

Research Article

I See Your Gesture: A VR-Based Study of Bidirectional Communication between Pedestrians and Automated Vehicles

Michael R. Epke , Lars Kooijman , and Joost C. F. de Winter 

Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, Netherlands

Correspondence should be addressed to Joost C. F. de Winter; j.c.f.dewinter@tudelft.nl

Received 20 January 2021; Revised 22 March 2021; Accepted 5 April 2021; Published 27 April 2021

Academic Editor: Nirajan Shiwakoti

Copyright © 2021 Michael R. Epke et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Automated vehicles (AVs) are able to detect pedestrians reliably but still have difficulty in predicting pedestrians' intentions from their implicit body language. This study examined the effects of using explicit hand gestures and receptive external human-machine interfaces (eHMIs) in the interaction between pedestrians and AVs. Twenty-six participants interacted with AVs in a virtual environment while wearing a head-mounted display. The participants' movements in the virtual environment were visualized using a motion-tracking suit. The first independent variable was the participants' opportunity to use a hand gesture to increase the probability that the AV would stop for them. The second independent variable was the AV's response "I SEE YOU," displayed on an eHMI when the vehicle yielded. Accordingly, one-way communication (gesture or eHMI) and two-way communication (gesture and eHMI combined) were investigated. The results showed that the participants decided to use hand gestures in 70% of the trials. Furthermore, the eHMI improved the predictability of the AV's behavior compared to no eHMI, as inferred from self-reports and hand-use behavior. A postexperiment questionnaire indicated that two-way communication was the most preferred condition and that the eHMI alone was more preferred than the gesture alone. The results further indicate limitations of hand gestures regarding false-positive detection and confusion if the AV decides not to yield. It is concluded that bidirectional human-robot communication has considerable potential.

1. Introduction

In current traffic, pedestrians and drivers use hand gestures and other bodily signals to inform, acknowledge, draw attention, or clarify situations [1–3]. These communication modes will no longer be available when automated vehicles (AVs) have taken over the driving task. Keferböck and Riener [4] argued that it is important to substitute today's pedestrian-vehicle communication with AVs that can detect pedestrians' gestures and actively communicate via external human-machine interfaces (eHMIs).

So far, a large number of studies have examined the effectiveness of eHMIs that display the AV's state and intentions (for reviews, see [5–7]); for example, a VR-based study by De Clercq et al. [8] found that pedestrians feel safer to cross in front of an AV with eHMI (e.g., text or front brake

light) compared to without eHMI. A variety of other studies also show that eHMIs provide performance improvement or enhance subjective clarity for pedestrians relative to control conditions without eHMI [9–13].

Although eHMIs have demonstrated their value in various experimental studies, it can be questioned whether the solution to the interaction between AVs and pedestrians should be sought in eHMIs. eHMIs have several drawbacks. In some concepts, the eHMI covers a large surface area of the AV [14, 15], which would entail technical complexity and high cost. Another point is that, in real traffic, the eHMI may not be noticed or understood [16].

In addition, it may be questioned whether eHMIs are the only way to achieve communication between pedestrians and AVs. Schieben et al. [17] provided a framework for communication between AVs and other road users and showed that various forms of interaction are conceivable, of

which eHMIs are one. Other communication strategies include the use of the infrastructure, the design of the vehicle shape, and the AV movements themselves (the latter is also known as implicit communication [18–22]). Moreover, the paper by Schieben et al. [17] makes clear that eHMIs do not necessarily have to show the AV's state and intentions. Informing other road users about the AV's perception of the environment and the AV's cooperation capabilities is a fruitful alternative. The current paper proposes a communication strategy that places responsibility on both the pedestrian and the AV. More specifically, we investigate whether pedestrians prefer to make their intention clear using an explicit gesture, and whether the AV should be made more intelligent by recognizing and responding to this gesture.

Current AVs are already capable of detecting pedestrians and other vulnerable road users. However, an ongoing difficulty remains in the prediction of pedestrians' intentions [23, 24]. Vinkhuyzen and Cefkin [25] noted that a "challenge . . . is the limitations of the technology in making observational distinctions that socially acceptable driving necessitates." Because AVs may have difficulty in reading pedestrians' natural body language that may signal crossing intent, it may be necessary to require pedestrians to use more explicit bodily communication, such as a hand gesture. It can be expected that the camera systems of future AVs will be able to detect a hand gesture.

The use of driver gestures has previously been studied for the control of in-vehicle information systems [26, 27] and maneuver-based AV control [28, 29]. It has been found that hand gesture usage is effective for letting manually driven vehicles stop for pedestrians [30]. However, to the best of our knowledge, no human factor studies have examined pedestrians expressing their intentions towards AVs through hand gestures.

The present study aimed to examine how pedestrians experience the use of hand gestures to increase the probability that the AV will stop for them. In addition to pedestrian-to-AV gesturing, this study examined how a subsequent response from the AV via its eHMI affects the pedestrians' experience. In real-life applications, our hand gesture concept would require the AV computer vision systems to recognize these gestures. Various studies have already been performed in this area, such as the detection of gestures made by police officers [31–33] or cyclists [34]. In the present lab-based study, we used a motion suit [9], which allowed us to measure the pedestrians' bodily state.

2. Materials and Methods

2.1. Participants. Twenty-six participants (4 females and 22 males) were recruited among students and PhD candidates at the TU Delft. They had a mean age of 26.0 years ($SD = 3.7$ years). All participants were living in the Netherlands at the time of the study but had nationalities from different parts of the world (i.e., Europe, Asia, North America, South America, and Africa). The participants were offered a compensation of €10 for participating in the study and signed a written informed consent form before starting the experiment. The

research was approved by the Human Research Ethics Committee of the TU Delft.

