.10$, $SD = 0.50$, $\alpha = 0.78$).

Participants were also satisfied with the operation of the locomotion technique ($M = 4.42$, $SD = 0.50$, $\alpha = 0.79$), as well as with the visual quality and the simulation fidelity of the VE ($M = 4.08$, $SD = 0.50$, $\alpha = 0.72$). Scores related to the perception of the FAVs, as measured by the PEQ, are reported in Table I ($\alpha = 0.76$). Participants were satisfied overall with the FAV simulation. Based on open feedback collected after the experiments, lower scores assigned to some questions were due to the non-yielding behavior of some vehicles which, as said, was intentionally introduced to simulate failures in pedestrian detection: in this respect, 75.0% of the participants stated that they felt awkward when, right in the middle of the crossing, the vehicle sometimes did not recognize them (and they ascribed this behavior to faulty, or not smart enough, FAVs). Few participants reported also that they found FAVs too polite with pedestrians compared to how a human driver would behave in such situations.

TABLE I
FAV simulation perception. Inverted items are marked with *.

Item	M	SD
You think that vehicles were driven by a human (1) or were fully autonomous (5)	3.75	0.72
The FAVs showed adequate decision making skills	4.17	0.56
The FAVs struggled in case of sudden, unexpected or abrupt pedestrian behavior*	2.08	1.04
The FAVs made mistakes*	2.08	0.96
The FAVs seemed intelligent	3.83	0.56
Overall, you are satisfied about FAV simulation	4.00	0.59

B. Interface comparison

Once the representativeness of the simulated scenario was validated, the analysis moved to comparing selected interfaces.

1) Subjective results: Ranking of the features analyzed in the PEQ were first analyzed. Results aggregated with the RPSS are reported (together with pairwise significances) in Table II. Considering the various features, all the interfaces were ranked significantly higher than B (apart from familiarity, as it could be expected); this finding suggests that the introduction of a vehicle-to-pedestrian interface was effective. Worth to mention are the different placements of the two AR interfaces; according to participants preferences, E significantly outperformed F for most of the features. The least significant differences were obtained for cognitive workload: only S was judged as significantly less demanding in terms of mental effort than B, M, F and E. In fact, S was also ranked frequently among the two best interfaces, largely overcoming E in features regarding immediateness (like ease of use and intuitiveness).

Very similar considerations can be made based on rankings assigned in the AIQ: by performing a consistency check between the two observations (AIQ, PEQ) of the same features, all reached significant levels with high correlation ($\rho \geq 0.83$), except ambiguity ($p = 0.16$) and latency ($p = 0.27$).

Like for mental demand, none of the NASA-TLX indicators (normalized to interface B on a per-sample basis) were found to be significant, with the exception of S that was considered as less demanding than B ($p = 0.01$).

The best and second-best interfaces along each dimension are highlighted in Table II. It can be concluded that, while both S and E stood out from the other interfaces, S required a lower effort for the user. Comparing the AR interfaces, participants were more effective in completing the crossing task (higher efficiency, lower latency) when using E. The blue arrow included in E could justify its significantly higher ranking in terms of visibility. Importantly, both S and P were

TABLE II
RANKINGS AND P-VALUES FOR PAIRWISE COMPARISONS. AGGREGATED RANKING IS CALCULATED BY COMBINING INDIVIDUAL FEATURE RANKINGS (OVERALL EXCLUDED). REMAINING RANKINGS ARE OBTAINED DIRECTLY FROM PEQ ANSWERS. INVERTED ITEMS ARE MARKED WITH *.

