Shrinkable Arm-based eHMI on Autonomous Delivery Vehicle for Effective Communication with Other Road Users

Xinyue Gui
The University of Tokyo
Tokyo, Japan
xinyueguikwei@gmail.com

Mikiya Kusunoki
JAIST
Ishikawa, Japan
s2420019@jaist.ac.jp

Bofei Huang
JAIST
Ishikawa, Japan
hbf105s@gmail.com

Stela H. Seo
Kyoto University
Kyoto, Japan
stela.seo@i.kyoto-u.ac.jp

Chia-Ming Chang
The University of Tokyo
Tokyo, Japan
info@chiamingchang.com

Haoran Xie
JAIST
Ishikawa, Japan
xie@jaist.ac.jp

Manabu Tsukada
The University of Tokyo
Tokyo, Japan
mtsukada@g.ecc.u-tokyo.ac.jp

Takeo Igarashi
The University of Tokyo
Tokyo, Japan
takeo@acm.org

Figure 1: a shrinkable arm (shrunk when lower down and expanded when raised up) attached to an autonomous delivery vehicle (aka delivery robot) to express its crossing request to a car driver, and the simulator used in evaluation (Study one/Scenario 1).

ABSTRACT

When employing autonomous driving technology in logistics, small autonomous delivery vehicles (aka delivery robots) encounter challenges different from passenger vehicles when interacting with other road users. We conducted an online video survey as a pre-study and found that autonomous delivery vehicles need external human-machine interfaces (eHMIs) to ask for help due to their small size and functional limitations. Inspired by everyday human communication, we chose arms as eHMI to show their request through limb motion and gesture. We held an in-house workshop to identify the arm's requirements for designing a specific arm with shrink-ability (conspicuous when delivering messages but not affect traffic at other times). We prototyped a small delivery robot with a shrinkable arm and filmed the experiment videos. We conducted two studies (a video-based and a 360-degree-photo VR-based) with 18 participants. We demonstrated that arm-on-delivery robots can increase interaction efficiency by drawing more attention and communicating specific information.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in interaction design**.

KEYWORDS

arm-based eHMI, transformation design, autonomous delivery vehicles, delivery robot

AutomotiveUI '24, September 22–25, 2024, Stanford, CA, USA

ACM ISBN 979-8-4007-0510-6/24/09
https://doi.org/10.1145/3640792.3675716

ACM Reference Format:
Xinyue Gui, Mikiya Kusunoki, Bofei Huang, Stela H. Seo, Chia-Ming Chang, Haoran Xie, Manabu Tsukada, and Takeo Igarashi. 2024. Shrinkable Arm-based eHMI on Autonomous Delivery Vehicle for Effective Communication with Other Road Users. In *16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '24), September 22–25, 2024, Stanford, CA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3640792.3675716

1 INTRODUCTION

Researchers in AutomotiveUI explored a broad range of external human-machine interface (eHMI), attached outside of the autonomous vehicle in different forms (text display, light, eye, etc.) [45] to convey various information (intent, perception, status, etc.) [17, 20] to other road users (pedestrians, cyclists, car drivers, etc.) [2, 58]. During the transition period of mixed traffic (associating various levels of autonomous vehicles with traditional vehicles), the eHMI design research direction has emerged with the development of autonomous driving technology. However, even with the researchers' significant efforts, we could not cover the entire communication space between road users and autonomous vehicles, such as autonomous delivery vehicles.

Other than the normal vehicle, autonomous driving systems draw attention to the service industry [36], such as transportation, entertainment, and logistics. Among these, we found our particular interest in logistics: unlike other applications, autonomous delivery vehicles can have various forms or sizes from passenger cars while utilizing the self-driving algorithm and possibly manipulating real-world objects. The industry has already noticed its usefulness and deployed autonomous delivery vehicles such as Starship, Uber Eat, and Tier 4 Kawasaki. These are categorized as autonomous delivery vehicles [22, 69] since vehicles often carry people, whereas delivery robots often carry goods. In the industry, people called them "delivery robots," and we use this term throughout the paper to distinguish them from the traditional vehicles used in evaluation (Figure 1).

Delivery robots move mainly on the sidewalk and do not occupy the traffic road [29]. They are designed to divert traffic during peak hours and address labor shortages [12]. As such, while delivery robots can vary in size, most are designed to be less than half the height of an average adult and narrow enough to move among people [1]. The different forms and sizes lead to different challenges in designing feasible eHMIs when they need to communicate with people. To understand the nature of delivery robots, we gathered and analyzed 50 video clips of delivery robots, including vlogs, marketing videos, and so on (see Appendix A: The Video Clip List). We found the delivery robot to be an outlier in the field of autonomous driving. They perform as autonomous vehicles, but at the same time act as pedestrians. Even so, they are more vulnerable than pedestrians due to their limited capabilities. For example, a delivery robot cannot press a button for a button-activated crosswalk. As such, delivery robots' eHMI needs to show different types of information from the current eHMI, which focuses on accident reduction and pedestrian safety by showing "stop" or "please go." That is, delivery robots need to show their request information, such as "help me."

Inspired by everyday human communication, we see that an arm is a good candidate for the delivery robots' eHMI. As a common sense, people raise or wave their hands to draw their communication partner first, using their body language to express their request if they are far away from each other [8]. Through an in-house workshop, we determined our arm shape. For the delivery robot, an arm is not necessary for navigation, but it is useful for asking for help, notifying car drivers, etc. If the delivery robot has a long arm, it is disturbing its navigation or causing discomfort to pedestrians. Thus, we determined to design a shrinkable arm. Then, we manufactured a physically shrinkable arm with a hand out of polypropylene and 3D printed it. We prototyped a delivery robot with a paper box and a controllable toy car and attached the arm to the box. The prototype delivery robot can be operated remotely using a controller, and the arm can be operated via Bluetooth.

We used the prototype to take the video and photo for our evaluation. Based on situations in real life (particularly clip 03 and clip 20 in Appendix), we designed two scenarios and corresponding user study environments: (1) delivery robot crosses the road (participant as a car driver in a car simulation) and (2) delivery robot asks for button pressing (participant as a pedestrian in a VR environment). We have two conditions: (1) baseline: the delivery robot with a static flag (the common appearance used in the industry [62]) and (2) proposed: the delivery robot with a shrinkable arm. We recruited 18 participants and conducted the user study with these two conditions and two scenarios. Our result proved the benefit of an arm-type eHMI for a delivery robot to communicate with other road users effectively. In summary, our research has three main contributions:

- a video survey to identify the potential challenges in delivery robot interaction with other road users and design two testbed scenarios.
- a design of shrinkable arm, allowing the delivery robot to show request information when expanding and can be shrunk in other time to avoid affecting other road users.
- a formal user study with two scenarios where scenario 1 showed that the arm can help the delivery robot cross the road effectively and scenario 2 showed that the arm can help the delivery robot express their request.

2 RELATED WORK

2.1 Delivery robot in environment

The current research on how delivery robots work in the environment is divided into two components. The first component is a sensing, detecting, and reacting module that allows delivery robots to complete tasks autonomously [35, 49]. The second is the moment of handover [40, 67]. This includes the design of an phone app interface and the distribution network system to efficiently manage multiple robots. This technical research is well-developed, as seen by many companies' delivery robots that have passed testing and are already in operation in the real world. However, we discovered that the system that operates flawlessly during their tests encounters a variety of unanticipated issues in real life (as described in Section 3).

We hypothesized that this situation occurred for two reasons. The first is because the "human factor" was not included during the testing procedure. Human behavior is difficult to pre-define in an experiment as we cannot predict every possible human behavior, and environmental changes are unpredictable (e.g., potholes in the road, icy roads, piles of snow) [52]. The second is because the delivery robot is somehow too weak to achieve their goals independently (e.g., small formfactors or collision-safe materials) [48]. In such situations, robots need to ask for people's help. For example, the sociable trash box robot [66], an automatic moving car with a trashcan, is primarily used to clean the streets. However, due to its minimalistic design and limited capability, it verbally asks passersby to grab garbage for it. Therefore, bidirectional communication is required. Not only should it receive information from the environment, but when the delivery robot is incapable of completing its task, it should also be able to send messages for specific requests.

Previous research supported a similar view based on an observational study [11, 57, 68]. These findings emphasize the need to investigate robots' interactional work throughout their deployment. For example, nonverbal help-seeking tactics for urban robots and bystander factors shape assistive behaviors. Other than the visual eHMI, audio interaction is feasible with the example that a delivery robot asks people to press the elevator button [28]. The speech-based method could be a good choice at the moment of handover (when the customer and robot are close and ready to interact with each other). However, when the robot is on delivery, most people are not paying attention to the robot (in many cases, they are using ear pods) or they are far away from the communication partners. In addition, the noise in the outdoor environment interrupted the robot's speech interaction toward road users. In such a situation, audio does not work and we proposed that the design of a visual-based eHMI to show request information about the delivery robot is necessary.

