Can Cars Gesture? A Case for Expressive Behavior Within Autonomous Vehicle and Pedestrian Interactions

Paul Schmitt, Nicholas Britten, JiHyun Jeong, Amelia Coffey, Kevin Clark, Shweta Sunil Kothawade, Elena Corina Grigore, Adam Khaw, Christopher Konopka, Linh Pham, Kim Ryan, Christopher Schmitt, and Emilio Frazzoli

***Abstract*—One of the major challenges that autonomous vehicles (AVs) face in an urban setting is communicating with other road users such as pedestrians. In this work, we investigated with what expressive behaviors we can endow AVs such that pedestrians readily recognize the underlying intent of the vehicles' movements. The purpose of our study was to test the impact of expressive stopping behaviors on pedestrians' decision to cross a road. We utilized a virtual reality (VR) environment in which participants would have to cross a street in the presence of an oncoming vehicle that may or may not stop. Next, we crafted several expressive AV behaviors conveying its intention to stop for the pedestrian. Then, for each expressive design we recorded how quickly a pedestrian determined that it was safe to cross the street. We also administered repeated surveys of their subjective experiences. Our findings suggest that expressive behaviors such as easing into a full stop or stopping farther away can help pedestrians make quicker decisions to cross the road. Additionally, stopping farther away from the pedestrian also resulted in higher subjective experience for sense of safety, confidence, and intention understanding. We propose further investigation into expressive behaviors such as easing into a stop and stopping farther away to convey yielding intentions to pedestrians in future work.[1]**

***Index Terms*—Autonomous vehicles, human-robot interaction, animatronics.**

Manuscript received September 2, 2021; accepted December 13, 2021. Date of publication December 24, 2021; date of current version January 12, 2022. This letter was recommended for publication by Associate Editor Drazen Brscic and Editor Angelika Peer upon evaluation of the reviewers' comments. This work was supported by Motional, Boston, MA, USA. *(Corresponding author: Paul Schmitt.)*

Paul Schmitt, Elena Corina Grigore, Adam Khaw, Christopher Konopka, Linh Pham, Kim Ryan, and Christopher Schmitt are with the Motional, Boston, MA 02210 USA (e-mail: paul.schmitt@motional.com; elena.corina.grigore@motional.com; adam.khaw@motional.com; christopher.konopka@motional.com; linh.pham@motional.com; kimberly.ryan@motional.com; paul.schmitt@nutonomy.com).

Nicholas Britten is with Virginia Tech, Blacksburg, VA 24061 USA (e-mail: brittenn@vt.edu).

JiHyun Jeong is with Cornell University, Ithaca, NY 14853 USA (e-mail: jihyun@infosci.cornell.edu).

Amelia Coffey is with Tufts University, Medford, MA 02155 USA (e-mail: amelia.coffey@tufts.edu).

Kevin Clark is with Bowdoin College, Brunswick, ME 04011 USA (e-mail: kclark@bowdoin.edu).

Shweta Sunil Kothawade is with Northeastern University, Boston, MA 02115 USA (e-mail: shweta.sunil.kothawade@motional.com).

Emilio Frazzoli is with ETH Zurich, 8092 Zurich 02139, Switzerland (e-mail: emilio.frazzoli@idsc.mavt.ethz.ch).

Digital Object Identifier 10.1109/LRA.2021.3138161

[1]As a contribution to the field, all VR files used in this research are being open sourced at https://nureality.org.

I. Introduction

In this study, we present the exploration of expressive autonomous vehicle (AV) movements for the purpose of intuitively conveying a vehicle's intentions to yield to a pedestrian trying to cross a street. Interacting with an AV can be confusing for pedestrians due to a lack of familiar social cues. For instance, human drivers and pedestrians often use intuitive non-verbal cues like gaze, eye contact, hand and arm movements to communicate their intentions to stop and yield [1], [2]. Vulnerable road users, like pedestrians, rely on these social cues to recognize the driver's intentions and coordinate their crossing actions. Ideally, interacting with AVs should feel equally familiar and intuitive. With the absence of a human in the driver's seat, we designed artificial cues for the AV that make its intentions easily interpretable.

Inspired by the animation industry's use of movement [3], we designed expressive cues for the AV to signal its intentions to stop and yield to the pedestrian. These expressive cues include actions such as easing the deceleration rate of a vehicle to a stop, projecting sounds that exaggerate a car's braking and lower RPM sounds, exaggerating a vehicle's nose dive when braking, and stopping farther away from the pedestrian. We compared the effect of these implicit expressive cues against a state-of-the-art LED interface [4] as well as a baseline human-driven vehicle and a baseline AV where the deceleration profile was modeled after drivers braking at a stop-controlled intersection [5].

A within-subjects study design was employed so participants would experience all vehicle behaviors three different times within the realistic virtual reality (VR) environment of an intersection with an unmarked crosswalk. Our results show that expressive behaviors such as easing the deceleration to a stop resulted in earlier pedestrian cross times and lead more pedestrians to cross the road prior to the vehicle fully stopping. Even though it did not result in a *much* earlier crossing time, stopping farther away still lead more pedestrians to cross before the vehicle stopped when compared to baseline vehicle behaviors Interestingly, pedestrians subjectively did not report feeling more safe, confident, nor having a higher intention understanding for the vehicle that decelerated expressively compared to the baseline vehicles. Instead, participants felt safer and felt they understood the vehicle's intentions better when the AV stopped farther away from the pedestrian. Participants also felt a higher degree of confidence in making the crossing decision when they initially experienced the vehicle that stopped farther away.