3. Materials

The experiment was conducted on a desktop computer running Windows 10 64-bit platform, from the brand Alienware, with Intel (R) Core (TM) i7-9700K CPU @ 3.60 GHz, NVIDIA GeForce RTX 2080 8 GB Graphics Card, 16 GB of RAM. The virtual environment was developed and run using Unity version 2018.4.6f1. The scripts and environment were adapted from De Clercq et al. [8] and Kooijman et al. [9]. The participants wore an Oculus Rift CV1 to experience the virtual environment. An Xsens Link motion suit was used to record the participant's body movements, which were mapped onto an avatar in the virtual environment. By means of the Oculus Rift and motion suit, the participants could look and walk around while being able to see their own body from a first-person perspective [9]. The motion suit was connected to a transmitter via which the data were sent to the desktop computer. Data received from the Xsens Link were handled by the software Xsens MVN Analyze Version 2019.0.0 Build 1627.

3.1. Design. The experiment was of a within-subject design, with the following independent variables:

- (1) Opportunity to use hand gestures (two levels: yes and no).
- (2) The eHMI message "I SEE YOU" when the AV yielded (two levels: eHMI upon yielding of the AV and no eHMI upon yielding of the AV).
- (3) The yielding behavior of the approaching AV (two levels: yielding [if the participant makes a gesture] and no yielding).

The first two independent variables formed a total of four conditions that were offered in blocks of ten trials. The third independent variable was varied within these blocks and was contingent on the participant's use of hand gestures. More specifically, participants experienced ten trials per condition, five of which involved an AV that yielded if the participant used a hand gesture and five of which involved an AV that never yielded. The four blocks of 10 trials were randomized per participant.

The four conditions were as follows:

- (i) Baseline, in which no hand gesture was allowed to be used by the participant and the vehicle would not respond via the eHMI. Even if the participant did use a hand gesture, the AV would not yield in response to the hand gesture. The AV yielded in a random 5 of the 10 trials; in the other 5 trials, the AV did not yield.
- (ii) eHMI, which was identical to Baseline but with the addition that the AV displayed "I SEE YOU" on its eHMI when the AV started to decelerate.



FIGURE 1: VR environment with AV displaying “I SEE YOU.” The participants’ body was visualized using an avatar. During the experiment, participants could see the avatar (e.g., hands and feet) from a first-person perspective while wearing the head-mounted display.

- (iii) Hand, in which a hand gesture was allowed to be used by the participant. By using a hand gesture, the participant could increase the likelihood that the approaching vehicles would stop for them. The hand gesture would only result in yielding if the AV was programmed to yield during that specific trial. If the hand gesture was used during a trial in which the AV was not programmed to yield, the AV would maintain speed without yielding. Thus, if the participant would raise their hand in all 10 trials, the AV would yield in 5 trials. If the participant never raised their hand, the AV would yield in 0 of the 10 trials. The eHMI was off in all trials.
- (iv) Combination, which was identical to the Hand condition but with the addition that the AV would respond by displaying “I SEE YOU” on its eHMI (see Figure 1) when the AV started to decelerate.

In summary, Baseline was identical to eHMI, and Hand was identical to Combination, except for the AV’s acknowledgment “I SEE YOU” in the eHMI and Combination conditions. In the Hand and Combination conditions, the participant could use a hand gesture to let the AV yield in 5 of 10 trials. In the Baseline and eHMI conditions, the AV always yielded in those 5 trials. Each participant interacted with a virtual AV in a total of 40 trials.

The AV had a constant approach speed of 50 km/h during all trials. In the Baseline and eHMI conditions, yielding AVs started to decelerate 50 m from the pedestrian and came to a standstill about 7 m from the pedestrian (the distance measured towards the center of the AV, and parallel to the road).

In all trials, the target vehicle was preceded by a lead vehicle that always maintained speed. Upon approach, the time gap between the lead vehicle and the target vehicle was 1.3 s. Figure 2 shows the distance between the AVs and the pedestrian, distinguishing between yielding and nonyielding target vehicles.

As noted above, participants were permitted to use a hand gesture in two of the four conditions (i.e., Hand and Combination). Hand gesture use was not obligatory, to test if the participants were willing to adopt this novel communication mode. The distance thresholds for the hand gesture were 30 and 50 meters from the pedestrian’s location. A hand gesture used while the AV was before the 50 m threshold or after the 30 m threshold would not cause the AV to yield. If the hand was raised while the AV was between these distance thresholds, the eHMI would turn on (in the Combination condition) and the AV would immediately start the deceleration (in the Hand and Combination conditions). If the AV yielded, it did so with a deceleration that depended on the distance to the pedestrian crossing and in such a way that the AV would come to a standstill before the pedestrian crossing.

Hand usage (a binary variable: no or yes) was identified for each simulation timestep based on whether one of the positions or angles of either arm would change to above a threshold value. The participants were not informed about these thresholds. The threshold values were predetermined during the development and pilot testing of the experiment. More specifically, hand usage was operationalized based on whether either of the following criteria were met:

- (1) The angle between the upper arm and the participant’s body (defined as an upright vector) was greater than 45° . Note that an angle of 0° would correspond to the upper arm hanging down towards the ground, an angle of 90° would correspond to the participant elevating his upper arm horizontally, and an angle of 180° would correspond to the participant pointing his upper arm to the ceiling. An angle of 45° was regarded as indicative of the fact that the participant had raised his arm.
- (2) The angle between the forearm and the direction of the upper arm was greater than 60° . An angle of 0° would correspond to a fully stretched arm. An angle of 60° was regarded as indicative of a bent arm.