2.2 The design space of eHMI

There is a body of research work in visual eHMI comparing various modalities [7, 10, 19]. Although multiple studies [4, 9] have confirmed that "text is one of the most effective eHMI methods," we cannot directly apply these findings to delivery robots. We explain the reasons in three dimensions.

The first is the implementation method. The same modality, but various implementation methods yield different results. For example, distinct user feedback was obtained for the same modality, text, when projection and display were implemented [9]. The modality of eyes, implemented by the 3D model displayed by video or physical robotics prototypes, provides users with a different experience [31, 65]. The second parameter is the message type [61]. Faa. et al. [24] previously compared three types of information that eHMI can convey: perception, intent, and status. Dey et al. [18] expanded to six types (intention, instruction, situational, awareness, warming, and advising) in "taming the eHMI." There are various historical examples that demonstrate that, even if the information content is identical, different types of information will produce distinct consequences [27]. The third aspect is the scene setting. "Pedestrians crossing the road" is a popular research scenario with the important research goal of maintaining pedestrians' safety using eHMI. In this setting, researchers conduct research from multiple perspectives, including pedestrians, cyclists, and drivers [3, 5, 6, 15]. The results reveal that pedestrians are eager to search for detailed information from eHMI to know autonomous vehicles' explicit intentions [17, 21, 24] (for example, "after you."). However, when autonomous driving technology enters the service industry, pedestrians may require different information communicated via eHMI [34, 56].

When we took these three dimensions together, we discovered that in different instances, road users require different communication formats. A certain modality may have a significant advantage in delivering a specific message type, but it could be ambiguous and ineffective in transmitting other information [16, 30]. We need a more comprehensive design framework. Colley et al. [13, 14] created a two-dimensional design space by combining the three-type concept dimension and the six-type scenario dimension. Based on this design space and the three dimensions discussed above, we built our eHMI prototype for the delivery robot. When we apply the scene setting to the delivery robot [47, 54], "text display" may not be a plausible modality since the text may be too small to read. We required a visible eHMI that could provide information to road users in multiple directions (for example, a pedestrian behind a delivery robot or a driver sitting inside a car). Thus, we decided to "build a physical prototype of our eHMI." Based on our pre-study, we examined and determined that the delivery robot's eHMI needed to show "request"-type information. In our daily lives, when we ask someone a question or seek assistance, we typically raise a hand. To sum up, we chose to create a physical arm-based eHMI for a delivery robot to communicate their request to other road users effectively.

2.3 Arm as a communication modality

There is research proposed for arm-based eHMI design for vehicles [50]. However, because it is not mainstream, there is a few research that may serve as a foundation for this discussion. To initiate our discussion, we first broaden our scope to human-computer interaction (HCI) and human-robot interaction (HRI) [59, 64]. We found that, unlike in the vehicle eHMI design, arms are widely used by robots as mostly task-completing tools rather than interactive communication tools. While many robotic arms are developed to manipulate real-world objects (e.g., grab, press) – that is, the arm acts as a body extension [23, 60, 63] – they can also be a communication tool to convey ideas with arm gestures or limb motions. We extend the theme of using an arm as a communication tool to autonomous vehicles by leveraging it to ask for people's help and convey its intention to other road users.

We have two categories of the robot's arm behavior types to communicate information with other road users. One is arm limb movement, which includes lifting and waving [32], and another is a hand gesture (such as the palm facing up or down) [51] or a finger change (such as pointing) [37]. Combining these two categories leads to research akin to employing robotic arms to transmit sign language [44]. When conducting arm-based eHMI for vehicles, we discovered that previous studies primarily used the second category [26, 46, 53]. However, given the small size of the delivery robot, we believe that an eHMI that combines arms and hands would be more appropriate and effective, as it can "raise" its arm to draw people's attention and subsequently the hand gesture to transmit its request.

We found that if a humanoid robot is being developed, they often have two arms, and much research converts commonly used human body language into robotic motion and applies it to the robot to evaluate if it can deliver the same information [25, 55]. However, if the arm-based eHMI is intended for vehicles, it cannot be directly replicated. Because of its distinctive appearance (the delivery robot resembles a little box rather than a human-like or anthropomorphic shape), experts are unsure whether it can be utilized by attaching an arm and performing similar behaviors [42]. As a result, even some typical sensing actions necessitate a thorough study to ensure their feasibility. Our work is an initial step in ensuring the feasibility of the arm-based eHMI on delivery robots.

3 PRE-STUDY: ONLINE VIDEO SURVEY

To analyze the real industry challenges, we conducted an online video survey with the 50 delivery-robot-related clips (Appendix A: The Video Clip List, later referred to as clip #). Several autonomous delivery companies have tested their delivery robots in specific locations, including Snack-E in Los Angeles (clip 01), Refraction AI in Austin (clip 23), Starship in Leeds (clip 16), and Uber Eats in Tokyo (clip 21). We obtain a wide range of information online. These videos highlight the convenience that technical improvements bring to people. However, several vlogs demonstrate the limitations of delivery robots in erroneous situations. We categorized them into three types.

The first involves the interaction of car drivers and delivery robots. The driver's on-road experience contradicts the appearance of the delivery robot. When pedestrians need to cross the street, they typically continue to gaze at the vehicles; in some nations, the pedestrian nods or bows to the driver. After waiting a while or receiving a similar answer from the driver, the pedestrian will cross the street. However, this does not happen on the delivery robot (clip 08, clip 16). For example, in the video (clip 03), a delivery robot hesitates to move forward or backward on the road's edge, leaving the driver clueless and hesitating to halt or drive. The two parties reached an impasse. Some delivery robots use the light to show their crossing intention, and while it seems to work well at night (clip 19).

The second concerns the interaction between pedestrians and delivery robots. Most of the time, delivery robots walk alongside pedestrians on the pavement (clip 15, clip 35). Other autonomous cars' interactions with pedestrians are limited to road crossings. However, the delivery robots are moving with pedestrians all over the sidewalk (clip 11, clip 18). The video shows that conflicts between delivery robots and pedestrians occur on occasion (clip 05, clip 06, clip 13).

The third is the limitation of delivery robots. For example, on Google Maps, we understand that driving routes for automobile navigation differ from walking routes for pedestrian navigation. The majority of driving routes consist of flat roads. The walking route will travel through various locations, including stairs, grass, and uneven terrain (clip 34, clip 36). Humans can walk those uneven terrains; however, delivery robots frequently get stuck in various locations (clip 10, clip 24, clip 32, clip 33) due to their small size and inflexible wheels. In addition, as delivery robots occasionally need to use an elevator to get upstairs or to cross the street with the button-activated pedestrian traffic light (clip 20), they frequently cannot reach their destination on their own.

To summarize, we hypothesize that, due to their size and functional limitations, delivery robots are often trapped in transit and need to maintain communication with their surroundings to resolve their issues. Delivery robots require an evident eHMI to communicate a help request with detailed content. According to the difficulties we found, we selected and adapted two scenarios for our research evaluation (clip 03, clip 20). The two scenarios (Figure 2) are about road crossing with different road conditions and interacting partners. In scenario 1, the delivery robot asked a car driver to let it cross the road on the pedestrian's first cross-situation (i.e., a crosswalk without a traffic light). In scenario 2, the delivery robot requested assistance from a walking pedestrian to press a button for a crosswalk traffic light.

4 PROTOTYPING THE DELIVERY ROBOT WITH A SHRINKABLE ARM

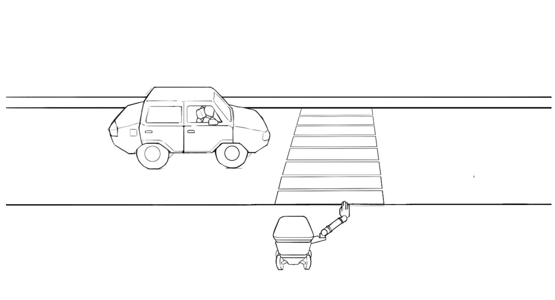
4.1 The delivery robot

While conducting the video survey, we looked into the size of delivery robots that have appeared on the market. The height of delivery robots manufactured by various companies varies. Large robots, such as 7-Eleven Snack-E, stand approximately 1.2 meters tall. Smaller robots, such as Starship, Kiwibot, and Uber Eats, stand between 0.5 and 0.6 meters tall. We chose to build a smaller one because they are more likely to experience problems and need an eHMI to draw people's attention and request their assistance. First, we chose a 45-cm-tall paper box, coated it with white paper, and mounted it on a four-wheel-drive toy car (Figure 3 (a)(b)). Our delivery robot's total height is around 50 cm.