	B	S	P	M	F	E	B-S	B-P	B-M	B-F	B-E	S-P	S-M	S-F	S-E	P-M	P-F	P-E	M-F	M-E	F-E
Overall	6	2	3	4	5	1	**0.00**	**0.00**	**0.04**	**0.02**	**0.00**	0.68	**0.02**	**0.03**	0.50	**0.05**	**0.07**	0.28	0.84	**0.00**	**0.01**
Safety	6	2	3	4	5	1	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	0.58	0.21	**0.04**	0.40	0.07	**0.01**	0.78	0.40	**0.04**	**0.00**
Familiarity	1	3	4	2	5	6	**0.00**	**0.00**	**0.04**	**0.00**	**0.00**	0.22	**0.00**	**0.02**	**0.00**	**0.00**	0.22	**0.00**	**0.00**	**0.00**	**0.01**
Intuitiveness	6	1	3	2	5	4	**0.00**	**0.05**	**0.03**	0.81	0.40	0.28	0.34	**0.00**	**0.03**	0.90	**0.03**	0.23	**0.02**	0.19	0.28
Mental workload*	6	1	2	3	4	5	**0.00**	0.13	0.30	0.73	0.82	0.11	**0.04**	**0.01**	**0.01**	0.64	0.25	0.20	0.49	0.42	0.91
Efficiency	6	2	3	4	5	1	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	0.58	0.08	0.22	0.13	0.22	0.49	**0.04**	0.58	**0.00**	**0.01**
Ambiguity*	6	1	2	3	5	4	**0.00**	**0.00**	**0.01**	0.08	**0.00**	0.63	0.28	**0.05**	0.47	0.55	0.15	0.81	0.40	0.72	0.23
Ease of use	6	1	2	3	5	4	**0.00**	**0.00**	**0.03**	0.16	**0.02**	0.25	**0.01**	**0.00**	**0.02**	0.12	**0.02**	0.20	0.44	0.80	0.30
Latency*	6	2	3	5	4	1	**0.00**	0.10	0.31	0.13	**0.00**	**0.01**	**0.00**	**0.01**	0.70	0.53	0.90	**0.03**	0.61	**0.01**	**0.03**
Visibility	6	4	5	3	2	1	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**	0.55	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.24	**0.00**	**0.02**
Aggregated	6	2	3	4	5	1	**0.00**	**0.00**	**0.02**	**0.04**	**0.00**	0.45	0.11	**0.05**	0.90	0.38	0.24	0.38	0.75	**0.05**	**0.04**

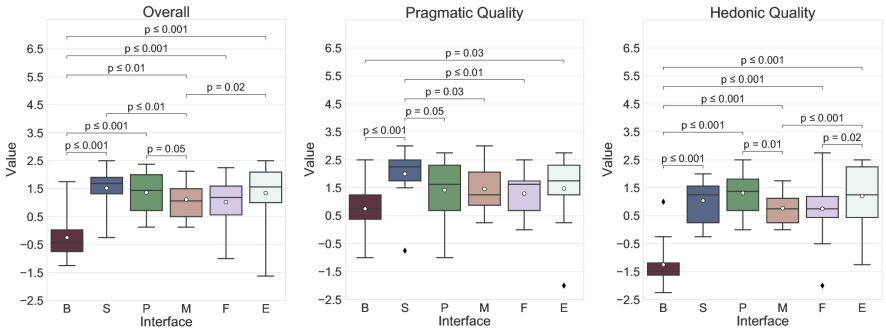


Fig. 3. Results of the UEQ-S: overall, as well as pragmatic and hedonic quality dimensions shown (p-values reported for significant pairwise comparisons).

TABLE III
TRUST AND SAFETY. MEANS (STANDARD DEVIATIONS), AND P-VALUES FOR PAIRWISE COMPARISONS. INVERTED ITEMS ARE MARKED WITH *.

	B	S	P	M	F	E	B-S	B-P	B-M	B-F	B-E	S-P	S-M	S-F	S-E	P-M	P-F	P-E	M-F	M-E	F-E
Trust	27.0(7.1)	37.3(3.2)	34.3(7.6)	33.4(6.6)	33.2(6.2)	36.3(6.1)	**0.00**	**0.00**	**0.01**	**0.03**	**0.00**	0.74	0.11	**0.03**	0.74	0.20	**0.05**	1.00	0.56	0.20	**0.05**
Safety	2.3(1.0)	4.2(1.2)	3.9(1.1)	3.5(0.3)	3.4(0.6)	4.2(0.6)	**0.00**	**0.00**	**0.00**	**0.02**	**0.00**	0.66	0.11	**0.03**	0.88	0.25	**0.05**	0.77	0.51	0.15	**0.04**
Hesitancy*	3.3(0.8)	2.2(0.4)	3.1(0.0)	2.1(1.0)	2.8(1.0)	2.1(1.4)	**0.01**	0.46	**0.00**	0.18	**0.00**	**0.04**	0.55	0.13	0.46	**0.01**	0.55	**0.01**	**0.04**	0.89	**0.03**
Cautiousness	4.9(0.6)	5.0(0.7)	4.2(0.9)	4.5(1.3)	4.8(0.9)	4.8(0.4)	0.27	**0.03**	**0.10**	1.00	1.00	**0.00**	**0.01**	0.27	0.27	0.58	**0.03**	**0.03**	0.10	0.10	1.00
Rashness	2.0(1.0)	3.4(1.2)	3.3(1.2)	3.2(1.2)	3.0(0.4)	3.6(1.2)	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.68	0.54	0.18	0.89	0.84	0.34	0.58	0.45	0.45	0.14

deemed less visible than the other interfaces (as expected), but for S this aspect was not considered detrimental to safety.