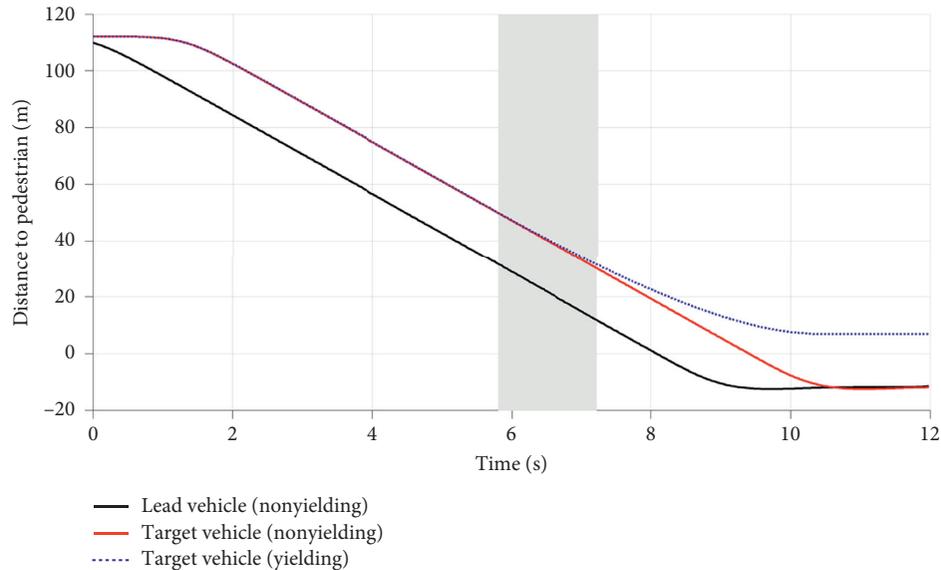


FIGURE 2: The approaching vehicles' distance to pedestrian (distance measured parallel to the road) in the Baseline and eHMI conditions. Vehicles approached from around a corner at about 115 m distance and turned left at a corner after passing the participant (pedestrian distance 12 m). The grey area represents the region where the AV could detect hand gestures in the Hand and Combination conditions (i.e., distance to pedestrian between 30 and 50 m).

- (3) The position of the hand was higher above the ground than the position of the elbow. When standing or walking, the height above the ground of the hands can be expected to be lower than that of the elbow. If the hand was higher than the elbow, this was regarded as a raised arm.

3.2. Procedure. Before starting the experiment, participants read and signed an informed consent form. The form mentioned that participants would be encountering 40 trials divided into four blocks of 10 trials. It was further mentioned that, during some parts of the experiment, the researcher would inform them that they could raise their hand to communicate to the AV that they want to cross. Participants were informed that each trial consisted of two AVs driving towards the pedestrian crossing. Participants were instructed to let the first AV pass and to make a step forward when they thought was a good time to do so. They could express this crossing intention before or after the target vehicle had passed.

After signing the form, participants completed a digital preexperiment questionnaire consisting of several demographic questions and four Likert-scale questions related to trust in automated vehicles and hand gestures.

During the experiment, the participants stood on the curb in front of a pedestrian crossing. The participants were instructed not to cross the road but only to make one step forward when they felt safe to cross the road. In this way, the participants had the task of making a crossing decision and were not merely observers of the approaching cars. Participants were asked not to express their crossing intention before the first AV (i.e., lead vehicle) had passed.

Before the Hand and Combination conditions, the participant was informed by the researcher as follows: "for the following ten interactions, you are allowed to use a hand

gesture if you want to, but you do not have to." Participants were told that the gesture involved raising their hand to show their intention to cross the road.

After each trial, the participant was asked two questions: "on a scale from 0 to 10, how difficult was it for you to predict the behavior of the car, where 0 is not difficult at all, and 10 is very difficult?" and "on a scale from 0 to 10, how sure were you that the car would see you, where 0 is not sure at all and 10 is completely sure?"

The participants ended the experiment with a post-experiment questionnaire containing the same trust-related questions used in the preexperiment questionnaire. The postexperiment questionnaire also asked participants to rank the four conditions depicted as follows:

- (i) *Baseline*. "No communication" accompanied by a screenshot of the car without eHMI.
- (ii) *eHMI*. "Communication via eHMI" accompanied by a screenshot of the car with the eHMI depicting "I SEE YOU."
- (iii) *Hand*. "No communication after hand gesture" accompanied by a screenshot of the car without eHMI.
- (iv) *Combination*. "Communication via eHMI after hand gesture," accompanied by a screenshot of the car with the eHMI depicting "I SEE YOU."

The participants completed the ranking four times: (1) based on the extent to which they felt safe to cross the road; (2) based on the extent to which they were sure the car had seen them; (3) based on the extent to which they believed their decision was affected by the fact that no eye contact with the driver was possible; (4) based on their general preference when interacting with AVs.

3.3. *Analyses.* Analyses were performed for the following variables:

- (i) Hand gesture usage, defined as whether the hand was raised in the 30–50 m distance interval.
- (ii) Hand-release time, defined as the first moment the hand was released after it had been raised. The hand-release time was expressed from the moment the AV had passed the 50 m distance threshold. In case the hand was never raised, no hand-release time was determined for that trial.
- (iii) Responses to the posttrial questions (difficulty and sureness, on a scale of 0 to 10).
- (iv) Responses to the postexperiment questions.

Statistical comparisons were performed by judging nonoverlapping confidence intervals, which were computed using a method for within-subject designs [35]. Furthermore, paired-samples *t*-tests were used to compare participants' scores between the experimental conditions. An alpha value of 0.05 was used.

4. Results

Participants performed a total of 1040 trials (26 participants \times 40 trials per participant). From those 1040 trials, 16 trials were discarded due to anomalies in the experiment or data recording.

Figure 3 shows the percentage of trials with hand gesture usage per condition. In the two conditions where a hand gesture was allowed (Hand and Combination), the participants used a hand gesture on average in, respectively, 72.1% ($SD = 24.6\%$) and 68.7% ($SD = 26.0\%$) of the trials. Figure 3 also shows that there were false positives for several participants in the Baseline and eHMI conditions.

Figure 4 shows at which moment the participants made hand gestures. In about 50% of the trials, participants had their hand raised already before the hand gestures could be picked up by the AV (i.e., distance between AV and pedestrian >50 m). Participants lowered their hand as the AV started to decelerate (Figure 4, top), especially for the Combination condition. Statistical analysis showed that the mean hand-release time since passing the 50 m mark at 5.80 s was 3.66 s ($SD = 1.46$ s) for the Hand condition and 2.85 s ($SD = 1.54$ s) for the Combination condition, a significant effect according to a paired-samples *t*-test, $t(22) = 3.68$, $p = 0.001$ (3 participants did not produce hand-release data because they did not raise their hand).