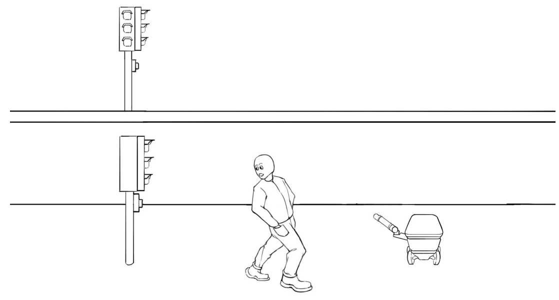
4.2 The arm and hand

We held an iterative in-house workshop to design the arm shape. We started with a regular digital model and 3D printed it (Figure 4, Type 1). On our first iteration, we realized an issue with our arm prototype: as the delivery robot is operating outside, having a long arm (i.e., in the state of "spread arms") on the body is not safe for other pedestrians or vehicles. As such, we set the default arm position to "lower down" and near the body. If the arm is excessively long and longer than the delivery robot, it would be unable to "lower down" entirely around the delivery robot when it is not active. If we shorten the arm's length, the arm would not be visible to other road users even if it were raised. As a result, we abandoned the traditional shape and experimented with the shrinkable/collapsable structure.

We took the origami idea that the arm can shrink in its default state and expand to show its request. Computational origami is a popular field for focusing on cylindrical structures to dynamically manipulate their volume. Researchers [39, 43] explored various origami mechanisms and assessed their advantages and disadvantages in terms of ease of use, fabrication time, and usefulness. Among them, the twisted tower structure attracts our attention because of its stability, pressure resistance, elasticity, and durability. We found other research that proved its usefulness in semi-soft robotics design [38, 41]. Based on these, we created a telescopic

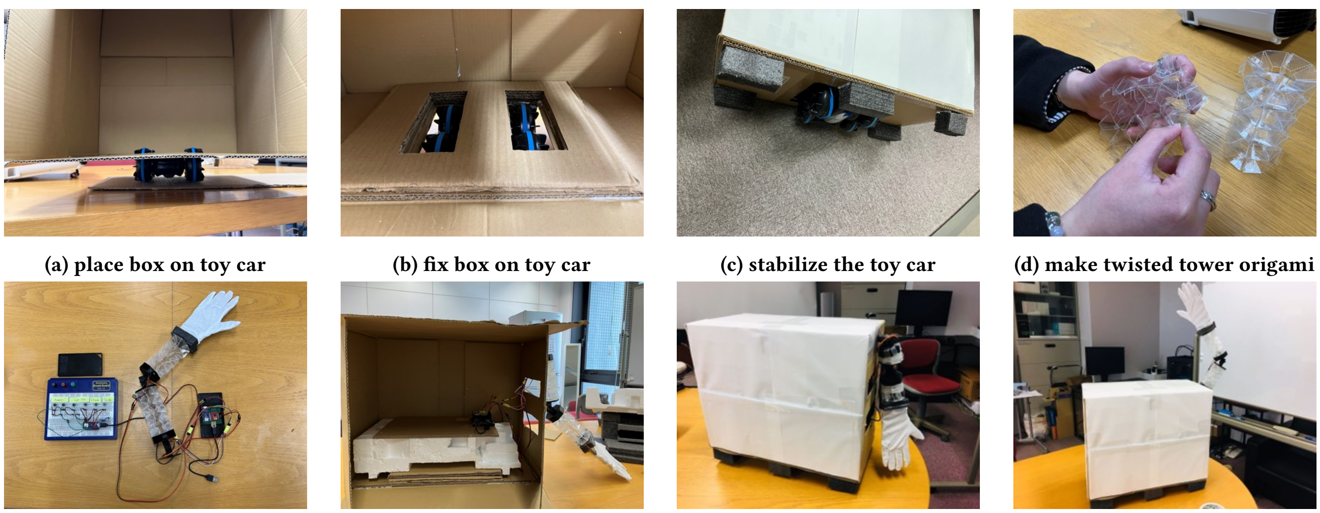


(a) Scenario 1: Request for road crossing



(b) Scenario 2: Request for button pressing

Figure 2: the two testbed scenarios for evaluation. Scenario 1 is the interaction between delivery robot and car driver. Scenario 2 is the interaction between delivery robot and pedestrian.



(a) place box on toy car (b) fix box on toy car (c) stabilize the toy car (d) make twisted tower origami

(e) arm and controller (f) attach arm on box (g) shrunk arm (h) expanded arm

Figure 3: The procedure of making the overall prototype

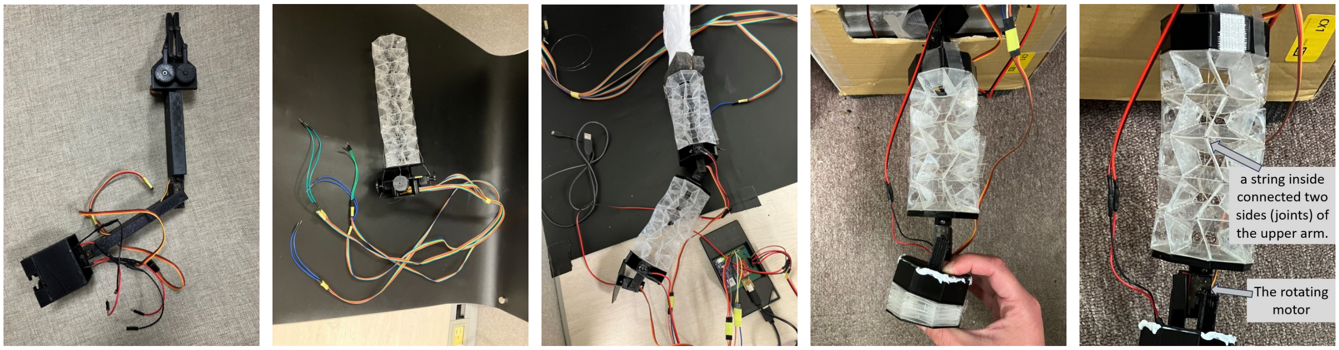
arm (Figure 4, Type 2). Our improved prototype encountered another challenge. When the arm is 20 cm or longer, it starts to bend. That is, even though the structure is stable, there is not enough strength to lift the entire length due to the material limitations. In addition, with this arm prototype, we could not have fine control over bending a specific point of the arm.

To address this challenge, we created a new design in which the limbs are composed of polypropylene with an origami structure (Figure 3 (d)), and the shoulder and elbow joints are 3D printed from a stronger material (Figure 4, Type 3). The total length of the arm with the hand is 42 cm when fully shrunk, making it suitable for use on any existing delivery robot (Figure 3 (g)). On the other hand, when the arm is fully expanded, it measures 60 cm long, reaching the average of a woman's arm length ranges (Figure 3 (h)). That is, our shrinkable robotic arm is suitable for delivery robots and our goal of communicating with other road users.

4.3 Assembly and control

Finally, we attached the shrinkable arm to the delivery robot prototype we built (Figure 3 (f)). We finalize our prototype by fine-tuning its center of gravity distribution and adding sponge to the box's bottom to lessen inertial impact (Figure 3 (c)). The robot's movement is driven by a remote controller, while the microcontroller's (Arduino Nano 33 BLE) Bluetooth controls the stretching and lifting of the arms (Figure 3 (e)). The shrink-ability function is achieved by the thread-driven method. There is a motor inside the joint, and a translucent thread is placed inside the origami cylinder. While the motor is activated, it curls up the thread to compress the origami cylinder. When it rotates in the opposite direction, it releases the thread, and the origami cylinder returns to its expanded state due to its own elastic (Figure 4 (e)). We tested the prototype indoors, simulating the state of a delivery robot transporting goods.

We devised two arm movements for the two testbed scenarios based on the arm's typical daily use. In scenario 1, the arm is raised when the delivery robot asks to cross metaphoring the request "may



(a) Type 1: normal arm **(b) Type 2: all in origami** **(c) Type 3: limb in origami** **(d) Type 3: partly shrunk** **(e) Type 3: thread-driven**

Figure 4: The in-house design workshop for our arm development. In type 3, forearm is collapsed / upper arm is expanded (the right)

I cross the street now?" and lowers down the arm when there is no such intent. In scenario 2, we present a text display that reads "could you help me to press the button" while raising the arm and pointing to the position of the traffic light's press-activated button.