We add to the field of pedestrian-Autonomous Vehicle interaction research by performing a study that compares animation inspired expressive vehicle behaviors against more traditional baseline driving conditions and a state-of-the-art e-HMI. Additionally, we release the Virtual Reality files used to perform this study in hopes of assisting the academic community to further explore the pedestrian-AV relationship.

II. Related Work

In this section, we discuss prior literature in designing interactions between autonomous vehicle and pedestrian interaction. We particularly focus on relevant concepts from the animation industry and in the field of Human-Robot Interaction for anthropomorphic and mobile robots. Much research has been performed in the subarea of legible robot motion planning [6]. Further, studies on the effects of exaggerating motion in order to optimize human-mobile robot interaction have yielded positive results [7], encouraging us to extend these concepts to AVs.

A. Pedestrian-AV Interaction

Prior research has widely explored visual mediums such as LED lights [8] and digital displays as a method for autonomous vehicles to communicate with varying degrees of success. For instance, [9] had color based warnings displays, while [10] visualized a smile, a hand waving gesture, and a pedestrian light to the signal intent of the car. [11] also tested various colors and motions of the external LED interface to signal the car's intent to yield. [12] even designed eyes on the car's headlights. Having AVs with eyes that looked at the pedestrian increased pedestrians' subjective feeling of safety and resulted in earlier decisions to cross the street compared to AVs with no eyes [12]. Cyan colored LED lights were perceived as neutral and a pulsing pattern was preferred to a sideways sweeping one to signal the car's intentions to stop [11]. And informative displays with symbols helped reduce pedestrians' decision time and supported pedestrian's comfort of an AV [10]. Of the many external displays including front brake lights, smiling digital display, sweeping LED lights and text displays, text was the least ambiguous for pedestrians [13]. However, many studies noted that the visual display would only be a secondary modality as many participants seemed to base their crossing decisions on the kinematics of the AV, its velocity and distance from the pedestrian [9], [10], [14]. Moreover, some participants had trouble understanding what information was being communicated through visuals, for example, finding eyes on cars confusing [12].

Audio cues have also been considered by researchers. Along with visual and physical cues, [14] tested explicit auditory cues such as using the speaker to broadcast messages like "about to stop". Similarly, [15] had words like "Cross Now" displayed and spoken out to the pedestrians. In other cases, artificial engine, tire and braking sounds were projected to communicate intentions of a silent electric vehicle [16]. Being able to hear the vehicle decelerate and accelerate was reported as crucial to pedestrian decision-making when visibility was low, such as in bad weather conditions [16].

A review of literature emphasizes vehicle's implicit communication through motion to be the most effective way to convey intent to pedestrians [1], [17]. Therefore, we decided to double down on this finding when investigating how a vehicle might expressively behave to convey intent to pedestrians.

B. Using Expressive Movements

Researchers have looked at designing motions that can convey robot intentions in the field of Human-Robot Interaction, often inspired by techniques and principles from the animations industry [18], [19]. Several studies have particularly referenced the 12 Principles of Animation to make robots more expressive. [20] prototyped human-robot interaction behaviors using animation techniques of anticipation and reaction to create human-readable behaviors that explain what the robot is doing and what it will do next. [21] was able to leverage the animation principle of timing to express the robot's internal states. Without adjusting path but by manipulating timing, the robot was able to express its level of confidence, naturalness and difficulty of task to users.

Another key framework for robot behaviors is legibility and predictability [22]. Legible robot behavior is expressive, such that a naive observer can infer the robot's goals quickly and confidently. Just as expressive motion can convey the robot's intentions to users [23], we can design expressive motions for autonomous cars to convey its intentions to pedestrians.

Drivers often convey intentions to stop and yield the right of way to pedestrians implicitly through non-verbal movements, braking and stopping far away from the pedestrian [24]. Movement can evoke visceral reactions from pedestrians that influence their crossing decision [25]. Findings from an in-the-wild Wizard-of-Oz (WoZ) study of an autonomous vehicle and pedestrians suggest that cars need not have explicit cues like visual lights and displays for the pedestrians to cross in front of the vehicle, unless the car violates pedestrian expectations like behaving erratically [26].

Building on this background, we collaborated with animators to apply principles from animation to create our expressive AV behaviors and then tested the efficacy of those behaviors in communicating with pedestrians.

III. Hypotheses

Our study aims to investigate whether expressive vehicle cues could affect a participant's decision on when to cross the street as well as their subjective experience while making such a decision. A robot's expressive motion can increase the legibility [22] of its goals, leading to a faster recognition of its intended future state. This earlier recognition of vehicle intention by pedestrians can lead to decisions to cross the street earlier, as opposed to waiting for the vehicle to reach a complete stop.

In addition to earlier decision-making, expressive and legible behaviors can influence the way pedestrians feel about interacting with the AV. If the pedestrians feel like they have a clear sense of what the AV is intending to do, they could feel safer and more confident to cross the street.

Fig. 1. A panoramic view of the virtual reality environment. Participants start each trial facing the four-way intersection and the flashing yellow traffic light. Upon hearing an auditory cue, they look to their left to see a white vehicle approaching.

Finally, expressive behaviors could be more intuitive for pedestrians, which would enable them to quickly learn to recognize the vehicle's intentions with limited repeated exposures. Hence, we test the following hypotheses:

- H1. Participants will cross the street at earlier times in relation to a vehicle's full stop when interacting with AVs with expressive behaviors.
- H2. Participants will have a better subjective experience interacting with AVs with expressive behaviors.
- H3. Pedestrians will more likely respond to expressive behaviors faster with less exposure.