Moving to the other dimensions characterising the user experience that were investigated more in depth through the UEQ-S, as reported in Fig. 3 all the interfaces performed significantly better than B, whereas M was significantly the worst interface (although not significantly with respect to F). Analyzing separately the dimensions addressed by the questionnaire, S was considered as significantly more pragmatic than the other interfaces, except E; moreover, focusing on the AR interfaces, only E was judged better than B in this respect. Concerning the hedonic dimension, M was judged worse than both E and P, E overcame F, and P was rated better than M.

With respect to overall usability, according to SUS results all the interfaces were rated from "good" to "excellent", and no significant differences were observed ($p = 0.11$): B ($M = 84.6$, $SD = 10.2$), S ($M = 90.6$, $SD = 6.2$), P ($M = 84.0$, $SD = 17.8$), M ($M = 79.8$, $SD = 10.4$), F ($M = 79.8$, $SD = 13.4$), E ($M = 76.1$, $SD = 23.2$).

Items of the AIQ pertaining trust and safety are reported in Table III. Based on the TS, all the interfaces were found to be more trustworthy than B; furthermore, S, P, and E scored significantly better than F. The same trend was observed also for the perceived safety. An interesting aspect to analyze is hesitancy ("I felt the need to wait for the vehicle to stop before starting to cross"): interfaces that provide an instantaneous indication of the predicted stop position (M and E) scored significantly better than those that do not offer such a feedback (B, P, F). The importance of providing the VRU with some explicit feedback that the vehicle recognized him or her (not necessarily indicating the predicted stop position) is confirmed by the fact that F did not reach significance when compared with B. Surprisingly, S obtained intermediate scores for this factor, performing significantly better than both B and P (though not differently than M, F and E). Related to aspects above there is the item regarding cautiousness, intended as the participants predisposition to cross when vehicles were still far away ("I felt safe to cross when FAVs were distant"): as one could expect, P performed significantly worse than other interfaces (S, F, E, and B), whereas S scored unexpectedly better than M. Based on open feedback, this result could be due to the fact that the interface induced the participants not to look at the vehicle but just at the indications shown close to their feet. Results concerning rashness, intended as the participants predisposition to cross when vehicles were close to them, are also reported, although significance was found only for the

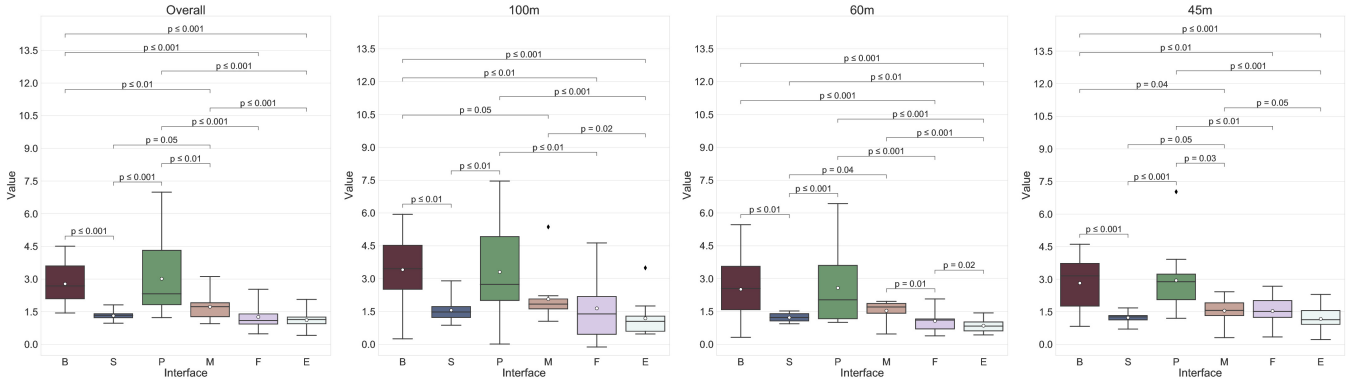


Fig. 4. Decision time (lower is better): overall, and for the various distances considered (p-values reported for significant pairwise comparisons).

comparisons with B.

Remaining aspects analyzed through subjective metrics did not showed significant differences for the various interfaces.

2) Objective Results: Data analysis on CT highlighted no significant difference among the various interfaces ($p = 0.31$). Despite previous speculations, this result is not surprising, since it may suggest that, once the VRU has concluded the negotiation with the FAV (decided to cross or not), the urgency to complete the task and the trust in the FAV are predominant on his or her behavior compared to other possible information provided by the vehicle's interface.