4.4 Participants and general procedure

We invited 18 people (all university students, 6 women, and 12 men) to participate in our user study with a $7 honorarium. The evaluation contains two studies and five steps: (1) the verbal explanation for Study One; (2) the trial and formal test for Study One; (3) a short break and verbal instruction for Study Two; and (4) the formal test for Study Two. (5) the post-study interview. The entire evaluation takes approximately 30 to 40 minutes.

5 STUDY ONE

5.1 Scenario one: delivery robot interacts with car driver

In scenario 1, we informed participants that they are a car driver driving a car forward (note that it is left-hand traffic). In this scenario, we explained that there are multiple vehicles lined up behind the participant's car, and there is a delivery robot at one end of the zebra crossing (Figure 1, middle). The delivery robot may decide to wait for the car driver (and all the other vehicles) to go first, or it may request the driver to allow it to cross the road first. We took the video that followed the story in four types from the driver's perspective (Figure 5). There are two options. Case 1: The delivery robot intends to cross. Case 2: The delivery robot intends to wait (and will not cross for the time being).

There are two conditions. 1) Baseline: A delivery robot with a flag: there is no difference between the two cases, as we observed from some videos [62]. 2) Proposed: A delivery robot with a shrinkable arm; for case 1, the arm is raised at the start of the video, whereas for case 2, the arm remains down. To avoid bias in the user's judgment, in all videos, we drove the car at a constant speed. In the study, we chose a static flag as a baseline, as this is how the delivery robot appears in industry. The flag is intended to make the delivery robot more visible in congested places. That is, the flag is not designed to communicate its intention. In this "crossing road" scenario, the baseline represents a no-eHMI state.

We built a simple in-car simulator using a monitor, a wheel, and a pedal. Participants watch the video on the 28-inch monitor (figure 1, right). Participants could see the delivery robot as 5 cm when it first appeared, then getting bigger as the car got closer. The participant's task is to carefully observe the delivery robot and determine whether it intends to cross the road now or not. We instructed the participants to step on the brakes to stop the car if they believed the delivery robot was asking them to let it cross. If they believe the delivery robot intends to wait, they do not need to take any action. As experienced drivers would instinctively step on the brake to slow down when they noticed an object in front of them, we asked the participant to step on the pedal only if they believed the delivery robot wanted to cross the road right now for an accurate measure. We performed a within-subject test for Study One. We repeated each type of video 12 times in a random-mixed order. In each task, we edited the video until the car was three meters away from the crossroad (before the delivery robot crossed). Each task is 6-second long. A green progress bar and a countdown at the top of the video indicate the amount of time remaining, and participants can only make their decision while the video is playing. Each participant was given a trial period to practice. There is no data recorded during the practice time. They then completed 48 tasks (12 for each type and 4 perspective types) in the formal study. When the user steps on the pedal, the system records their decisions and calculates the accurate decision rate.

5.2 Result: correct decision rate

In scenario 1, we first calculated the correct decision rate of whether the car driver (participant) chooses to stop the car or not after observing the delivery robot's request intention. The paired t-test revealed statistically significant differences ($p<0.001$) between the flag (M = 0.5116, SD = 0.069) and the shrinkable arm eHMI (M = 0.8102, SD = 0.21; Figure 6, left). The results showed that participants understood the delivery robot's cross request more accurately with the raised arm compared to the flag.

We separated the scenarios according to the delivery robots' two options (request to cross or wait till cars pass by). The paired

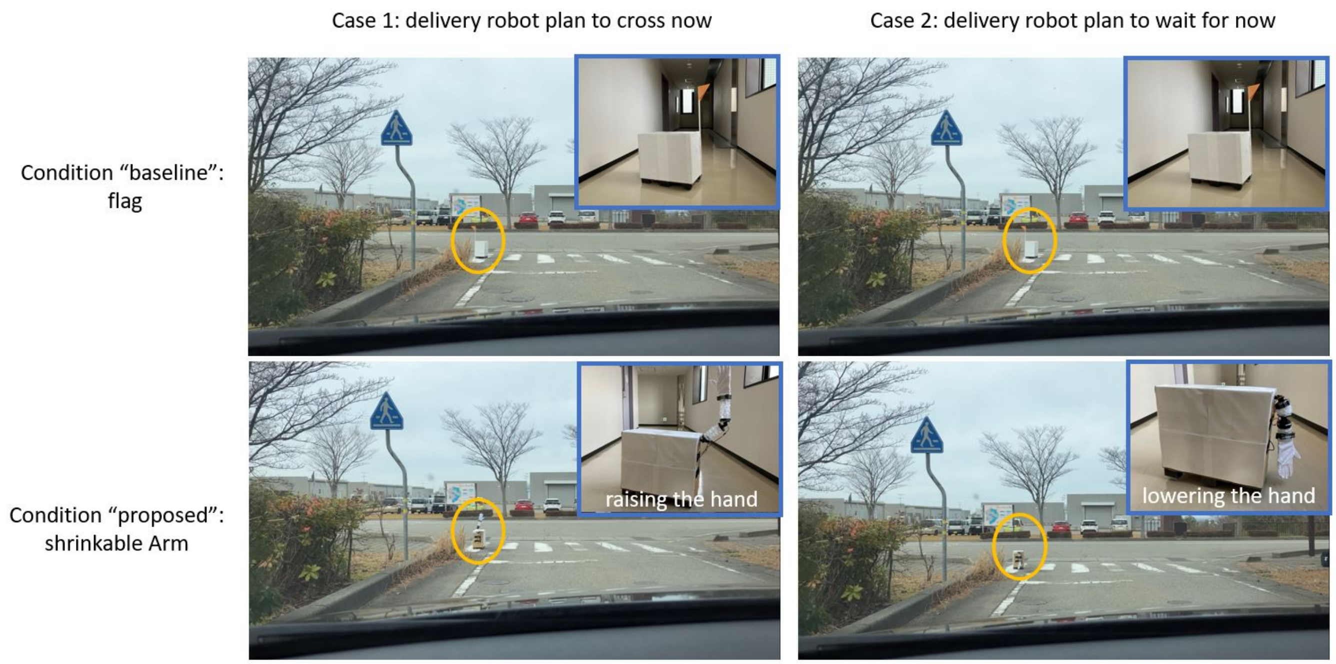


Figure 5: Four types (two conditions * two cases) of videos we took for our study one. Note that yellow circle highlights the robot on the scene and we placed close-up picture of the robot on the right-top of each scene.

t-test revealed statistically significant differences (*p<0.001*) in case 1, where the delivery robot requests to cross the road using the flag (*M=0.3056, SD=0.31*) or using the shrinkable arm-based eHMI (*M=0.8611, SD=0.24*; Figure 6, right). However, we did not find any statistically significant differences in case 2, where the delivery robot waits for the cars to pass by with the flag (*M=0.7176, SD=0.34*) or the shrinkable arm-based eHMI (*M=0.7593, SD=0.36*; Figure 6). The data indicates that our proposed arm-based eHMI helps the "delivery robot intends to cross the road."

6 STUDY TWO

6.1 Scenario two: delivery robot interacts with pedestrian

In scenario two, we chose an area in front of a train station. At this intersection, some pedestrians exit the train station and turn right rather than crossing. A delivery robot is waiting next to the intersection (to avoid blocking the road), hoping for assistance in pressing the button for the crosswalk traffic light to cross the road. The robot uses the eHMI to initially capture people's attention and then request help. We captured two 360-degree photos at the pedestrian position (Figure 7, middle, display in equirectangular projection format). In the "baseline" condition, when the display above the robot's head states, "Please help me to press the button," it stands with the flag. In the "proposed" condition, it raised its arm and pointed to the traffic light's button location while the screen displayed the same text. In this study two, our main goal is to compare our shrinkable arm to the widely used baseline flag. We believe that our shrinkable arm is more understandable than the flag.

In the study, we only inform the participant that they are a casual pedestrian walking along the street. Then, we asked them to wear a VR headset (Figure 7, right). After the participant initially saw the 360-degree photo, we asked them to speak freely about their next actions, behaviors, or thoughts. If they are puzzled or wish to continue walking forward without taking any action, we ask them to tell us to stop the study. This is an observational experiment. Unlike scenario 1, where the car would do a certain movement (i.e., fixed video flow), scenario 2 has highly flexible pedestrian movement. That is, if the video starts with a certain movement, the participants can guess the next event, which impacts their thought process. Hence, we would like to increase their freedom in thinking about their next behaviors by providing a static 360-degree image instead of video.