In comparison, the difference in hand use was not seen in the nonyielding trials, in which the eHMI was always off (Figure 4, bottom). More specifically, the mean hand-release time since passing the 50 m mark at 5.80 s was 2.91 s ($SD = 1.25$ s) for the Hand condition and 2.76 s ($SD = 1.09$ s) for the Combination condition, a nonsignificant effect, $t(24) = 0.64$, $p = 0.526$ (1 participant did not produce hand-release data because the hand was not raised).

Figure 5 shows the mean responses of participants' difficulty in predicting the AV's behavior and the sureness of being seen by the AV. A distinction is made between trials in

which the AV did and did not yield. The eHMI made it easier for participants to predict the AV's behavior (Figure 5(a)) and assured them of being seen (Figure 5(b)). As can be seen from the nonoverlapping confidence intervals, the effects were generally statistically significant; for example, for yielding AVs, a paired-samples *t*-test between the Baseline and Hand conditions' difficulty scores indicated a significant effect: $t(25) = 3.94$, $p < 0.001$.

The results further showed that, in the case of nonyielding AVs, the Hand condition made it more difficult for participants to predict the AV's behavior compared to the Baseline condition, a significant difference according to a paired-samples *t*-test, $t(25) = 2.96$, $p = 0.007$.

The results of the pre- and postexperiment questionnaire regarding trust in AVs are provided in Table 1. It can be seen that, after the experiment, participants exhibited a higher trust in the idea that hand gestures can be used to interact with AVs as compared to that before the experiment. Before the experiment, participants were skeptical towards the idea that self-driving vehicles will respond to hand gestures ($M = 5.15$ on a scale from 1 to 10).

The postexperiment ranking of the four experimental conditions is provided in Table 2. The Combination condition was ranked highest regarding safety, the sureness of being seen, and general preference. This was followed by the eHMI condition, Hand condition, and Baseline condition. The eHMI condition received higher rankings than the Hand condition. In other words, hand gestures alone (i.e., without receptive eHMI) were not highly rated. The presence of the eHMI, in the eHMI condition as well as in the Combination condition, appeared to alleviate the perceived effect of the lack of eye contact.

5. Discussion

In the present study, we examined the efficacy of one-way and two-way communication in pedestrian-AV interactions in a VR environment. The experiment results showed that pedestrians' hand gesture use was moderately high, at about 70%. In other words, most participants were willing to use a hand gesture to make the AV stop. Possibly, some participants did not want to take the effort to raise their hand since no benefit could be obtained by doing so. In comparison, in real traffic, pedestrians may gain time or increase their safety if an approaching vehicle stops for them. It is also possible that participants in our experiment were trying out the use and nonuse of hand gestures.

The experiment further showed that the eHMI, which provided the confirmatory message "I SEE YOU," improved the perceived predictability of the AV's behavior as compared to no eHMI, as demonstrated by the relatively sharp decline in the percentage of hand gesture usage for the Combination condition in Figure 3. The latter effect can be explained by the fact that the eHMI turned on if and only if the vehicle yielded. The eHMI further caused pedestrians to lower their hands early (see results for the Combination condition vs. the Hand condition). In other words, the eHMI's affirmative message prevented pedestrians from holding their hand in the air for an unnecessarily long time.

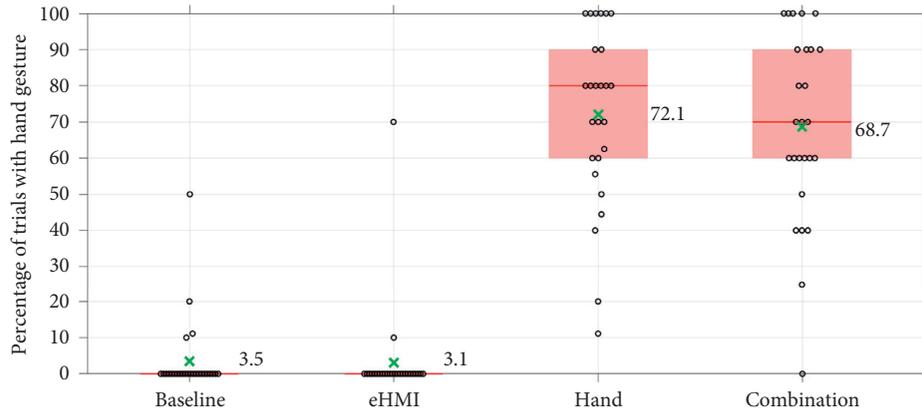


FIGURE 3: Boxplot depicting the percentage of trials with gesture usage per condition. Each marker represents a participant. The box runs from the 25th to the 75th percentile. The horizontal red line represents the median. The green “x” and reported number next to it correspond to the mean of participants.