More insights can be obtained by focusing on the distribution of DT in Fig. 4. Considering all the distances together (overall), only P did not score significantly better than B, and both of them showed significantly higher DT values compared to the other interfaces (S, M, F, E). Furthermore, M performed significantly worse than both the AR interfaces (F, E) as well as of S, and no difference was spotted among them (S, F, E). Further considerations can be made by analyzing results for the various distances. Participants showed significantly lower DTs for M compared with B at 100m and 45m, but not at 60m. Differences with respect to other vehicle-mounted interfaces tend to reduce at higher distances. This behavior can be observed for M against S (100m) and for M against P (100m, 60m). The only interface able to outperform M at any distance was E. F was able to consistently perform better than M just at 60m; however, at that distance, F was significantly worse than E. At intermediate distances, the only interface capable of consistently offer significantly lower DTs was E. A significantly high correlation ($\rho = 0.89$, $p = 0.02$) was found between the overall DT and perceived latency (AIQ), which is a confirmation of the agreement between objective and subjective results.

Another indicator that is worth discussing is the speed at which the FAV was moving when the pedestrian started crossing the road (SAC). Even though this speed is undoubtedly linked to DT, this is also influenced by other factors, as stated in [11]. It is also more unbiased that other metrics like, e.g., the DAC, which depends on the vehicle distance class. As shown in Fig. 5, considering all the distances (overall) the rank from best to worst is as follows: E, F, S, M, B, and P. All pairwise comparisons were significant, except E-F ($p = 0.29$) and B-M ($p = 0.86$). Importantly, the fact that SAC for P was even lower than for B confirms the subjective finding concerning cautiousness. Furthermore, P is the only interface for which SACs for the three distances were not statistically different ($p = 0.09$). For all the other interfaces, SAC values are significantly lower at 100m than both at 60m and 45m ($p \leq 0.001$). Only F had significantly degraded performance at small distances (60m with respect to 45m).
Coming back to Fig. 5, digging into the behavior of SAC for the different distances it is possible to note that the B-M, B-S, and S-M pairs gain significance merely at small distances (45m). On the contrary, B-F, B-P, and F-P pairs gain significance at medium and large distances (60m and 100m). A possible interpretation for these findings could be that the closer the vehicle is, the more important is for the VRU to have a clear confirmation that it has been detected, and this relevance fades out as the distance grows. Moreover, although the trend for M-F could appear as contradictory, it could be easily explained by the fact that participants tended to take a confirmatory look at the vehicle for small distances after receiving the feedback from the road interface, hence delaying the time they actually started the crossing. Another interesting aspect is that both the AR interfaces performed better than the other interfaces (S included) at 100m and 60m; then, at 45m, F fell behind both S and E, thus providing further evidence of the importance to provide VRUs with a detection feedback, especially at small distances. A possible explanation for this result could be that AR interfaces are characterized by a higher visibility, as confirmed by a correlation analysis with PEQ subjective visibility: $\rho_{100m} = -0.89$ ($p = 0.02$), $\rho_{60m} = -0.83$ ($p = 0.04$), $\rho_{45m} = -0.77$ ($p = 0.08$). It is worth observing that also SACs at 45m were found to be significantly correlated ($\rho = 0.83$, $p = 0.04$) with the AIQ rashness item, suggesting that the latter could be a good metric for subjective observations.

Concerning aborted crossings and collisions, no significant differences were found among the various interfaces. Finally, regarding efficiency (i.e., the number of valid crossings normalized to the simulation time), Fig. 6 indicates that the most inefficient interface was P, which resulted even statistically equivalent to B. S, M, and E were found to be comparable too, whereas E appeared significantly more efficient than F.