Consider that once the participants understand the delivery robot's request, there is no need for multiple tasks in this study. Since we would like to avoid a learning effect by changing the order of presenting the scenarios, we used a between-subjects design. That is, each participant completed only one task for one condition. The experimenter noted down everything that the participants said after putting on the VR headset until they decided to stop. Following the experiment, the experimenter studied the recorded utterances to determine whether the participant understood the delivery robot's request to take action or reasons if they refused to assist. Following the two studies, participants were invited to a semi-structured interview to explain their decisions and thoughts.

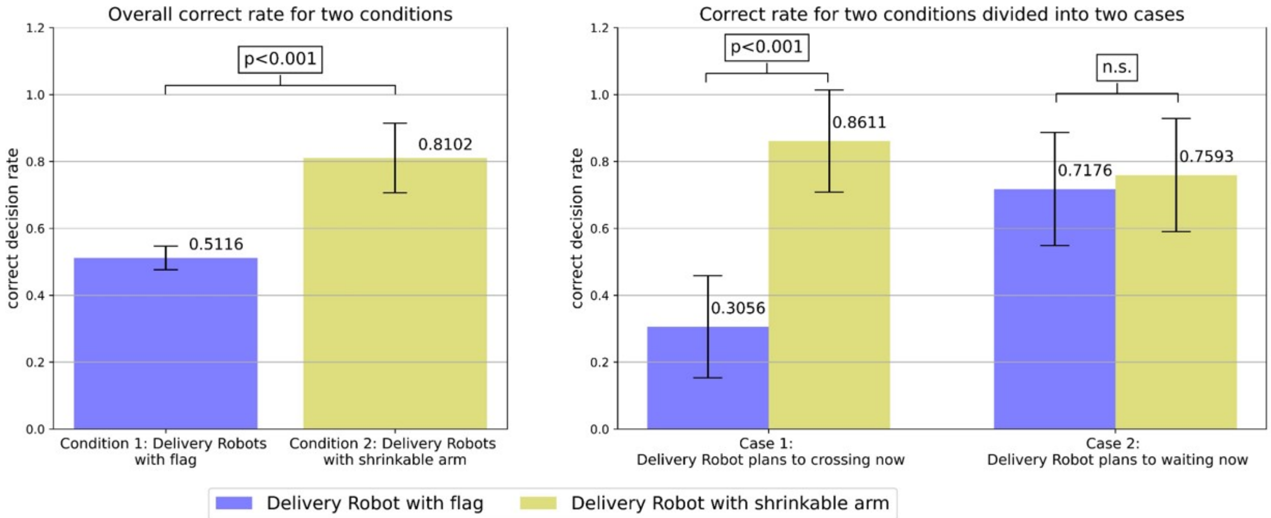


Figure 6: the overall correct decision rate for two conditions on the left, and the correct decision rate divided into two cases.

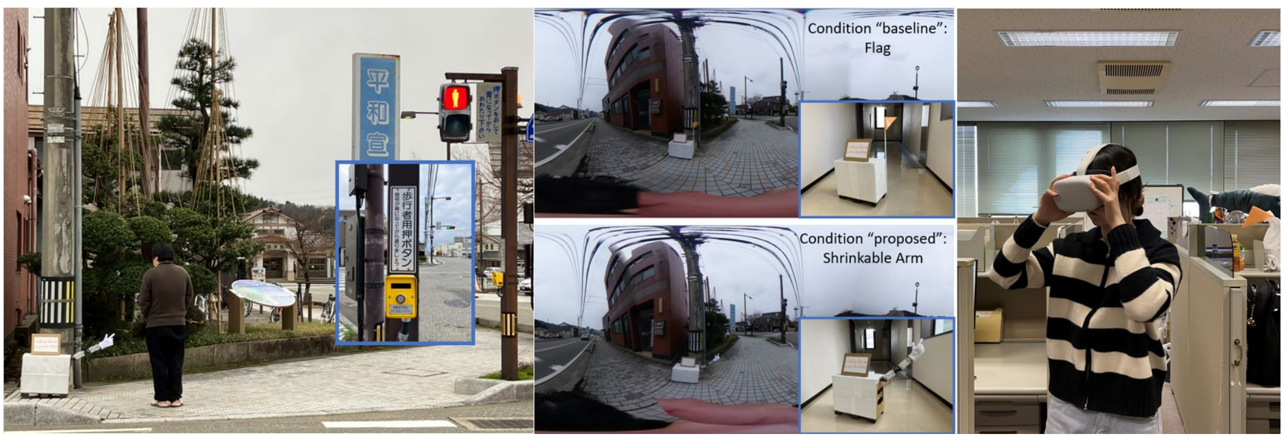


Figure 7: Left is the overall view of the scenario from the third-person perspective. The 360 degrees participant view is taken from the person next to the robot in this photograph. Middle is two photos for two conditions. Right shows that the participant is wearing a VR headset and conducting the study two.

6.2 Result: the participants' action in three categories.

The on-site researcher noted down participants' words and categorized them into three types. In type 1 (the blue area on the pie chart in Figure 8), participants did not see the delivery robot and responded to walk directly. In type 2 (the orange area on the pie chart), the participant discovered the delivery robot, either the flag (baseline) or the shrinkable arm with the pointing gesture hand (proposed), but they were unable to understand the request. In type 3 (green area on the pie chart), the participant saw the delivery robot, comprehended the request, and replied to press the button on the traffic light to help the delivery robot cross the street.

Each type represents to what extent each participant understands the request of the robot. Type 1 is zero understandability, type 2 is minimum understandability, and type 3 is full understandability. We analyzed using a repeated measures ANOVA to compare the effect of people's understandability of the robot's intent for each condition (flag and shrinkable arm) with the planned contrast (shrinkable arm is much more understandable than flag). We found a large-sized main effect of people's understandability of the robot's intent ($F = (1, 8) = 16$, $P = 0.00395$, $\eta^2 = 0.17$) with the planned contrast of a shrinkable arm ($M = 1.3333$) being more understandable than a flag ($M = 0.6667$).

In type 1, four participants did not notice the delivery robot with the flag, and two did not notice the robot with the shrinkable arm. The results showed that our arm-based eHMI can draw more pedestrians' attention while they are walking. In types 2 and 3, from the pool of people who noticed the robot, one out of five (*20%*) participants with the baseline (the robot with the flag) and five out of seven (*71.4%*) with the proposed method (the robot with

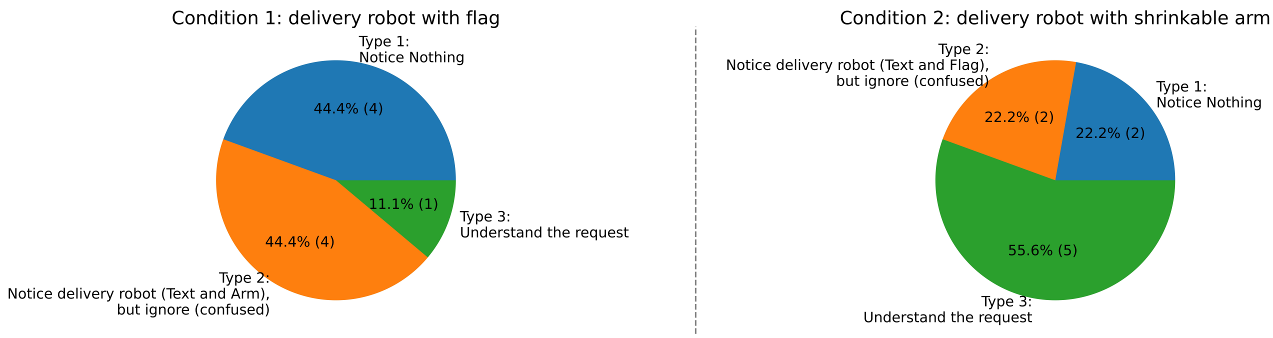


Figure 8: participants' action in three categories in pie chart

the shrinkable arm) understood the delivery robot's request and expressed a desire a desire to help them.