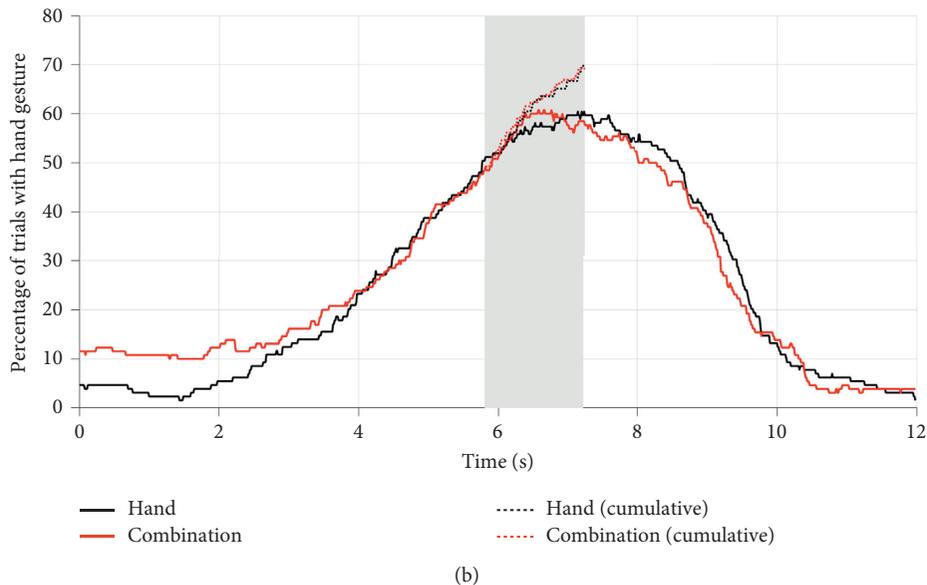
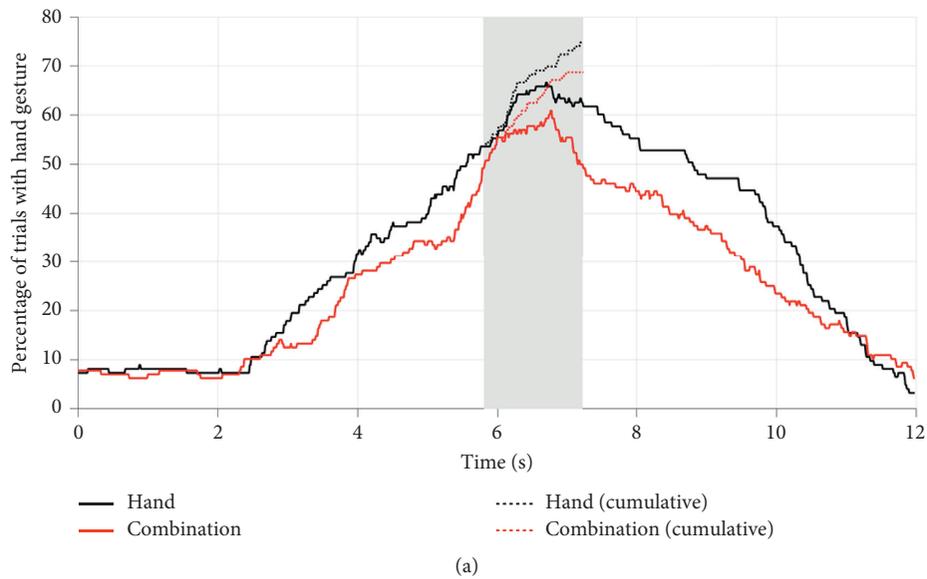


FIGURE 4: Percentage of trials in the Hand and Combination conditions in which participants had their hand raised, as a function of elapsed time. (a) Only trials in which the AV could yield are considered ($n = 123$ for the Hand condition and $n = 128$ for the Combination condition). (b) Only trials in which the AV did not yield are considered ($n = 129$ for the Hand condition and $n = 130$ for the Combination condition). The dotted lines represent the percentage of trials in which the hand gesture was used up to that point. The grey area represents the region in which hand gestures could be picked up by the AV (distance to pedestrian between 30 and 50 m).

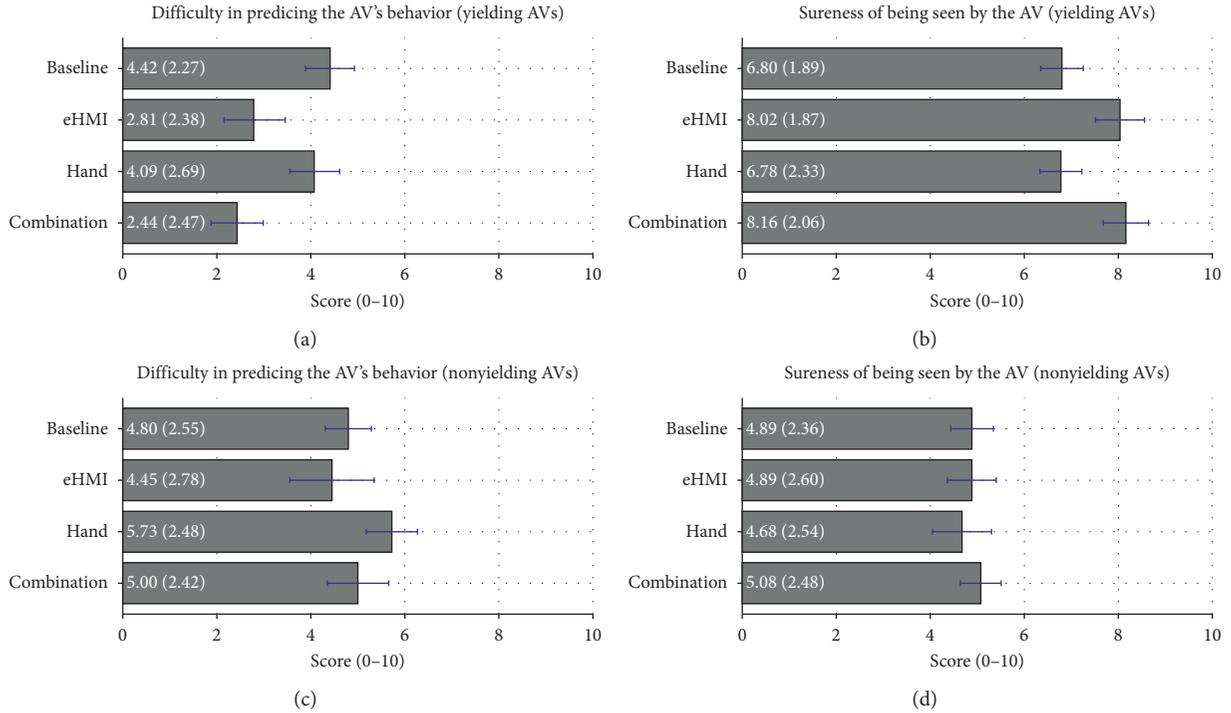


FIGURE 5: Mean responses to posttrial questions, distinguishing between yielding and nonyielding AVs. Error bars represent 95% confidence intervals (computed according to [35]). Also shown are the means (standard deviations across participants in parentheses) in numeric form. $n = 26$, except for the Combination condition in *a* and *b*, where $n = 25$, because 1 participant never raised the hand.