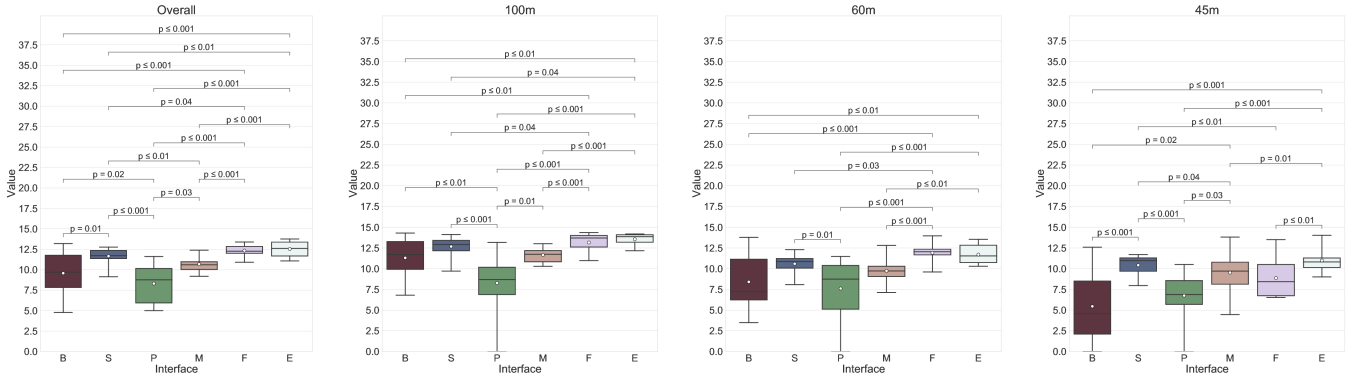


Fig. 5. Speed at crossing (higher is better): overall, and for the various distances considered (p-values reported for significant pairwise comparisons).

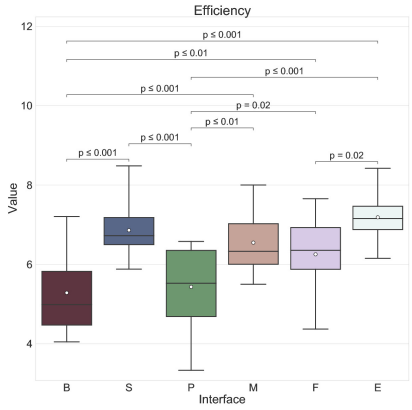


Fig. 6. Objective efficiency (higher is better): no distinction was made based on distances (p-values reported for significant pairwise comparisons).

Objective efficiency was significantly correlated with the AIQ subjective efficiency ($\rho = 0.83$, $p = 0.04$).

C. Key findings

Hereafter, the key aspects of each interface as emerging from collected measurements are summarized, by considering also open feedback collected through the interviews.

1) Smile (S): It proved overall to be one of two best interfaces. It is based on a very simple concept, which results in high efficiency and intuitiveness, and low mental effort. However, at large distances it is poorly visible (all the participants mentioned this issue); furthermore, 33% of the participants would have also preferred it to include an additional state ("frown") to signal when the FAV has seen them (horn activated) but cannot stop safely, and 17% of them found not ideal the lack of a predicted stop position indication.

2) Projected (P): It suffers as well from visibility issues; however, its main drawback is the limited efficiency, which can be attributed to its semantics. Even though it is considered very pleasant (hedonic quality), participants tend to wait for the green indication (that is showed only when the vehicle has come to a full stop); according to 50% of the participants, the red crosswalk sign is not a clear indication that the FAV actually detected the pedestrian (they suggested to use the yellow color to that purpose, mimicking traffic lights). Lastly, 92% of the participants found the vehicle-mounted LED panel useless or even did not notice it all.

3) Smart roads (M): Even though it was judged as very familiar and visible, overall it did not score well compared to other interfaces. The main negative aspect, pinpointed by 92% of the participants, was the fact that, differently than all the other interfaces, it is not visually entangled to the FAV: hence, once a change in the interface is noticed, participants tend to double-check the incoming vehicle to ensure that the interface feedback is coherent with vehicle dynamics (i.e., it is braking). Moreover, 33% of the participants would have preferred an additional state indicating whether the vehicle was successfully connected to and communicating with the smart road, in order to distinguish the case of faulty pedestrian detection from faulty connections (which were not simulated).

4) Safe roads (F): Although, based on objective metrics, it scored similarly to the extended version, from the subjective viewpoint it was often rated worse than the other interfaces. This result could be explained by the fact that, being the two AR interfaces very similar to each other and at the same time very different from all the others, participants may have faced psychological biases, such as the Weber–Fechner law [43] and the distinction bias [44], making them overestimate the actual differences. The main objective difference was in the SAC at small distances, which was a clear indication of the greater safety provided by the extended version compared to the original design. By considering also poor ratings concerning mental workload, this interface is not particularly appealing compared to considered alternatives. Notwithstanding, 50% of the participants expressed their appreciation for the hints provided by the green/red regions drawn on the road.

5) Safe roads extended (E): It was considered as one of the two best interfaces, although more complex, mentally demanding, and less intuitive compared to the first design. Participants appreciated in particular its high visibility and the visual connection to the vehicle, as well as the indication used to simultaneously provide the pedestrian detection feedback and the predicted stop position. It was the only interface deemed capable of fostering trust at any distance. It is worth remarking that 92% of the participants would have preferred fewer indications (only one participant would have liked more information), and 50% of them considered not fundamental the arrows showing the vehicle's dynamics (yellow and red);

17% of the participants found the red tick useful, but not the yellow arrow (as it provides a duplicated information).