7 DISCUSSION

Based on the quantitative data and post-interview responses, we found that our participants typically do not choose to stop the car unless specifically requested to do so. We hypothesized that it was because of the participants' differing attitudes about pedestrians and delivery robots: "*If a pedestrian is waiting here, I will always stop. However, if it's a small box, I would sooner pass than stop.*" [P3]. At the same time, the participants mentioned other reasons for their unstopping choice: "*If it's a pedestrian, when I make gestures or make eye contact in the car, they will respond to me, but I'm not sure if the car can detect it or if they have no reaction.*" [P14]. These interview results validated the importance of employing eHMI for delivery robots to let drivers know of the delivery robot's request or that the delivery robot has detected their behavior and understood their intention. Based on the interview results, several people demonstrate how the arm is advantageous in our test scenarios. We conclude with four reasons:

First, using the arm can encourage participants to think based on common sense. An arm can induce people to unconsciously think of their daily life behaviors since it resembles a real human arm, even if the whole delivery robot is not human-shaped. During the interview, seven participants stated that it was common sense for them that "*seeing a little box reminded them of a child crossing the street with its hand outstretched.*" [P12]

Second, using the arm, even though people have different interpretations, they make the correct decision. Positive findings were obtained from our tests with scenario 1. However, part of the reason is that "children crossing the road with their hands raised" is a widespread practice in Japan [33]. Since the road crossing situation and the evaluation were taken in Japan, we chose the arm-lowering/raising gesture instead of other common gestures (e.g., waving) in Study 1. In addition, waving can have a different meaning, such as "Please come here!" when facing a taxi. We noticed that people from diverse backgrounds have varied driving habits and experiences; the outcomes of this experiment may vary. For example, in scenario 1, "raising hands" can mean a request to cross the road, a strategy to protect oneself, or a call to others. All these interpretations lead the driver to believe that they must stop. We hypothesized that the reason is that "raising hands" is global common sense to draw people's attention. However, to receive a generalized result, it is worthy to conduct a demography investigation of how people from different cultures apply the gesture in their daily on-road lives and carry out a comparison study with various gestures.

Third, when using the arm, it is more visible. In scenario 1, participants indicated that the flag makes them feel as if there is something here. However, most of the time, the flag is indistinguishable from the surrounding landscape. "*The flag is too small; when it does not have any behavior, I will pass it because it just merges into the environment.*" [P15]. This is also evident in scenario 2, where four out of nine participants who tested with baseline photo did not find the robot. If using the LED board, which is more noticeable, the result may differ. However, handwritten texts would not impact our comparison as both cases are the same for the display. If we had LED displays for both, it would not be much different since we are comparing the flag and the shrinkable arm. In addition, we cannot say that our shrinkable arm has no impact on the LED display, as we did not compare.

Fourth, using the arm, the request shown can be more specific. We analyzed comments from participants who noticed the robot but did not understand the robot's request. Some participants ignored the robot and stated, "*I do not believe it is asking for my assistance.*" [P7] and some could not understand where the button is: "*I am attempting to locate the button on the box itself or around the box.*" [P15] That is, we identified the arm's special feature. Unlike other eHMIs, which primarily draw attention to the delivery robot itself, the arm can direct people's attention to a specific location by pointing ("look at that"). As "help me press the button" is ambiguous (i.e., which button and where), the delivery robot needs to direct pedestrians' attention to the traffic light (rather than itself, the delivery robot) to specify which button to press. The arm becomes dominant in scenario 2.

8 LIMITATIONS AND FUTURE WORK

This study is the first stage in determining the possibility of using arms to display the request on a delivery robot. In our pre-video survey, we identified three sorts of limitations and found various examples of scenarios. We simply chose two scenarios. However, according to our findings, we need to explore more potential difficulties and develop various types of testbed scenarios for fine-tuning user studies.

Our prototype is currently too fragile to hold multiple electronics (e.g., a heavy battery, a monitor) along with the robotic arm. To trade off, we used cardboard to display the message in Study 2. In addition, in the evaluation, we used the 360-degree photo (not video) in Study 2. That is, the robot and the eHMI are all static, which has limitations in indicating intentions. In the future, it will be worthwhile to conduct the study in a natural setting where people interact with the delivery robot continuously. Since real robots can hold batteries and LED displays on them, we can show text messages on the LED screen. In addition, we can modify the color or brightness of the text so that it expresses the message emotionally and visibly. Meanwhile, field study can enable the participant to meet the robot without prior knowledge, where we can receive a more authentic result. That is, we can observe how the participant perceives and reacts to this delivery robot in the real world toward the continuous arm movement and LED display with animations.

Our prototype is well-suited for delivery robots up to 50–60 cm tall, but may not be suitable for robots taller than 1 m. However, our approach and evaluation demonstrate that the arm can be useful as an eHMI. Meanwhile, the current arm prototype is built using 3D printing and handmade origami. To increase its robustness and durability in the future, it is worthwhile to improve the arm to meet industrial standard by analyzing existing robots in the service industry to identify an appropriate range of the arm's length and feasible materials. In this study, we only applied one limb motion (raising up) and one hand gesture (pointing). The arm's flexibility distinguishes it from other anthropomorphic modalities (such as eyes). Because the upper and lower limbs can move with six degrees of freedom and the hand has five fingers, combining them allows for significantly more forms of information to be expressed than other modalities. More comprehensive research on the arm and hand should be undertaken in the future.

9 CONCLUSION

In this study, we present an arm-based eHMI to address issues by leveraging people's common sense in gestures and body motion. We build a delivery robot prototype with a shrinkable robotic arm through an iterative in-house workshop. The evaluation results show that the arm can draw more attention and convey detailed information. This is an initial study in which we demonstrate the benefits of using an arm as an eHMI on a delivery robot in two scenarios. We believe that our work would shed light on other arm-based eHMI studies in autonomous vehicles by incorporating more complicated arm motions and gestures or applying arms to various types of autonomous driving vehicles.

ACKNOWLEDGMENTS

This work was supported by JST CREST Grant Number JPMJCR22M4, Japan.

REFERENCES

[1] Tier 4. 2023. *When will automatic delivery robots that solve the labor shortage in logistics start running around town?* Retrieved April 11, 2024 from https://www.walkingspacedx.go.jp/post-197/

[2] Ammar Al-Taie, Frank Pollick, and Stephen Brewster. 2022. Tour de Interaction: Understanding Cyclist-Driver Interaction with Self-Reported Cyclist Behaviour. In *Adjunct Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* 127–131.

[3] Christopher E Anderson, Amanda Zimmerman, Skylar Lewis, John Marmion, and Jeanette Gustat. 2019. Patterns of cyclist and pedestrian street crossing behavior and safety on an urban greenway. *International journal of environmental research and public health* 16, 2 (2019), 201.

[4] Pavlo Bazilinskyy, Dimitra Dodou, and Joost De Winter. 2019. Survey on eHMI concepts: The effect of text, color, and perspective. *Transportation research part F: traffic psychology and behaviour* 67 (2019), 175–194.

[5] Pavlo Bazilinskyy, Dimitra Dodou, and Joost De Winter. 2020. External Human-Machine Interfaces: Which of 729 colors is best for signaling 'Please (do not) cross'?. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC).* IEEE, 3721–3728.

[6] Pavlo Bazilinskyy, Lars Kooijman, Dimitra Dodou, Kirsten Mallant, Victor Roosens, Marloes Middelweerd, Lucas Overbeek, and Joost de Winter. 2022. Get out of the way! Examining eHMIs in critical driver-pedestrian encounters in a coupled simulator. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* 360–371.

[7] Juan Carmona, Carlos Guindel, Fernando Garcia, and Arturo de la Escalera. 2021. eHMI: Review and guidelines for deployment on autonomous vehicles. *Sensors* 21, 9 (2021), 2912.

[8] Justine Cassell, Tim Bickmore, Lee Campbell, Hannes Vilhjalmsson, Hao Yan, et al. 2000. Human conversation as a system framework: Designing embodied conversational agents. *Embodied conversational agents* (2000), 29–63.

[9] Chia-Ming Chang, Koki Toda, Takeo Igarashi, Masahiro Miyata, and Yasuhiro Kobayashi. 2018. A video-based study comparing communication modalities between an autonomous car and a pedestrian. In *Adjunct proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications.* 104–109.

[10] Chia-Ming Chang, Koki Toda, Daisuke Sakamoto, and Takeo Igarashi. 2017. Eyes on a Car: an Interface Design for Communication between an Autonomous Car and a Pedestrian. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications.* 65–73.

[11] Vivienne Bihe Chi, Shashank Mehrotra, Teruhisa Misu, and Kumar Akash. 2024. Should I Help a Delivery Robot? Cultivating Prosocial Norms through Observations. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems.* 1–7.

[12] Jen Jen Chung, Carrie Rebhuhn, Connor Yates, Geoffrey A Hollinger, and Kagan Tumer. 2019. A multiagent framework for learning dynamic traffic management strategies. *Autonomous Robots* 43 (2019), 1375–1391.

[13] Mark Colley and Enrico Rukzio. 2020. A design space for external communication of autonomous vehicles. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* 212–222.

[14] Mark Colley and Enrico Rukzio. 2020. Towards a design space for external communication of autonomous vehicles. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–8.