IV. Experiment

We designed a within-subjects experiment in virtual reality where, in each of the trials, a vehicle approaches the participants as they are intending to cross the street at an intersection.

A. Setup

The virtual reality environment depicts a four-way urban intersection (Fig. 1). There is no clear pedestrian crossing zone and a flashing yellow traffic light is visible. Flashing yellow lights indicate that cars need not stop, but should proceed carefully [27]. The pedestrian could look around using the VR headset, but they could not move within the environment. The oncoming vehicle was a white van modeled after a 2019 Chrysler Pacifica. The color white was selected for its global popularity [28].

We contracted a professional animation studio to create a more realistic environment and vehicle behavior to elicit higher fidelity reactions from participants. Numerous details were added throughout the environment such as road texturing, aged bricks, parked cars, swaying trees, city noise etc. The virtual environment ran with Unreal Engine 4.25 and a HTC Vive Cosmos Elite setup (Cosmos Elite head-mounted display, Vive controller, and two base tracking stations) at 90 Hz with a resolution of 2880 x 1700. We used an HP OMEN Obelisk 875–1022 desktop with an Intel i9-9900 k processor, a Nvidia Geforce RTX 2080ti Graphics Processor, and 16 GB DDR4 RAM. The public release of our VR files used in the study can be found at https://nureality.org.

B. Participants

We recruited 60 adults (M = 43.8, SD = 15.28, male = 52.7% and female = 47.3%) from the Boston, MA area to participate in the study. All participants self-reported crossing the street at least once per week on average, and did not require the use of an assistive mobility device. The study protocol followed enhanced COVID measures to protect the participants' and experimenter's safety. Approval to conduct this study was obtained from Motional's Institutional Review Board on July 17, 2020. Participants received a $75 Amazon gift card as compensation for their time.

C. Procedure

Participants first gave written informed consent and completed a pre-survey on their demographics, mobility, and crossing patterns. Then, they received training on the VR environment and task.

The task consisted of participants experiencing up to 33 trials of a vehicle approaching an intersection that the participant is trying to cross. Participants were instructed to signal the first moment they felt it was safe to cross the road. Each participant had 21 trials where the vehicle stopped in one of seven ways (see section IV-D) three separate times. They also had up to 12 trials where the vehicle rolled through without stopping (see section IV-D3). The order of these trials were randomized for each participant. After completing each trial, participants verbally answered a three item questionnaire about their subjective experience.

Participants completed a single practice trial prior to beginning the study. We administered the Misery Scale questionnaire [29] half-way through the study to ensure that participants were not experiencing motion sickness.

In the beginning of each trial, participants faced the intersection and the traffic light. We notified the participants using a short auditory cue to look to their left and see the approaching vehicle. The short auditory cue was added after pilot study observations. We found that the initial position of vehicles at the start of each scenario was too far from the pedestrian that participants would choose to cross the street without looking at the vehicle's approach. Implementing the short auditory cue to delay participants' exposure to the vehicle made it necessary for pedestrians to negotiate crossing decisions with the approaching

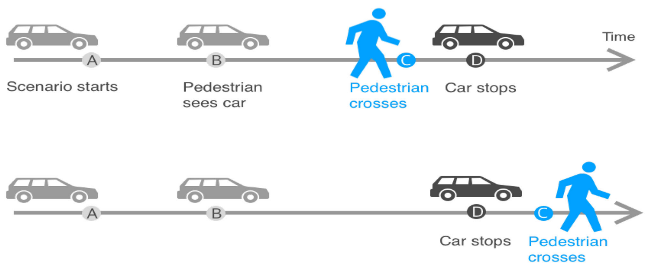


Fig. 2. Scenario starts at Point A. The pedestrian looks over to the car, at Point B, when the car reaches 55 meters from the crossing point. We mark the time when the pedestrian decides to cross the road as Point C. (Thus, B to C, measured in distance, is 55 meters.) The time when the vehicle comes to a full stop is at Point D. Pedestrians can cross before the car stops (above) or afterwards (below).

Fig. 3. Vehicle stops 5 meters farther away (C5).

Fig. 4. Left depicts the appearance of a vehicle with its nose down as it is braking. Right depicts the car with its nose raised with an overlay of its previous nose dive state in red.

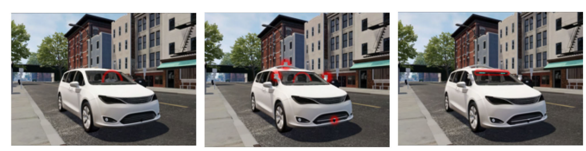


Fig. 5. A baseline human-driven vehicle (left, C1) and a baseline AV (mid, C2), and a state-of-the-art AV with cyan LED lights that move in a sweeping motion (right, C3). AVs have visible sensors and no one in the driver's seat. The differences in vehicle conditions are highlighted in red for clarity.

TABLE I
A SUMMARY TABLE OF FEATURES FOR EACH VEHICLE TYPE CONDITION

	AV	Baseline Decel.	LED eHMI	Expr. Decel.	Stop Short	Expr. Sound	Nose dive
C1		✓					
C2	✓	✓					
C3	✓	✓	✓				
C4	✓			✓			
C5	✓	✓			✓		
C6	✓	✓				✓	
C7	✓	✓				✓	✓

vehicle. We also made the decision to control for the distance at which the pedestrian first saw the vehicle based on prior studies that suggested pedestrians are more sensitive to the approaching vehicle's distance than time until arrival [30], [31]. In all conditions, participants first saw the approaching vehicle at 55 meters out (at Point B in Fig. 2).

Participants marked when they would cross the street by pressing the trigger on the Vive controller. The time they initiated crossing (Point C in Fig. 2) was recorded as time in seconds from the start of the trial.