D. Considerations and remarks

It is worth observing that, despite the efforts put in carrying out a fair and representative comparison, results and commentary reported above can be considered as valid only for the scenario configured for the experiments. Different settings, encompassing, e.g., a multi-lane and/or two-way traffic could introduce important challenges for the considered interfaces, which would require further investigations. For instance, one could validate speculations on the fact that performance of interfaces providing at a glance a feedback about safe zones for the whole road width (like M, F, E) may be boosted in such a scenario. Similarly, one could study whether having multiple agents (either avatars controlled by the simulation or other user-controlled pedestrians) crossing the road simultaneously to the user-controlled pedestrian may lead to different communication patterns between the FAVs and the VRUs. In fact, this configuration could be extremely penalizing for interfaces conceived only for one-to-one communication like, e.g., S (as a VRU could hardy determine whether the smile was directed to him or her or to other agents). However, S could be possibly implemented in practice not as a vehicle-mounted one but as an AR interface: thus, the design would become similar to experimented solutions in which each VRU gets a different feedback for the same vehicle. In this case, future experiments should focus on digging into possible impact of current technological limitations of wearable AR devices, e.g., related to their limited field of view. Still concerning the evaluation perspective, it shall be observed that the sample size of the user study reported in this work and its characteristics (e.g., the fact that participants were all Italians and accustomed to cross in unregulated conditions, etc.) may not be fully representative of all the potential end-user categories. Hence, future works should also consider cultural factors and personal behaviors of study participants, since it can be easily expected that they could have a non-negligible impact on users' preferences.

VI. Conclusions

In this paper, a careful selection of the most promising state-of-the-art and newly proposed interfaces for FAV-to-VRU interaction were compared through a user study by considering a common experimental scenario represented by pedestrian crossing. Comparison was performed in an immersive VE, in which a single, user-controlled VRU was requested to cross a one-way road under non-regulated conditions.

Results obtained using subjective and objective metrics outlined the importance to provide the users with a clear feedback about the pedestrian detection process, and to ensure high visibility. Moreover, the study proved the potential of AR-based interfaces in supporting effective vehicle-to-pedestrian communication (and, to the best of the authors' knowledge, represents the most extensive analysis of such interfaces). In particular, one of the newly proposed AR-based interfaces (namely, E) outperformed the other designs for what it concerns the above requirements; it was also characterized by a higher efficiency, and it was the only interface to be judged as capable of inducing in the users a high sense of safety independently of the vehicle distance. However, state-of-the-art designs leveraging anthropomorphic features displayed on vehicle-mounted LED panels (like S) proved to be characterized by a lower cognitive effort and a higher intuitiveness (and ease of use, in general). Despite these findings, none of the considered interfaces stood out for all the analyzed dimensions. Nevertheless, the outcomes of this study provide precious indications that could be used to shape interaction paradigms and technologies of future FAVs ecosystem.

Acknowledgements

The authors would like to thank Edoardo Demuru for his contribution on the system implementation. This research was partially supported by the VR@POLITO initiative.

References

[1] Y. Qiao, Y. Cheng, J. Yang, J. Liu, and N. Kato, "A mobility analytical framework for big mobile data in densely populated area," *IEEE Trans. on Vehic. Techn.*, vol. 66, no. 2, pp. 1443–1455, 2016.

[2] "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," in *SAE Technical Paper*, 2018.

[3] Y. Xun, J. Liu, N. Kato, Y. Fang, and Y. Zhang, "Automobile driver fingerprinting: A new machine learning based authentication scheme," *IEEE Trans. on Ind. Informatics*, vol. 16, no. 2, pp. 1417–1426, 2019.

[4] J. Wang, J. Liu, and N. Kato, "Networking and communications in autonomous driving: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1243–1274, 2018.

[5] L. Morra, F. Lamberti, F. G. Pratticó, S. La Rosa, and P. Montuschi, "Building trust in autonomous vehicles: Role of virtual reality driving simulators in HMI design," *IEEE Trans. on Vehic. Techn.*, vol. 68, no. 10, 2019.

[6] A. Rasouli and J. K. Tsotsos, "Autonomous vehicles that interact with pedestrians: A survey of theory and practice," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 900–918, 2019.

[7] C.-M. Chang, K. Toda, D. Sakamoto, and T. Igarashi, "Eyes on a car: An interface design for communication between an autonomous car and a pedestrian," in *Proc. 9th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications*, 2017, pp. 65–73.

[8] Y. Li, M. Dikmen, T. G. Hussein, Y. Wang, and C. Burns, "To cross or not to cross: Urgency-based external warning displays on autonomous vehicles to improve pedestrian crossing safety," in *Proc. of 10th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications*, 2018, pp. 188–197.