TABLE 1: Responses to pre- and postexperiment questions regarding trust in automated vehicles and hand gestures (1 = strongly disagree and 10 = strongly agree) ($n = 26$).

Question	Preexperiment	Postexperiment	Paired t -test
1. I have trust in self-driving vehicles	6.04 (1.95)	6.58 (1.72)	$t(25) = -1.59, p = 0.124$
2. I have trust that self-driving vehicles will notice me	6.19 (1.86)	6.27 (1.71)	$t(25) = -0.19, p = 0.849$
3. I have trust that self-driving vehicles will respond to hand gestures	5.15 (2.26)	6.77 (2.03)	$t(25) = -2.87, p = 0.008$
4. I have trust that hand gestures can be used to communicate with a self-driving vehicle	6.04 (2.44)	7.62 (1.92)	$t(25) = -3.12, p = 0.005$

TABLE 2: Mean rank when sorting the four experimental conditions ($n = 26$).

	Baseline	eHMI	Hand	Combination
Feeling of safety to cross (1 = most safe to cross)	3.58	1.85	3.23	1.19
Sureness of being seen (1 = surest that the car has seen me)	3.62	1.92	3.04	1.23
Affected by lack of eye contact (1 = most affected)	2.15	2.73	2.35	2.85
General preference (1 = most preferred)	3.46	1.73	3.42	1.35

In case the AV did not yield, the hand gesture use made it subjectively more difficult to predict the AV's behavior compared to the Baseline condition. A possible explanation is that the use of a hand gesture did not guarantee that the AV would yield (i.e., in 50% of the trials, the AV was programmed not to yield, regardless of hand gesture usage). These findings suggest that pedestrians may have difficulty in future traffic if only a portion of AVs is responsive to their hand signals. Such a situation could be likely, as future traffic is likely to consist of AVs of different brands having different computer vision abilities. In addition, it may be the case that

approaching vehicles are unable to stop, for example, because the traffic rules forbid this or because of traffic behind them.

Based on an analysis of communication between road users in today's traffic, Lee et al. [22] concluded that road users rarely use explicit communication such as hand gestures. They also pointed out that "there may be limited requirement for automated vehicles to adopt explicit communication solutions." The present study was undertaken from a different point of view. In our study, the pedestrians provided an explicit hand gesture, which was

detected and used by the simulated AV. The underlying idea was that future AVs will have difficulty in reading pedestrians' implicit body language and that explicit gestures are therefore needed. Besides, future AVs may have to detect explicit gestures of vulnerable road users for safety reasons (e.g., detection of an extended arm of cyclists [34, 36]) and to comply with traffic rules (e.g., being responsive to signals used by traffic police [31–33]). Furthermore, as others [8, 10, 37] have shown as well, the present study demonstrated that an eHMI makes visible the invisible. That is, it may be hard for pedestrians to detect the initiation of deceleration of an approaching vehicle; an eHMI makes such information salient, thereby improving subjective clarity.

Our study showed that false positives are of some concern. It is possible that pedestrians may be confused and raise their hand even for AVs that cannot stop for them, or that the AV may detect a hand gesture, even if the pedestrian did not intend to gesture to that particular AV. In other words, the hand gesture feature we introduced could add complexity to future traffic. More generally, the introduction of solutions to problems (in this case, the lack of communication in traffic) can create new problems (in our case, false positives and occasional confusion), a phenomenon Sheridan [38] referred to as “fixes to fixes” (p. 146).

It would be interesting to examine whether the eHMI should be isomorphic to the participant's gesture in order to establish a more efficient dialogue between pedestrian and AV. Several researchers have already proposed eHMIs in the form of a gesture; for example, Fridman et al. [39] and Hudson et al. [40] tested an upraised hand, whereas Mahadevan et al. [41] proposed an animated hand above the vehicle. It would be interesting to determine whether gesture-based eHMIs are more effective than the present text-based eHMI.

In the present study, the AV in the Baseline and eHMI conditions yielded in a random 50% of the trials. This appears to be a realistic percentage relative to contemporary manual driving. An observational study by Sucha et al. [3], for example, found that “36% of the drivers failed to yield to pedestrians in situations where they were obliged to.” A study in China by Zhuang and Xu [30] found that 63.5% of drivers did not even change speed when a pedestrian stood on the curb, and only 3.5% of the vehicles yielded. In our study, the AV could either yield or not yield, with no other behaviors possible. Further research could include other types of AV behaviors, such as showing its intention via lateral movement [42], slowing down but not stopping in order to let the pedestrian cross [3], or adapting to the pedestrian's behavior (e.g., braking if the pedestrian walks up to the curve and not braking if the pedestrian is showing hesitant behavior). Further research could also focus on the long-term consequences of the use of hand gestures. The results of our pre- versus postexperiment questions showed that, at the end of the experiment, the participants had gained trust in the idea of communicating with AVs through hand gestures. In the longer term, hand gestures may lead to misuse, where pedestrians raise their hand and cross the street without waiting for the AV to confirm [43]. Aside from regulations, there would be a need

for research into a standardized set of hand gestures for pedestrian-AV interaction, considering cultural differences in gesture use [44, 45]. Standardized gestures would make it easier for AV developers to train their computer vision systems to recognize these gestures. For research into standardization, large groups of users from diverse target groups, including children and older persons, would be needed.

6. Conclusions and Outlook

This lab-based study concludes that pedestrian gestures can be used to let an approaching AV stop. Furthermore, an eHMI on the AV depicting the message “I SEE YOU” makes the encounter clearer for the pedestrian. The present study further demonstrated the value of bidirectional communication: if the AV confirms it has seen the pedestrian, the pedestrian knows that their hand can be released again.

In the past, driving involved keeping the hands on the steering wheel. In modern times, such as with most Level 2 automated driving systems, drivers still have their hands on the steering wheel continuously and may perform shared control [46, 47]. In the future, automated driving systems may still require some human involvement, but this may take the form of intermittent control of maneuvers and prediction-level interventions through gestures or touchscreens [28, 48–50]. The present study suggests that gestures on behalf of the pedestrian could also have a role in future traffic.