D. *Experimental Conditions*

Table I shows the different modalities and behaviors for vehicle type in each condition. Vehicle conditions C1, C2, and C3 are our baseline conditions (Section IV-D1). C1 has a human driver in a conventional vehicle. C2 has an AV with no visible occupants. C3 has an AV that has a LED light bar display on its windshield. Vehicle conditions C4 through C7 has our expressive behaviors (Section IV-D2). C4 exhibits our expressive deceleration behavior. C5 stops 5 m short. C6 projects expressive sound. C7 combines expressive nose dive and expressive sound.

1) Baselines: We devised three baseline conditions (Fig. 5). All baseline vehicles followed a deceleration curve modeled after human drivers stopping at a stop-controlled intersection in a natural setting [5]. We estimated the deceleration curve by fitting a polynomial function to extrapolate the velocity and distance values by time.

- **Human Driver** A conventional human-driven vehicle baseline (C1) was modeled after a white 2019 Chrysler Pacifica, with a single occupant looking straight ahead in the left driver's seat.
- **AV** An AV baseline (C2) had the same white van equipped with visible LIDAR sensors mounted on its side and roof as well as a RADAR sensor on the front of the car. There were no visible occupants in the car.
- **LED eHMI** The state-of-the-art AV (C3) had a LED light bar mounted on its windshield. When yielding, cyan colored LEDs displayed a sweeping motion from the center to the edge of the bar following the work of [4].

2) Expressive Behavior Conditions: Expressive vehicle conditions exhibit one or more of the following expressive behaviors:

- **Expressive Deceleration** Following the animation's principle of "slow in and slow out" [3], we collaborated with animators to modify the baseline deceleration so that it slowly eases out of the braking motion to a full stop (C4). The AV exaggerates this motion by initiating the braking sequence earlier and it achieves a peak deceleration value 10% greater and 10 meters earlier than the baseline scenario.
- **Stop Short:** The vehicle followed the same deceleration pattern of the baseline vehicles, but stopped 5 meters farther away from the pedestrian (C5, Fig. 3). The team selected a distance that would be readily discernible to a pedestrian observer, but not so large as to confuse intention. The sizing of a vehicle was felt to meet these requirements,

so the length of the virtual AV, 5 meters, was selected. In practice, we had the vehicle start the scenario 5 meters back and exposed the pedestrians to the vehicle when it reached 55 meters from the pedestrian crossing point. It then travelled for another 50 meters following the baseline vehicle's deceleration pattern, and stopped 5 meters farther away from the crossing point.

- **Expressive Sound** We added exaggerated brake noises and low engine RPM sounds that increases in volume as the car approaches closer to the pedestrian. C6 tests this behavior in isolation, while C7 tests a vehicle with both expressive nose dive and expressive sound.
- **Expressive Nose Dive** In line with "anticipation" and "exaggeration" principles of animation [3], the vehicle approached the intersection with its nose down, to mimic heavy braking. Fig. 4 depicts the vehicle's undercarriage compressed against its front tires, appearing to absorb the forward momentum of the car (left).
  From initial testing, we discovered that participants had difficulty discerning nose dip behavior from the baseline. The authors feel this is due to the movement range constrained by the limits of a typical active suspension (approximately 1 inch in each direction). Additionally, movement was limited in order to minimize potential discomfort felt by the riders due to unnatural movements. We chose to combine the Nose Dip with C6 because we wanted to see if combing the two subtler behaviors could have any compounding effects.

3) Control: We added a control condition where vehicles did not yield but rolled through the intersection. We collaborated with animators to devise a convincing vehicle behavior that approached while braking but ultimately rolled through the intersection without fully stopping. The control condition was intended to counter learning effects across trials. Without this condition, participants might assume that the vehicle will stop in every trial and learn to cross earlier regardless of vehicle behavior. Participants were exposed to the control condition up to twelve times throughout the study. The cars in the control conditions were a mix of human-driven conventional vehicles used in C1 and AVs used in C2. We did not record any data for these trials.

E. Measures

1) Pedestrian Cross Time: We measure the gap between the time the vehicle came to a full stop (Point D in Fig. 2) and the time the pedestrian decided to cross the street (C). We call this metric $\Delta\ Crossing\ time$. Since we control for the distance at which the participant is exposed to the car, different deceleration inevitably leads to a difference in the time it takes the car to come of a full stop (D). To control for this variance, we define a metric called $Exposure\ time$ to indicate how long the pedestrian was exposed to the vehicle for, from exposure (B) to crossing (C). We define pedestrian cross time as $\Delta\ Crossing\ time$ over $Exposure\ time$.

$$Pedestrian\ Cross\ Time = \frac{\Delta\ Crossing\ time}{Exposure\ time} = \frac{(D-C)}{(C-B)}.$$

2) Crossed Before Stop: We extract a binary categorical variable "Crossed Before Stop" to indicate whether the pedestrian crossed prior to the vehicle's stop or after. If $D - C > 0$, we note 1 to indicate the pedestrian crossed before the vehicle's stop. If not, we note 0 to indicate the pedestrian crossed after the vehicle came to a stop.

3) Subjective Experience Survey: A panel of domain experts from the research team (consisting of product managers, HRI researchers and user experience designers) identified safety, confidence, and intention understanding as likely antecedents for trust in the AV. Participants rated the level of agreement with the following statements on a 7-point Likert scale (1= I strongly disagree, 7 = I strongly agree): "I felt safe interacting with the vehicle" (sense of safety); "I felt confident with my decision" (decision confidence); "I understood the intention of the vehicle before it arrived at the intersection" (intention understanding). We opted for rating single statements for each construct due to concerns of participant fatigue over multiple trials.