[9] A. Habibovic *et al.*, "Communicating intent of automated vehicles to pedestrians," *Frontiers in Psychology*, vol. 9, 2018.

[10] T. Lagström and V. M. Lundgren, "AVIP – autonomous vehicles' interaction with pedestrians – An investigation of pedestrian-driver communication and development of a vehicle external interface," 2016.

[11] A. Löcken, C. Golling, and A. Riener, "How should automated vehicles interact with pedestrians? a comparative analysis of interaction concepts in virtual reality," in *Proc. 11th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications*, 2019, pp. 262–274.

[12] T. T. Nguyen, K. Holländer, M. Hoggenmueller, C. Parker, and M. Tomitsch, "Designing for projection-based communication between autonomous vehicles and pedestrians," in *Proc. 11th Int. Conf. on Automotive User Int. and Interactive Vehic. Appl.*, 2019, pp. 284–294.

[13] S. Kitayama, T. Kondou, H. Ohyabu, M. Hirose, H. Narihiro, and R. Maeda, "Display system for vehicle to pedestrian communication," in *SAE Technical Paper*, 2017.

[14] E. Florentine, M. A. Ang, S. D. Pendleton, H. Andersen, and M. H. Ang Jr, "Pedestrian notification methods in autonomous vehicles for multi-class mobility-on-demand service," in *Proc. 4th Int. Conf. on Human Agent Interaction*, 2016, pp. 387–392.

[15] Y. M. Lee *et al.*, "Understanding the messages conveyed by automated vehicles," in *Proc. 11th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications*, 2019, pp. 134–143.

[16] L. Graziano, "AutonoMI autonomous mobility interface," 2014.

[17] M.-P. Böckle, A. P. Brenden, M. Klingegård, A. Habibovic, and M. Bout, "SAV2P: Exploring the impact of an interface for shared automated vehicles on pedestrians' experience," in *Proc. of 9th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications*, 2017, pp. 136–140.
[18] S. Deb, L. J. Strawderman, and D. W. Carruth, "Investigating pedestrian suggestions for external features on fully autonomous vehicles: A virtual reality experiment," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 59, 2018.
[19] M. Matthews, G. Chowdhary, and E. Kieson, "Intent communication between autonomous vehicles and pedestrians," 2017. [Online]. Available: https://arxiv.org/abs/1708.07123
[20] K. De Clercq, A. Dietrich, J. P. Núñez Velasco, J. de Winter, and R. Happee, "External human-machine interfaces on automated vehicles: Effects on pedestrian crossing decisions," *Human Factors*, vol. 61, no. 8, 2019.
[21] "Drive.ai solves autonomous cars' communication problem." [Online]. Available: https://tinyurl.com/ycjw4ro2
[22] U. Gruenefeld, S. Weiß, A. Löcken, I. Virgilio, A. L. Kun, and S. Boll, "VRoad: Gesture-based interaction between pedestrians and automated vehicles in virtual reality," in *Proc. 11th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*, 2019.
[23] C. G. Burns, L. Oliveira, P. Thomas, S. Iyer, and S. Birrell, "Pedestrian decision-making responses to external human-machine interface designs for autonomous vehicles," in *IEEE Intelligent Vehicles Symp.*, 2019.
[24] J. Mairs, "Umbrellium develops interactive road crossing that only appears when needed," 2017. [Online]. Available: https://tinyurl.com/y4vng5p8
[25] K. Mahadevan, S. Somanath, and E. Sharlin, "Communicating awareness and intent in autonomous vehicle-pedestrian interaction," in *Proc. CHI Conf. on Human Factors in Computing Systems*, 2018, pp. 1–12.
[26] H. Nishiyama, T. Ngo, S. Oiyama, and N. Kato, "Relay by smart device: Innovative communications for efficient information sharing among vehicles and pedestrians," *IEEE Vehicular Technology Magazine*, vol. 10, no. 4, pp. 54–62, 2015.
[27] L. Cancedda, A. Cannavò, G. Garofalo, F. Lamberti, P. Montuschi, and G. Paravati, "Mixed Reality-based user interaction feedback for a hand-controlled interface targeted to robot teleoperation," in *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, 2017, pp. 447–463.
[28] M. Hesenius, I. Börsting, O. Meyer, and V. Gruhn, "Don't panic! Guiding pedestrians in autonomous traffic with augmented reality," in *Proc. 20th Int. Conf. on HCI with Mobile Devices and Services*, 2018.
[29] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "Airsim: High-fidelity visual and physical simulation for autonomous vehicles," in *Field and Service Robotics*, 2017. [Online]. Available: https://arxiv.org/abs/1705.05065
[30] A. Pillai, "Virtual reality based study to analyse pedestrian attitude towards autonomous vehicles," 2017.
[31] G. Johansson and K. Rumar, "Drivers' brake reaction times," *Human Factors*, vol. 13, no. 1, pp. 23–27, 1971.
[32] I. Doric *et al.*, "A novel approach for researching crossing behavior and risk acceptance: The pedestrian simulator," in *Proc. 8th Int. Conf. on Automotive User Int. and Interactive Vehic. Appl.*, 2016, pp. 39–44.
[33] M. Beggiato, C. Witzlack, and J. F. Krems, "Gap acceptance and time-to-arrival estimates as basis for informal communication between pedestrians and vehicles," in *Proc. 9th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications*, 2017, pp. 50–57.
[34] A. Cannavò, D. Calandra, F. G. Pratticò, V. Gatteschi, and F. Lamberti, "An evaluation testbed for locomotion in virtual reality," *IEEE Trans. Vis. Comput. Graphics*, p. (in press), 2020.
[35] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The Int. Journal of Aviat. Psychology*, vol. 3, 1993.
[36] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *Int. J. Cognitive Ergonomics*, vol. 4, no. 1, pp. 53–71, 2000.
[37] J. Brooke, "SUS – A quick and dirty usability scale," *Usability Evaluation in Industry*, 1996.
[38] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Advances in Psychology*, 1988, vol. 52, pp. 139–183.
[39] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Design and evaluation of a short version of the User Experience Questionnaire (UEQ-S)," *Int. J. Interact. Multimedia Artif. Int.*, vol. 4, no. 6, pp. 103–108, 2017.
[40] R. S. Kalawsky, "VRUSE – A computerised diagnostic tool for usability evaluation of virtual/synthetic environment systems," *Applied Ergonomics*, vol. 30, no. 1, pp. 11–25, 1999.
[41] T. W. Schubert, "The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness." *Zeitschrift für Medienpsychologie*, vol. 15, no. 2, pp. 69–71, 2003.
[42] G. Erdélyi, L. Piras, and J. Rothe, "Bucklin voting is broadly resistant to control," 2010. [Online]. Available: https://arxiv.org/abs/1005.4115
[43] S. Dehaene, "The neural basis of the Weber–Fechner law: A logarithmic mental number line," *Trends in Cognitive Sciences*, vol. 7, no. 4, pp. 145–147, 2003.
[44] C. K. Hsee and J. Zhang, "Distinction bias: Misprediction and mischoice due to joint evaluation." *Journal of Personality and Social Psychology*, vol. 86, no. 5, p. 680, 2004.