Data Availability

The data and MATLAB script used to reproduce the figures in this paper are available via <https://doi.org/10.4121/14406944>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors thank Dr. Pavlo Bazilinskyy for discussion in the early stages of this research. This research was supported by grant 016.Vidi.178.047 (How should automated vehicles communicate with other road users?), which was financed by the Netherlands Organisation for Scientific Research (NWO).

References

- [1] A. Rasouli, I. Kotseruba, and J. K. Tsotsos, “Agreeing to cross: how drivers and pedestrians communicate,” in *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium*, Los Angeles, CA, USA, June 2017.
- [2] B. J. Crowley-Koch, R. van Houten, and E. Lim, “Effects of pedestrian prompts on motorist yielding at crosswalks,” *Journal of Applied Behavior Analysis*, vol. 44, pp. 121–126, 2011.

- [3] M. Sucha, D. Dostal, and R. Risser, "Pedestrian-driver communication and decision strategies at marked crossings," *Accident Analysis & Prevention*, vol. 102, pp. 41–50, 2017.
- [4] F. Keferböck and A. Riener, "Strategies for negotiation between autonomous vehicles and pedestrians," in *Proceedings of the Mensch und Computer 2015 Workshopband*, pp. 525–532, Stuttgart, Germany, September 2015.
- [5] P. Bazilinskyy, D. Dodou, and J. de Winter, "Survey on eHMI concepts: the effect of text, color, and perspective," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 67, pp. 175–194, 2019.
- [6] D. Dey, A. Habibovic, A. Löcken et al., "Taming the eHMI jungle: a classification taxonomy to guide compare and assess the design principles of automated vehicles' external human-machine interfaces," *Transportation Research Interdisciplinary Perspectives*, vol. 7, Article ID 100174, 2020.
- [7] 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, pp. 900–918, 2019.
- [8] 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: The Journal of the Human Factors and Ergonomics Society*, vol. 61, no. 8, pp. 1353–1370, 2019.
- [9] L. Kooijman, R. Happee, and J. de Winter, "How do eHMIs affect pedestrians' crossing behavior? A study using a head-mounted display combined with a motion suit," *Information*, vol. 10, no. 12, p. 386, 2019.
- [10] Y. B. Eisma, S. van Bergen, S. M. ter Brake, M. T. T. Hensen, W. J. Tempelaar, and J. C. F. de Winter, "External human-machine interfaces: the effect of display location on crossing intentions and eye movements," *Information*, vol. 11, p. 13, 2020.
- [11] S. M. Faas, L.-A. Mathis, and M. Baumann, "External HMI for self-driving vehicles: which information shall be displayed?" *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 68, pp. 171–186, 2020.
- [12] Y. E. Song, C. Lehsing, T. Fuest, and K. Bengler, "External HMIs and their effect on the interaction between pedestrians and automated vehicles," in *Intelligent Human Systems Integration*, W. Karwowski and T. Ahram, Eds., Springer, Cham, Germany, 2018.
- [13] F. Weber, R. Chadowitz, K. Schmidt, J. Messerschmidt, and T. Fuest, "Crossing the street across the globe: a study on the effects of eHMI on pedestrians in the US Germany and China," in *HCI in Mobility Transport and Automotive Systems. HCII 2019. Lecture Notes in Computer Science*, H. Krömker, Ed., Springer, Cham, Germany, 2019.
- [14] M. Rettenmaier, J. Schulze, and K. Bengler, "How much space is required? Effect of distance, content, and color on external human-machine interface size," *Information*, vol. 11, no. 7, p. 346, 2020.
- [15] D. Schlackl, K. Weigl, and A. Riener, "eHMI visualization on the entire car body: results of a comparative evaluation of concepts for the communication between AVs and manual drivers," in *Proceedings of the Conference on Mensch und Computer*, Magdeburg, Germany, September 2020.
- [16] M. Cefkin, J. Zhang, E. Stayton, and E. Vinkhuyzen, "Multi-methods research to examine external HMI for highly automated vehicles," in *HCI in Mobility Transport and Automotive Systems. HCII 2019. Lecture Notes in Computer Science*, H. Krömker, Ed., pp. 46–64, Springer, Cham, Germany, 2019.
- [17] A. Schieben, M. Wilbrink, C. Kettwich, R. Madigan, T. Louw, and N. Merat, "Designing the interaction of automated vehicles with other traffic participants: design considerations based on human needs and expectations," *Cognition, Technology & Work*, vol. 21, no. 1, pp. 69–85, 2019.
- [18] W. Tabone, J. de Winter, C. Ackermann et al., "Vulnerable road users and the coming wave of automated vehicles: expert perspectives," *Transportation Research Interdisciplinary Perspectives*, vol. 9, Article ID 100293, 2021.
- [19] D. Moore, R. Currano, G. E. Strack, and D. Sirkin, "The case for implicit external human-machine interfaces for autonomous vehicles," in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 295–307, Utrecht, Netherlands, September 2019.
- [20] M. Clamann, M. Aubert, and M. L. Cummings, "Evaluation of vehicle-to-pedestrian communication displays for autonomous vehicles," in *Proceedings of the Transportation Research Board 96th Annual Meeting*, Washington, NJ, USA, January 2017.
- [21] D. Dey, M. Martens, B. Eggen, and J. Terken, "Pedestrian road-crossing willingness as a function of vehicle automation, external appearance, and driving behaviour," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 65, pp. 191–205, 2019.
- [22] Y. M. Lee, R. Madigan, O. Giles, and L. Garach-Morcillo, "Road users rarely use explicit communication when interacting in today's traffic: implications for automated vehicles," *Cognition Technology & Work*, 2020.
- [23] P. Pandey and J. V. Aghav, "Pedestrian-autonomous vehicles interaction challenges: a survey and a solution to pedestrian intent identification," in *Advances in Data and Information Sciences*, M. Kolhe, S. Tiwari, M. Trivedi, and K. Mishra, Eds., Springer, Signapore, 2020.
- [24] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: a survey," *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 895–935, 2020.
- [25] E. Vinkhuyzen and M. Cefkin, "Developing socially acceptable autonomous vehicles," *Ethnographic Praxis in Industry Conference Proceedings*, vol. 2016, no. 1, pp. 522–534, 2016.
- [26] L. Graichen, M. Graichen, and J. F. Krems, "Effects of gesture-based interaction on driving behavior: a driving simulator study using the projection-based vehicle-in-the-loop. Human Factors," *The Journal of the Human Factors and Ergonomics Society*, 2020.
- [27] C. A. Pickering, K. J. Burnham, and M. J. Richardson, "A research study of hand gesture recognition technologies and applications for human vehicle interaction," in *Proceedings of the 2007 3rd Institution of Engineering and Technology Conference on Automotive Electronics*, Coventry, UK, June 2007.
- [28] H. Detjen, S. Geisler, and S. Schneegass, "Maneuver-based control interventions during automated driving: comparing touch voice and mid-air gestures as input modalities," in *Proceedings of the IEEE International Conference on Systems Man and Cybernetics (SMC)*, Toronto, Canada, April 2020.
- [29] I. Vijayan, M. H. Laur, and J. P. Absmeier, "Automated vehicle operation based on gesture to pedestrian," (U.S. Patent Application No. 14/987,188), 2017.
- [30] X. Zhuang and C. Wu, "Pedestrian gestures increase driver yielding at uncontrolled mid-block road crossings," *Accident Analysis & Prevention*, vol. 70, pp. 235–244, 2014.
- [31] F. Guo, J. Tang, and C. Zhu, "Gesture recognition for Chinese traffic police," in *Proceedings of the 2015 International*