V. Results

Of the 60 participants recruited, seven participants withdrew from our study before completion. and their data was excluded in our analysis. In total, we analyzed data for 1,113 pedestrian crossing trials from our 53 participants. We used R [32] to conduct our analysis.

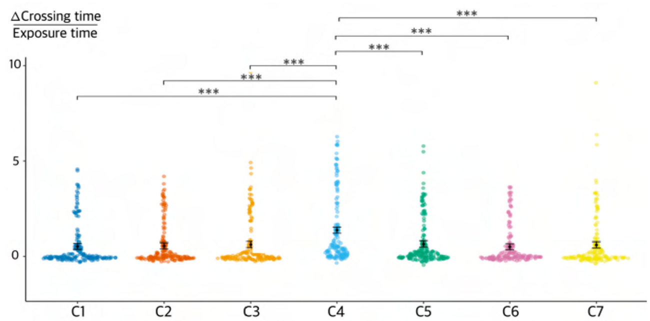
A. Crossing Behavior

The dependent variables are pedestrian cross time (Section IV-E1) and crossed before stop (Section IV-E2). The independent variables are vehicle condition and exposure number. Vehicle condition is a nominal variable that indicates which of the seven vehicle behaviors the participant was interacting with (C1-C7). Exposure number indicates whether it was the first, second or third time the participant was exposed to the particular vehicle condition.

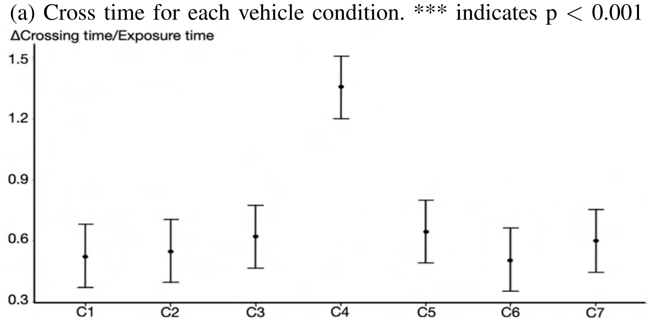
1) Pedestrian Cross Time: The distribution of pedestrian cross time violated normality assumptions determined by the Shapiro-Wilk test [33], [34]. Therefore, we used linear mixed models to analyze pedestrian cross time [35]. We fit a linear mixed model using vehicle condition and exposure number as fixed factors and unique participant ID as a random factor. There was a main effect for vehicle condition on pedestrian cross time ($F(6,1040) = 12.9671$, $p<0.001$). We did not find a main effect for exposure number ($p=0.22$), nor an interaction effect between our independent variables ($p=0.88$).

We performed post-hoc pairwise comparisons and adjusted all p-values using the Holm–Bonferroni method [36]. Post-hoc tests revealed a significant contrast for C4 compared to our baseline conditions ($p<0.001$). This means that using expressive deceleration (C4) led to earlier crossing decisions by pedestrians (Fig. 6). This result supports hypothesis H1.

We also believe that the lack of exposure effects or interaction effects support hypothesis H3. Even from the first exposure, pedestrians were able to interpret the vehicle's expressive deceleration and responded to it by crossing early.



(a) Cross time for each vehicle condition. *** indicates $p < 0.001$



(b) Zoomed in estimated means and standard errors from (a)

Fig. 6. Pedestrians make earlier crossing decisions when interacting with the vehicle with expressive deceleration (C4).

TABLE II
PERCENTAGE OF PEDESTRIANS THAT CROSSED THE STREET BEFORE EACH VEHICLE CAME TO A FULL STOP

	C1	C2	C3	C4	C5	C6	C7
Crossed before stop (%)	44.7	50.9	49.6	84.9	63.5	52.2	48.4

2) Crossed Before Stop: The dependent variable, crossed before stop, follows a binomial distribution which violates assumptions of parametric models. As a result, we used the non-parametric Aligned Rank Transform (ART) tool implemented in R [37]. We took vehicle condition and exposure number as fixed factors and unique participant ID as random factor. The fitted model converged and met all assumptions. ANOVA of Aligned Rank Transformed Data with Kenward-Roger F tests noted a main effect for vehicle condition ($F(6,1040) = 29.95$, $p<0.001$). We also observed a main effect for exposure number ($p=0.05$). There was not an interaction effect between our independent variables ($p=0.25$).

Post-hoc comparisons were performed using contrast tools developed by [38]. P-values were corrected for multiple tests using the Holm–Bonferroni method. More pedestrians crossed the street before the vehicle fully stopped in the expressive deceleration condition (C4) and the stop short condition (C5) compared to the baseline vehicles (C1,C2,C3) ($p<0.001$). The results support our hypothesis H1. To illustrate, 84.9% and 63.5% of the pedestrians crossed before C4 and C5 came to a complete stop, respectively.In contrast, 44.7%, 50.9%, and 49.6% of the pedestrians crossed before C1, C2, C3 came to a complete stop (Table II).