F. Gabriele Pratticò received the M.Sc. degree in computer engineering from Politecnico di Torino, Turin, Italy, in 2017. Currently, he is a Ph.D. student at the Dipartimento di Automatica e Informatica of Politecnico di Torino, where he carries out research in the areas of extended reality, human-machine interaction, educational and training systems, and user experience design.



Fabrizio Lamberti is a Full Professor with the Dipartimento di Automatica e Informatica of Politecnico di Torino, Turin, Italy, where he has the responsibility for the VR@POLITO hub. His research interests include computer graphics, human-machine interaction, and intelligent computing. He is serving as an Associate Editor for IEEE Transactions on Computers, IEEE Transactions on Learning Technologies, IEEE Transactions on Consumer Electronics, IEEE Consumer Electronics Magazine, and the International Journal of Human-Computer Studies.



Alberto Cannavò received the B.Sc. degree from University of Messina, Italy, in 2013. Then, he received the M.Sc. and the Ph.D. degrees in computer engineering from Politecnico di Torino, Italy, in 2015 and 2020, respectively. Currently, he is a Postdoctoral Fellow at the Dipartimento di Automatica e Informatica of Politecnico di Torino. His fields of interest include computer graphics and human-machine interaction.



Lia Morra received the M.Sc. and Ph.D. degrees in computer engineering from Politecnico di Torino, Turin, Italy, in 2002 and 2006, respectively. She is currently a Senior Postdoctoral Fellow with the Dipartimento di Automatica e Informatica of Politecnico di Torino. Her research interests include computer vision, pattern recognition, and machine learning.



Paolo Montuschi is a Full Professor with the Dipartimento di Automatica e Informatica and a Member of the Board of Governors of Politecnico di Torino, Italy. His research interests include computer arithmetic, computer graphics, and intelligent systems. He is serving as 2019 Acting (interim) Editor-in-Chief of the IEEE Transactions on Emerging Topics in Computing and as the 20172019 IEEE Computer Society Awards Chair. He is a Life Member of the International Academy of Sciences of Turin and of IEEE Eta Kappa Nu.