[15] Koen De Clercq, Andre Dietrich, Juan Pablo Núñez Velasco, Joost De Winter, and Riender Happee. 2019. External human-machine interfaces on automated vehicles: Effects on pedestrian crossing decisions. *Human factors* 61, 8 (2019), 1353–1370.

[16] Joost de Winter and Dimitra Dodou. 2022. External human–machine interfaces: Gimmick or necessity? *Transportation research interdisciplinary perspectives* 15 (2022), 100643.

[17] Debargha Dey, Azra Habibovic, Melanie Berger, Devanshi Bansal, Raymond H Cuijpers, and Marieke Martens. 2022. Investigating the Need for Explicit Communication of Non-Yielding Intent through a Slow-Pulsing Light Band (SPLB) eHMI in AV-Pedestrian Interaction. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* 307–318.

[18] Debargha Dey, Azra Habibovic, Andreas Löcken, Philipp Wintersberger, Bastian Pfleging, Andreas Riener, Marieke Martens, and Jacques Terken. 2020. 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* 7 (2020), 100174.

[19] Debargha Dey, Azra Habibovic, Bastian Pfleging, Marieke Martens, and Jacques Terken. 2020. Color and animation preferences for a light band eHMI in interactions between automated vehicles and pedestrians. In *Proceedings of the 2020 CHI conference on human factors in computing systems.* 1–13.

[20] Debargha Dey, Brent Temmink, Daan Sonnemans, Karijn Den Teuling, Lotte van Berkel, and Bastian Pfleging. 2021. FlowMotion: Exploring the Intuitiveness of Fluid Motion Based Communication in eHMI Design for Vehicle-Pedestrian Communication. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* 128–131.

[21] Debargha Dey and Jacques Terken. 2017. Pedestrian interaction with vehicles: roles of explicit and implicit communication. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications.* 109–113.

[22] Dimensions.com. 2023. *Autonomous Delivery Vehicles.* Retrieved April 14, 2024 from https://www.dimensions.com/collection/autonomous-delivery-vehicles

[23] Giulia Dominijanni, Solaiman Shokur, Gionata Salvietti, Sarah Buehler, Erica Palmerini, Simone Rossi, Frederique De Vignemont, Andrea d'Avella, Tamar R Makin, Domenico Prattichizzo, et al. 2021. Enhancing human bodies with extra robotic arms and fingers: The Neural Resource Allocation Problem. *arXiv preprint arXiv:2103.17252* (2021).

[24] Stefanie M Faas, Lesley-Ann Mathis, and Martin Baumann. 2020. External HMI for self-driving vehicles: Which information shall be displayed? *Transportation research part F: traffic psychology and behaviour* 68 (2020), 171–186.

[25] Ylva Ferstl, Sean Thomas, Cédric Guiard, Cathy Ennis, and Rachel McDonnell. 2021. Human or Robot? Investigating voice, appearance and gesture motion realism of conversational social agents. In *Proceedings of the 21st ACM international conference on intelligent virtual agents.* 76–83.

[26] Paul DS Fink, Velin Dimitrov, Hiroshi Yasuda, Tiffany L Chen, Richard R Corey, Nicholas A Giudice, and Emily S Sumner. 2023. Autonomous is not enough: designing multisensory Mid-Air gestures for vehicle interactions among people with visual impairments. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* 1–13.

[27] Lex Fridman, Bruce Mehler, Lei Xia, Yangyang Yang, Laura Yvonne Facusse, and Bryan Reimer. 2017. To walk or not to walk: Crowdsourced assessment of external vehicle-to-pedestrian displays. *arXiv preprint arXiv:1707.02698* (2017).

[28] Danilo Gallo, Prescillia Leslie Bioche, Jutta Katharina Willamowski, Tommaso Colombino, Shreepriya Gonzalez-Jimenez, Herve Poirier, and Cecile Boulard. 2023. Investigating the Integration of Human-Like and Machine-Like Robot Behaviors in a Shared Elevator Scenario. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction.* 192–201.

[29] Steven R Gehrke, Christopher D Phair, Brendan J Russo, and Edward J Smaglik. 2023. Observed sidewalk autonomous delivery robot interactions with pedestrians and bicyclists. *Transportation research interdisciplinary perspectives* 18 (2023), 100789.

[30] Xinyue Gui, Koki Toda, Stela Hanbyeol Seo, Chia-Ming Chang, and Takeo Igarashi. 2022. "I am going this way": Gazing Eyes on Self-Driving Car Show Multiple Driving Directions. In *Proceedings of the 14th international conference on automotive user interfaces and interactive vehicular applications.* 319–329.

[31] Xinyue Gui, Koki Toda, Stela Hanbyeol Seo, Felix Martin Eckert, Chia-Ming Chang, Xiang'Anthony Chen, and Takeo Igarashi. 2023. A field study on pedestrians' thoughts toward a car with gazing eyes. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems.* 1–7.

[32] Gianpaolo Gulletta, Wolfram Erlhagen, and Estela Bicho. 2020. Human-like arm motion generation: A review. *Robotics* 9, 4 (2020), 102.

[33] irasutoya. 2015. *Illustration of elementary school students crossing the crosswalk with their hands raised.* Retrieved April 11, 2024 from https://www.irasutoya.com/2015/12/blog-post_416.html

[34] TIER IV. 2024. *Media.* Retrieved April 11, 2024 from https://tier4.jp/en/media/

[35] Jong-Hann Jean, Chen-Fu Wei, Zheng-Wei Lin, and Kuang-Yow Lian. 2012. Development of an office delivery robot with multimodal human-robot interactions. In *2012 Proceedings of SICE Annual Conference (SICE).* IEEE, 1564–1567.

[36] Shyam Sundar Kannan, Ahreum Lee, and Byung-Cheol Min. 2021. External Human-Machine Interface on Delivery Robots: Expression of Navigation Intent of the Robot. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN).* IEEE Press, 1305–1312.

[37] Sotaro Kita. 2020. Cross-cultural variation of speech-accompanying gesture: A review. *Speech Accompanying-Gesture* (2020), 145–167.

[38] Mikiya Kusunoki, Linh Viet Nguyen, Hsin-Ruey Tsai, Haoran Xie, et al. 2024. Integration of Origami Twisted Tower to Soft Mechanism Through Rapid Fabrication Process. In *2024 IEEE/SICE International Symposium on System Integration (SII).* IEEE, 01–02.

[39] Mikiya Kusunoki and Haoran Xie. 2023. UX Study for Origami-Inspired Foldable Robotic Mechanisms. In *Proceedings of the Asian HCI Symposium 2023.* 28–34.

[40] Daegyu Lee, Gyuree Kang, Boseong Kim, and D Hyunchul Shim. 2021. Assistive delivery robot application for real-world postal services. *IEEE Access* 9 (2021), 141981–141998.

[41] Kiju Lee, Yanzhou Wang, and Chuanqi Zheng. 2020. Twister hand: Underactuated robotic gripper inspired by origami twisted tower. *IEEE Transactions on Robotics* 36, 2 (2020), 488–500.

[42] Rui Li, Hongyu Wang, and Zhenyu Liu. 2021. Survey on mapping human hand motion to robotic hands for teleoperation. *IEEE Transactions on Circuits and Systems for Video Technology* 32, 5 (2021), 2647–2665.

[43] Tao Liu, Yanzhou Wang, and Kiju Lee. 2017. Three-dimensional printable origami twisted tower: Design, fabrication, and robot embodiment. *IEEE Robotics and Automation Letters* 3, 1 (2017), 116–123.

[44] Sheng-Yen Lo and Han-Pang Huang. 2016. Realization of sign language motion using a dual-arm/hand humanoid robot. *Intelligent Service Robotics* 9 (2016), 333–345.

[45] Andreas Löcken, Carmen Golling, and Andreas Riener. 2019. How should automated vehicles interact with pedestrians? A comparative analysis of interaction concepts in virtual reality. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications.* 262–274.

[46] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. 2018. Communicating awareness and intent in autonomous vehicle-pedestrian interaction. In *Proceedings of the 2018 CHI conference on human factors in computing systems.* 1–12.

[47] Jennifer E Martinez, Dawn VanLeeuwen, Betsy Bender Stringam, and Marlena R Fraune. 2023. Hey? ! What did you think about that Robot? Groups Polarize Users' Acceptance and Trust of Food Delivery Robots. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction.* 417–427.

[48] Miraikan. 2023. *Living with 'Weak Robots'.* Retrieved April 14, 2024 from https://www.miraikan.jst.go.jp/en/exhibitions/future/hellorobots/?tabs=2#lab-2

[49] Bilge Mutlu and Jodi Forlizzi. 2008. Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction.* 287–294.