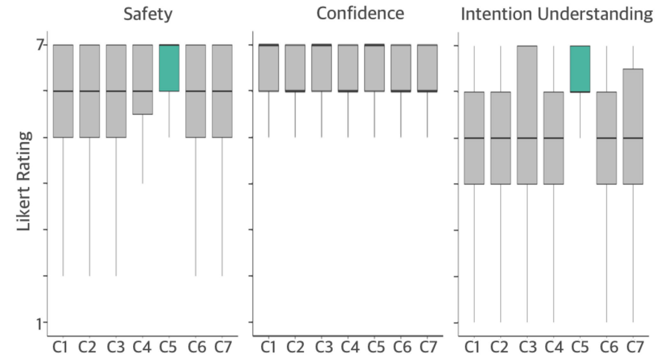


Fig. 7. Boxplots of Likert ratings for safety, confidence, and intention understanding. Stopping farther away (C5, highlighted) scored higher than all other vehicle behaviors for safety and intention understanding ($p<0.001$). Pedestrian confidence level was higher when they observed the car stopping farther away for the first time (C5, E1) compared to several other vehicles conditions and exposures ($p<0.05$). Outliers are not included in this plot.

Moreover, post-hoc comparisons indicated that second exposure to a vehicle condition resulted in a higher rate of crossing before stop compared to the first exposure to a vehicle condition ($p=0.04$). Pedestrians seemed to be learning to cross earlier on the second exposure compared to the first. There were not significant differences between any of the other pairs of exposures. The learning effect was observed in general, and not particular to any of the vehicle behaviors. In fact, expressive deceleration (C4) and stop short (C5) behaviors were able to elicit higher rate of crossing before stop even from the first exposure. Hence, the results support hypothesis H3.

B. *Subjective Experience*

Participants' subjective experiences were recorded using an ordinal 7-point Likert rating (Fig. 7). Since survey response options violate assumptions for parametric tests, we conducted ANOVA analyses using aligned rank transformed data with Kenward-Roger F tests. For each survey item, we fit a model with vehicle condition and exposure number as fixed factors and unique participant ID as a random factor. All reported p-values for post-hoc comparisons have been adjusted using the Holm–Bonferroni method. Results for each survey item is presented in Fig. 7.

1) Sense of Safety: ANOVA analysis for pedestrians' sense of safety indicated a main effect of vehicle condition condition ($F(6,1040) = 18.2848$, $p<0.001$). We also observed a main effect for exposure number on pedestrians' sense of safety ($F(2,1040) = 3.8433$, $p=0.02$). There was no notable interaction effect between our fixed factors at the $p<0.05$ level. Post-hoc comparisons (Fig. 7) revealed that vehicle stopping short (C5) resulted in a higher rating of safety compared to all other vehicles ($p<0.001$). Expressing yielding intentions by stopping short was effective at making pedestrians feel safer. This result supports hypothesis H2.

Post-hoc comparisons of exposure levels indicated a trend of participants feeling slightly less safe interacting with the vehicle by the third exposure compared to the first ($p=0.052$) and second ($p=0.03$) exposures. We did not expect this decrease in pedestrians' sense of safety and speculate on possible explanations for this effect in Section VI.

2) Decision Confidence: We observed an interaction effect between vehicle condition and exposure number on the pedestrian's confidence level ($F(12,1040) = 2.0579$, $p=0.017$). Post-hoc comparisons revealed that a pedestrian's first exposure to the vehicle stopping short (C5) resulted in higher ratings of decision confidence compared to several other vehicle condition and exposure combinations ($p<0.05$). The result for decision confidence partially support hypothesis H2.

3) Intention Understanding: We found a main effect for vehicle condition on pedestrians' level of vehicle intention understanding ($F(6,1040) = 21.5418$, $p<0.001$). There was not a clear main effect for exposure number ($p=0.07$) nor an interaction effect between our fixed factors ($p=0.4$). Post-hoc comparisons revealed that vehicle stopping short (C5) resulted in a higher rating of intention understanding compared to other vehicles ($p<0.001$). The result partially supports hypothesis H2.

VI. Discussion

In order to determine if expressive behaviors could communicate easily recognizable stopping intentions, we tested several novel expressive behaviors against our baselines. Based on the results we investigate which expressive behaviors are worth exploring further. We evaluated each expressive behavior on whether and to what extent it influenced a pedestrian's decision to cross early; and to what extent pedestrians felt safe, confident, and as though they understood the vehicle's intentions. Note that the effectiveness of each expressive behavior should be interpreted with the limitations of our study setup and measures in mind.

First, expressive deceleration (C4) resulted in an earlier pedestrian cross time. We define early crossing behavior in relation to a vehicle coming to a full stop. Since the vehicle with expressive deceleration "eases in" to a stop, it took longer to come to a full stop when travelling the same amount of distance. In order to test whether the expressive behavior can result in earlier decision in relation to the initial time of exposure, future work could create expressive deceleration behavior that controls for the time it takes for a vehicle to stop as well as the distance it travels. Our reasoning for not following the above direction is that the high rate of initial deceleration necessary to control for stopping time and distance could be uncomfortable to passengers and may look unnatural to observers.

Furthermore, stopping short (C5) resulted in higher percentage of pedestrians crossing before the vehicle stopped. While pedestrians may not have crossed substantially earlier than the vehicle's stop, they did do so narrowly. We think that investigating stop short behavior (C5) further in future work could be promising. Pedestrians in this study were exposed to C5 for less amount of time than vehicles in other conditions as they only watched the vehicle travel the last 50 meters instead of 55. One one hand it's possible that the results are due to pedestrians responding to the vehicle in C5 having a slightly lower velocity when it was 55 meters away compared to baseline vehicles at that same distance. However on the other hand, despite not being exposed to the vehicle for as long, pedestrians successfully made crossing decisions prior to the vehicle stopping. It could be that stopping farther away is a legible enough behavior to facilitate an efficient and early crossing decision. Further work is needed to pinpoint which specific aspects of the behavior pedestrians respond best to.