- Conference on Virtual Reality and Visualization (ICVRV)*, pp. 64–67, Xiamen, China, November 2015.
- [32] J. He, C. Zhang, X. He, and R. Dong, “Visual recognition of traffic police gestures with convolutional pose machine and handcrafted features,” *Neurocomputing*, vol. 390, pp. 248–259, 2020.
- [33] Z. Fang, W. Zhang, Z. Guo, R. Zhi, B. Wang, and F. Flohr, “Traffic police gesture recognition by pose graph convolutional networks,” in *Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1833–1838, Las Vegas, NV, USA, November 2020.
- [34] J. F. P. Kooij, F. Flohr, E. A. I. Pool, and D. M. Gavrila, “Context-based path prediction for targets with switching dynamics,” *International Journal of Computer Vision*, vol. 127, no. 3, pp. 239–262, 2019.
- [35] R. D. Morey, “Confidence intervals from normalized data: a correction to Cousineau (2005),” *Tutorials in Quantitative Methods for Psychology*, vol. 4, no. 2, pp. 61–64, 2008.
- [36] H. Kretzschmar and J. Zhu, “Cyclist hand signal detection by an autonomous vehicle,” U.S. Patent No. 9,014,905, 2015.
- [37] T. Singer, J. Kobbert, B. Zandi, and T. Q. Khanh, “Displaying the driving state of automated vehicles to other road users: an international virtual reality-based study as a first step for the harmonized regulations of novel signaling devices,” *IEEE Transactions on Intelligent Transportation Systems*, 2020.
- [38] T. B. Sheridan, *Humans and Automation: System Design and Research Issues*, Wiley-Interscience, Hoboken, NJ, USA, 2002.
- [39] L. Fridman, B. Mehler, L. Xia, Y. Yang, L. Y. Facusse, and B. Reimer, “To walk or not to walk: crowdsourced assessment of external vehicle-to-pedestrian displays,” in *Proceedings of the Transportation Research Board Annual Meeting*, Washington, NY, USA, January 2019.
- [40] C. R. Hudson, S. Deb, D. W. Carruth, J. McGinley, and D. Frey, “Pedestrian perception of autonomous vehicles with external interacting features,” in *Advances in Human Factors and Systems Interaction. AHFE 2018*, I. Nunes, Ed., Springer, Cham, Germany, 2019.
- [41] K. Mahadevan, S. Somanath, and E. Sharlin, “Communicating awareness and intent in autonomous vehicle-pedestrian interaction,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal, Canada, April 2018.
- [42] A. Sripada, P. Bazilinskyy, and J. C. F. de Winter, *Automated Vehicles that Communicate Implicitly: Examining the Use of Lateral Position within the Lane*, Paper Submitted for Publication, 2021.
- [43] A. Millard-Ball, “Pedestrians, autonomous vehicles, and cities,” *Journal of Planning Education and Research*, vol. 38, no. 1, pp. 6–12, 2018.
- [44] B. Färber, “Communication and communication problems between autonomous vehicles and human drivers,” in *Autonomous driving*, M. Maurer, J. Gerdes, B. Lenz, and H. Winner, Eds., Springer, Berlin, Germany, 2016.
- [45] S. Gupta, M. Vasardani, and S. Winter, “Conventionalized gestures for the interaction of people in traffic with autonomous vehicles,” in *Proceedings of the 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, San Francisco, CA, USA, October 2016.
- [46] E. R. Teoh, “What’s in a name? Drivers’ perceptions of the use of five SAE level 2 driving automation systems,” *Journal of Safety Research*, vol. 72, pp. 145–151, 2020.
- [47] D. A. Abbink, M. Mulder, and E. R. Boer, “Haptic shared control: smoothly shifting control authority?” *Cognition, Technology & Work*, vol. 14, no. 1, pp. 19–28, 2012.
- [48] C. Wang, M. Krüger, and C. B. Wiebel-Herboth, ““Watch out!”: prediction-level intervention for automated driving,” in *Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Washington, NJ, USA, September 2020.
- [49] F. Flemisch, F. Diederichs, R. Meyer et al., “Vorreiter: manoeuvre-based steering gestures for partially and highly automated driving,” *Human-Computer Interaction Series*, pp. 231–304, 2020.
- [50] M. Kauer, M. Schreiber, and R. Bruder, “How to conduct a car? A design example for maneuver based driver-vehicle interaction,” in *Proceedings of the IEEE Intelligent Vehicles Symposium*, La Jolla, CA, USA, November 2010.