[50] Cassidy Myers, Thomas Zane, Ron Van Houten, and Vincent T Francisco. 2022. The effects of pedestrian gestures on driver yielding at crosswalks: A systematic replication. *Journal of applied behavior analysis* 55, 2 (2022), 572–583.

[51] Yukiko I Nakano, Fumio Nihei, Ryo Ishii, and Ryuichiro Higashinaka. 2024. Selecting Iconic Gesture Forms Based on Typical Entity Images. *Journal of Information Processing* 32 (2024), 196–205.

[52] Huynh AD Nguyen and Quang P Ha. 2023. Robotic autonomous systems for earthmoving equipment operating in volatile conditions and teaming capacity: a survey. *Robotica* 41, 2 (2023), 486–510.

[53] Yoichi Ochiai and Keisuke Toyoshima. 2011. Homunculus: the vehicle as augmented clothes. In *Proceedings of the 2nd Augmented Human International Conference.* 1–4.

[54] Divyasha Pahuja, Saahil Sabnis, and Uday Nair. 2024. Delivery Bot: Enhancing Pedestrian Awareness, Willingness and Ability to Help Delivery Robots Encountering Obstructions. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction.* 1249–1252.

[55] Xiang Pan, Malcolm Doering, and Takayuki Kanda. 2024. What Is Your Other Hand Doing, Robot? A Model of Behavior for Shopkeeper Robot's Idle Hand. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction.* 552–560.

[56] Hannah RM Pelikan and Malte F Jung. 2023. Designing robot sound-in-interaction: The case of autonomous public transport shuttle buses. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction.* 172–182.

[57] Hannah RM Pelikan, Stuart Reeves, and Marina N Cantarutti. 2024. Encountering Autonomous Robots on Public Streets. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction.* 561–571.

[58] Dirk Rothenbücher, Jamy Li, David Sirkin, Brian Mok, and Wendy Ju. 2015. Ghost driver: a platform for investigating interactions between pedestrians and driverless vehicles. In *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* 44–49.

[59] Elaheh Sanoubari, Stela H Seo, Diljot Garcha, James E Young, and Verónica Loureiro-Rodríguez. 2019. Good Robot Design or Machiavellian? An In-The-Wild Robot Leveraging Minimal Knowledge of Passersby's Culture. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* IEEE, 382–391.

[60] Stela H Seo, Jihyang Gu, Seongmi Jeong, Keelin Griffin, James E Young, Andrea Bunt, and Susan Prentice. 2015. Women and Men Collaborating with Robots on Assembly Lines. (2015).

[61] Jinjuan She, Jack Neuhoff, and Qingcong Yuan. 2021. Shaping pedestrians' trust in autonomous vehicles: an effect of communication style, speed information, and adaptive strategy. *Journal of Mechanical Design* 143, 9 (2021), 091401.

[62] the japan times. 2024. *Uber Eats Japan begins deliveries with self-driving robots.* Retrieved April 11, 2024 from https://www.japantimes.co.jp/business/2024/03/06/tech/inoue-uber-eats-robot/

[63] TheTechAnonGuy. 2024. *Willow X.* Retrieved April 11, 2024 from https://twitter.com/TheTechAnonGuy/status/1770890716585083038

[64] Andrea Thomaz. 2023. Robots in Real Life: Putting HRI to Work. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction.* 3–3.

[65] Yiyuan Wang, Senuri Wijenayake, Marius Hoggenmüller, Luke Hespanhol, Stewart Worrall, and Martin Tomitsch. 2023. My eyes speak: Improving perceived sociability of autonomous vehicles in shared spaces through emotional robotic eyes. *Proceedings of the ACM on Human-Computer Interaction* 7, MHCI (2023), 1–30.

[66] Yuto Yamaji, Taisuke Miyake, Yuta Yoshiike, P Ravindra S De Silva, and Michio Okada. 2011. Stb: Child-dependent sociable trash box. *International Journal of Social Robotics* 3 (2011), 359–370.

[67] Zuozhong Yin, Jihong Liu, Bin Chen, and Chuanjun Chen. 2021. A delivery robot cloud platform based on microservice. *Journal of Robotics* 2021, 1 (2021), 6656912.

[68] Xinyan Yu, Marius Hoggenmüller, and Martin Tomitsch. 2024. From Agent Autonomy to Casual Collaboration: A Design Investigation on Help-Seeking Urban Robots. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–14.

[69] Jianqi Zhang, Xu Yang, Wei Wang, Jinchao Guan, Ling Ding, and Vincent CS Lee. 2023. Automated guided vehicles and autonomous mobile robots for recognition

and tracking in civil engineering. *Automation in Construction* 146 (2023), 104699.

A APPENDIX A: THE VIDEO CLIP LIST

Here is the video list containing 50 clips with their titles:

A.0.1 Clip 1. Food delivery robots under attack from vandals, thieves

A.0.2 Clip 2. A food delivery robot was seen driving through a taped off crime scene in LA

A.0.3 Clip 3. hard life of a robot trying to cross the street

A.0.4 Clip 4. robot survival: can food delivery robot survive in LA

A.0.5 Clip 5. Robot delivery: Isaish food delivery robot picks up and order

A.0.6 Clip 6. Man picks fight with food delivery robot in Hollywood

A.0.7 Clip 7. Are Police spying on you with delivery robots?

A.0.8 Clip 8. Delivery robots zip around the streets of Cambridge

A.0.9 Clip 9. I got my first robot food delivery

A.0.10 Clip 10. AI FAILURE: Delivery Robot Knocks Itself Out w/The Aid of Artificial Intelligence

A.0.11 Clip 11. The first wave of urban robots is here | Challengers

A.0.12 Clip 12. Self Driving Robots: The SOLUTION to Last-Mile Delivery?

A.0.13 Clip 13. Robot cause tension

A.0.14 Clip 14. Japan's best robot restaurant

A.0.15 Clip 15. robot tours walk of fame

A.0.16 Clip 16. Coop shopping delivery robot invading cross gates street in Leeds

A.0.17 Clip 17. robot scam

A.0.18 Clip 18. Found a food delivery robot

A.0.19 Clip 19. delivery robot using green cross code

A.0.20 Clip 20. Helping a delivery cab cross the road

A.0.21 Clip 21. 'Kawaii' food delivery robot service launches in Tokyo

A.0.22 Clip 22. Robots are delivering food to your door

A.0.23 Clip 23. How Austin's new food-delivery robots work and why some are speaking out

A.0.24 Clip 24. Pleasanton becomes second U.S. city to use robots to deliver groceries

A.0.25 Clip 25. Robot food delivery at George Mason could become the future

A.0.26 Clip 26. Self-driving robots making food deliveries

A.0.27 Clip 27. Domino's Droid Robot Delivers Pizza Right To Your Doorstep

A.0.28 Clip 28. A Day in the Life of a Starship Robot

A.0.29 Clip 29. Starship Campus Delivery Service with Robots

A.0.30 Clip 30. Starship Robot Delivery in Sunnyvale, CA

A.0.31 Clip 31. Meals on wheels: testing a pizza delivery robot

A.0.32 Clip 32. Angry delivery robot try to kill itself

A.0.33 Clip 33. Security Robot falls into a pot hole

A.0.34 Clip 34. delivery robots get stuck in snow

A.0.35 Clip 35. Starship Robots and other pavement users (With Audio Description and ASL overlay)

A.0.36 Clip 36. The Robot That's Roaming San Francisco's Streets to Deliver Food | WIRED

A.0.37 Clip 37. Are Robotic Rovers the Future of Food Delivery?

A.0.38 Clip 38. Watch Robot Food Delivery Debut on the University of Kentucky Campus

A.0.39 Clip 39. Ford's Delivery Robot Walks On Two Legs Like A Human

A.0.40 Clip 40. Robots roam the University of Wisconsin campus delivering food to hungry students

A.0.41 Clip 41. Kiwi's robots deliver food to hungry Berkeley students

A.0.42 Clip 42. Robots at your service: Automated food delivery on Wayne State University campus

A.0.43 Clip 43. Watch How These Robots are Revolutionizing Grocery Delivery in Manchester!

A.0.44 Clip 44. Deligo Food Delivery Robot- The New Trend in food distribution in restaurants.

A.0.45 Clip 45. Robots Are Delivering Food on This College Campus

A.0.46 Clip 46. Autonomous pizza delivery robot hits the road

A.0.47 Clip 47. China's Alibaba unveils autonomous logistics robot

A.0.48 Clip 48. Self-driving delivery robots hit Japanese streets | Tech It Out

A.0.49 Clip 49. Rude robot

A.0.50 Clip 50. Starship Robots blocked by DHL Van | Robots deliver food in Milton Keynes