Another reason to further investigate stopping short behavior is because it resulted in higher ratings in subjective experience. More specifically, stopping farther away resulted in a higher sense of safety and intention understanding. Finally, this behavior might seem unintuitive to trailing drivers; while such interact are beyond the scope of this research, it's important to explore in future work.

Despite the pedestrians' earlier crossing response to expressive deceleration conditions (C4), expressive deceleration did not help pedestrians to feel as safe, confident, or clear about the vehicle's intentions as C5 did. Perhaps pedestrians report favorable experiences with the vehicle that stopped short because they were more accustomed to it. Stopping short is one of the social norms that human drivers tend to follow when yielding to pedestrians on crosswalks [24]. In line with this point, further exploration of driver norms in the presence of pedestrians would be a useful direction for designing AV behaviors. For example, another frequently observed behavior is a human driver decelerating early to allow for the pedestrian to cross the street without having to completely stop the car [24].

Next, we observed some exposure effects on pedestrians' crossing before stop behavior and their feelings of safety and confidence. Pedestrians were learning to cross earlier than a vehicle's stop on their second exposure compared to their first. However, we did not observe the same type of learning trend on the third exposure. Therefore, we infer that our control condition was somewhat effective at countering pedestrians' learning of our experimental setup.

Conversely, participants reported feeling least safe on their third exposure to a vehicle condition. This may be due to an overcorrecting effect from our control condition. Observing more vehicles roll through the intersection may have caused people to feel less safe. Future experiments perhaps would benefit from a decreased number of control condition trials.

Moreover, effective expressive behaviors were recognizable from the first exposure. There was no interaction effect between exposure and vehicle condition on crossing behavior. No particular condition was easier to learn to respond to within three trials. Therefore, future work could focus on testing instantly recognizable behavior and less on behaviors that pedestrians might implicitly learn to adapt to.

Additionally, we observed somewhat of a novelty effect for C5 on a person's confidence to cross the street. The first exposure to the stopping short behavior led to higher confidence levels in decisions to cross. Stopping short was perhaps the most visually recognizable behavior and thus could have had a stronger novelty effect for pedestrians.

As an interesting remark, we did not find a significant difference between pedestrians interacting with the AV baseline (C2) versus the human driver baseline (C1). This is somewhat surprising, and may relate to comments made in debriefing sessions by a few of the participants that they did not even notice the absence of a driver.

In summary, we contribute to the pedestrian-AV literature by i) creating expressive movements inspired by animation techniques while balancing feasibility of behaviors and ii) testing the effectiveness of such behaviors in a realistic virtual reality environment. Furthermore, we publicly release the virtual reality environment to aid future investigations. By collaborating with a professional animation studio, our VR environment was able to create immersive experiences and elicit visceral responses from the participants. We hope open-sourcing these files will inspire and aid researchers to conduct exciting new research by applying animation techniques to the domain of designing behaviors for autonomous vehicles.

Acknowledgment

The authors thank Motional and the team at CHRLX for all of their support. All authors are or were with Motional at the time they contributed to this work.

References

[1] A. Rasouli, I. Kotseruba, and J. K. Tsotsos, "Agreeing to cross: How drivers and pedestrians communicate," in *Proc. 4th IEEE Intell. Veh. Symp.*, 2017, pp. 264–269.

[2] B. J. Crowley-Koch, R. V. Houten, and E. Lim, "Effects of pedestrian prompts on motorist yielding at crosswalks," *J. Appl. Behav. Anal.*, vol. 44, no. 1, pp. 121–126, 2011.

[3] J. Lasseter, "Principles of traditional animation applied to 3D computer animation," in *Proc. 14th Annu. Conf. Comput. Graph. Interactive Techn.*, 1987, pp. 35–44.

[4] J. Shutko, T. Bray, A. Schaudt, S. McLaughlin, and V. Williams, "Evaluation of AV external communication in the wild," in *Proc. Taskforce Auton. Veh. Signalling Requirements*, Doc. AVSR-02-23e, 2018. [Online]. Available: https://wiki.unece.org/display/trans/AVSR+-+2nd+session%2C+2018-12-20

[5] Z. R. Doerzaph, "Development of a threat assessment algorithm for intersection collision avoidance systems," Ph.D. dissertation, Virginia Polytech., Inst. State Univ., 2007. [Online]. Available: https://vtechworks.lib.vt.edu/handle/10919/30013

[6] A. Dragan, "Legible robot motion planning," Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, USA, Jun. 2015.

[7] M. J. Gielniak and A. L. Thomaz, "Enhancing interaction through exaggerated motion synthesis," in *Proc. 7th Annu. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, New York, NY, USA:Association for Computing Machinery, 2012, pp. 375–382. [Online]. Available: https://doi.org/10.1145/2157689.2157813

[8] A.-C. Hensch, I. Neumann, M. Beggiato, J. Halama, and J. F. Krems, "Effects of a light-based communication approach as an external HMI for automated vehicles—A wizard-of-oz study," *Trans. Transport Sci.*, vol. 10, no. 2, pp. 18–32, Jan. 2020.

[9] Y. Li, M. Dikmen, T. G. Hussein, Y. Wang, and C. Burns, "To cross or not to cross," in *Proc. 10th Int. Conf. Automot. User Interfaces Interactive Veh. Appl.*, 2018, pp. 188–197.

[10] K. Holländer, A. Colley, C. Mai, J. Häkkilä, F. Alt, and B. Pfleging, "Investigating the influence of external car displays on pedestrians' crossing behavior in virtual reality," in *Proc. 21st Int. Conf. Hum.- Comput. Interact. Mobile Devices Serv.*, 2019, pp. 1–11.

[11] D. Dey, A. Habibovic, B. Pfleging, M. Martens, and J. Terken, "Color and animation preferences for a light band eHMI in interactions between automated vehicles and pedestrians," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–13.

[12] 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. Automot. Interfaces Interactive Veh. Appl.*, 2017, pp. 65–73.

[13] 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," *Hum. Factors*, vol. 61, no. 8, pp. 1353–1370, 2019.

[14] K. Mahadevan, S. Somanath, and E. Sharlin, *Communicating Awareness and Intent in Autonomous Vehicle-Pedestrian Interaction*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–12.

[15] M. Matthews, G. Chowdhary, and E. Kieson, "Intent communication between autonomous vehicles and pedestrians," 2017. [Online]. Available: https://arxiv.org/abs/1708.07123

[16] A. Pillai, "Virtual reality based study to analyse pedestrian attitude towards autonomous vehicles," M.S thesis, KTH Royal Inst. Technol., 2017. [Online]. Available: http://www.diva-portal.se/smash/get/diva2:1153080/FULLTEXT01.pdf

[17] A. Rasouli and J. K. Tsotsos, "Autonomous vehicles that interact with pedestrians: A survey of theory and practice," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 900–918, Mar. 2020.

[18] G. Hoffman and W. Ju, "Designing robots with movement in mind," *J. Hum.-Robot Interact.*, vol. 3, no. 1, pp. 89–122, Mar. 2014.

[19] T. Schulz, J. Torresen, and J. Herstad, "Animation techniques in human-robot interaction user studies," *ACM Trans. Hum.-Robot Interact.*, vol. 8, no. 2, pp. 1–22, Jun. 2019.

[20] L. Takayama, D. Dooley, and W. Ju, "Expressing thought: Improving robot readability with animation principles," in *Proc. 6th Int. Conf. Hum.-Robot Interact.*, 2011, pp. 69–76.

[21] A. Zhou, D. Hadfield-Menell, A. Nagabandi, and A. D. Dragan, "Expressive robot motion timing," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, 2017, pp. 22–31.

[22] A. D. Dragan, K. C. Lee, and S.S. Srinivasa, "Legibility and predictability of robot motion," in *Proc. 8th ACM/IEEE Int. Conf. Hum.-Robot Interact.*, 2013, pp. 301–308.

[23] T. Ribeiro and A. Paiva, "Animating the Adelino robot with Erik: The expressive robotics inverse kinematics," in *Proc. 19th ACM Int. Conf. Multimodal Interact.*, 2017, pp. 388–396.

[24] M. Risto, C. Emmenegger, E. Vinkhuyzen, M. Cefkin, and J. Hollan, "Human-vehicle interfaces: The power of vehicle movement gestures in human road user coordination," in *Proc. Driving Assesment Conf.*, 2017, vol. 9, pp. 186–192.

[25] R. Zimmermann and R. Wettach, "First step into visceral interaction with autonomous vehicles," in *Proc. 9th Int. Conf. Automot. User Interfaces Interactive Veh. Appl.*, 2017, pp. 58–64.

[26] D. Rothenbucher, J. Li, D. Sirkin, B. Mok, and W. Ju, "Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles," in *Proc. 25th IEEE Int. Symp. Robot Hum. Interactive Commun.*, 2016, pp. 795–802.

[27] M. R. of Motor Vehicles, "Rules of the road," Dec. 2020. [Online]. Available: https://www.mass.gov/doc/chapter-4-rules-of-the-road-0

[28] BASF Coatings GmbH, *Basf color report 2019 for automotive oem coatings*, 2019. [Online]. Available: https://www.basf.com/global/documents/en/news-and-media/news-releases/2020/01/BASF_Color%20Report%202019_Press.pdf

[29] J. E. Bos, S. N. MacKinnon, and A. Patterson, "Motion sickness symptoms in a ship motion simulator: Effects of inside, outside, and no view," *Aviation, Space, Environ. Med.*, vol. 76, no. 12, pp. 1111–1118, 2005.

[30] M. Beggiato, C. Witzlack, S. Springer, and J. Krems, "The right moment for braking as informal communication signal between automated vehicles and pedestrians in crossing situations," in *Proc. Int. Conf. Appl. Hum. Factors Ergonom.*, 2017, pp. 1072–1081.

[31] J. A. Oxley, E. Ihsen, B. N. Fildes, J. L. Charlton, and R. H. Day, "Crossing roads safely: An experimental study of age differences in gap selection by pedestrians," *Accident Anal. Prevention*, vol. 37, no. 5, pp. 962–971, 2005.

[32] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statist. Comput., Vienna, Austria, 2020. [Online]. Available: https://www.R-project.org/

[33] J. P. Royston, "An extension of Shapiro and Wilk's w test for normality to large samples," *J. Roy. Stat. Soc.: Ser. C (Appl. Statist.)*, vol. 31, no. 2, pp. 115–124, 1982.

[34] P. Royston, "Remark as R94: A remark on algorithm as 181: The w-test for normality," *J. Roy. Stat. Soc. Ser. C (Appl. Statist.)*, vol. 44, no. 4, pp. 547–551, 1995.

[35] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using LME4," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, 2015, doi: 10.18637/jss.v067.i01.

[36] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Statist.*, vol. 6, no. 2, pp. 65–70, 1979.

[37] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only Anova procedures," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2011, pp. 143–146.

[38] L. Elkin, M. Kay, J. Higgins, and J. Wobbrock, "An aligned rank transform procedure for multifactor contrast tests," 2021, Accessed: Nov. 29, 2021. [Online]. Available: https://arxiv.org/abs/2102.11824