Scenarios Exploration: How AR-Based Speech Balloons Enhance Car-to-Pedestrian Interaction

Xinyue Gui[1(✉)], Chia-Ming Chang[1], Stela H. Seo[2], Koki Toda[1], and Takeo Igarashi[1]

[1] The University of Tokyo, Tokyo, Japan
xinyueguikwei@gmail.com, info@chiamingchang.com, toda.koki@mail.u-tokyo.ac.jp, takeo@acm.org
[2] Kyoto University, Kyoto, Japan
stela.seo@i.kyoto-u.ac.jp

Abstract. Previous studies considered the text-based external human-machine interface (eHMI) as the most effective method for self-driving vehicles. However, classic eHMIs (LED displays, windshields, and grounded projectors) cannot meet these requirements because of various limitations (e.g., placement, space, and limited communication channels). With rapid improvements in the Internet of Things, there are more options to improve communication between self-driving vehicles and pedestrians. We introduce our novel eHMI concept, the AR-based speech balloon, which displays text in a balloon shape and floats around self-driving vehicles through smart glasses. In this paper, we present future research scenarios of people wearing smart glasses. We define the mockups, describe the three main problems, and explain how an AR-based speech balloon can alleviate them in six corresponding scenarios.

Keywords: AR Comic · Car-to-Pedestrian Interaction · Scenario-based design

1 Introduction

Regarding technological convergence, various types of technologies collaborate and improve each other to keep up with cutting-edge innovations. Technological development is a long process; therefore, in the distant future, we foresee a situation in which pedestrians interact with mixed traffic consisting of manual and self-driving vehicles [1]. Current external human- machine interface (eHMI) design has emerged with the development of self-driving vehicles to improve pedestrian safety and communication efficiency in car-to-pedestrian interactions. In the classic eHMI, researchers have investigated different types of communication modalities (e.g., text, animated gestures, eyes, and facial expressions) [2], implemented them on a display or physically on a car, and evaluated them.

C. Stephanidis et al. (Eds.): HCII 2024, CCIS 2118, pp. 223–230, 2024.
https://doi.org/10.1007/978-3-031-61963-2_21

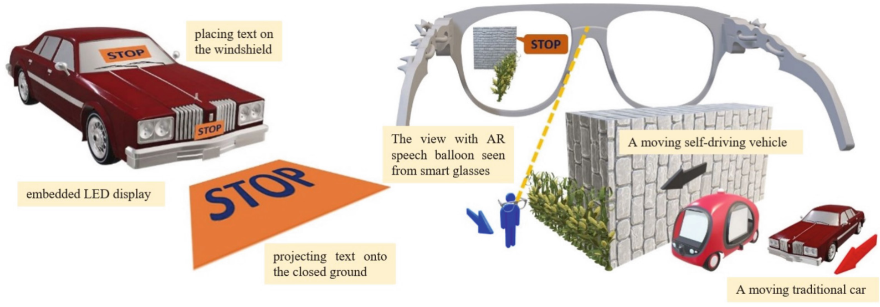

Fig. 1. The classic text-based eHMIs, and our proposal: the self-driving vehicle will stop after detecting an pedestrian, while the traditional car may not. The self-driving vehicle should notify the pedestrian of this potential danger from the blind spot.

According to extensive survey study and comparison experiments, text-based eHMI has been proven to be the most effective approach [3].

However, as technology advances and more information need to be presented, the classic text-based eHMI may not be able to meet this demand. The main problem is the conflict between information complexity and device constraints. The current implementation for displaying text is shown in Fig. 1 left. We define them as "classic text-based eHMIs" through this paper. With advances in technologies, such as the Internet of Things (IoT), object and event detection and response (OEDR), and wireless networks, self-driving vehicles in a network can gather, filter, sort, and selectively display messages for different pedestrians. However, this leads to three major problems for classic text-based eHMIs.

First, there are hardware limitations. Potential dangers, such as an approaching car from a blind spot, can be detected and identified by the sensor system of the self-driving vehicle. However, existing classic text-based eHMIs have limitations (e.g., placement and visibility in bad weather). They cannot provide information to pedestrians in time, particularly when the self-driving vehicle is not visible. Second, information asymmetry exists. As pedestrians cannot see what the self-driving vehicle senses (with robust sensing systems), it may be difficult for pedestrians to trust information from the vehicle. Third, when there are multiple pedestrians or cars, the classic text-based eHMI cannot send these processed messages to the appropriate pedestrian. A self-driving vehicle must pitch specific information to a specific individual. In this paper, we further discuss these problems with six corresponding scenarios and introduce our solution, an AR-based speech balloon that can improve the classic text-based eHMI in car-to-pedestrian interactions.

In manga culture, the speech balloon is a typical comic that can convey different message types (speaking and thinking) and even emotions by employing various balloon shapes [4]. Furthermore, speech balloons can progressively break up long informative words into shorter ones, thereby enhancing communication effectiveness [5]. As AR technology advances, such a futuristic implementation is nearing its realization. For example, AR technology assigns an unique ID to every party in the digital world. Using these IDs, AR objects can be placed on a person with scalable, flexible, and spatial placements. Hence, we used speech balloons with text, which is the most effective modality for self-driving vehicles, as revealed by a previous study.

In this study, we identified problems with the classic text-based eHMI and explained them by illustrating six scenarios. We then described how AR speech balloons could alleviate these problems in these scenarios. Each scenario contained two components: the overall view and smart glasses view. The overall view provided a general explanation of the scenario, including how the cars and pedestrians tend to move. We took a picture from the third-person omniscient perspective. We then added 3D models to illustrate the storyline. The smart glass view displayed the scene through the glasses from the pedestrian's perspective. To create this view, we took an appropriate picture of each scenario from the first-person perspective (pedestrian perspective) in the real world. We then added a speech balloon. In these two perspectives, self-driving vehicles differ (3D model in overall view and real car in smart glasses view).

We created a mockup of the AR speech balloon and used the "basic" bubble shape from comic books, an oval with a "tail" [6]. The tail of the bubble points toward the speaker or sound source so that the person can tell which car is communicating with them. Further, we applied the holographic effect (using Adobe Photoshop) to the basic bubble by adding various levels of motion blur to several layers and rendering with a halftone pattern filter and displacement maps (Figures in the Scenario Illustration section). Because the speech balloon should not interfere with the pedestrian's sight, we chose to make it translucent so that the pedestrian could see other objects in the surrounding environment through the speech balloon using AR glasses. The balloon's color and size should be modified to increase the visibility of the text.

2 Scenario Illustration

In introduction, we lists three problems. For each problem, we explained by two scenarios and the corresponding solution based on the proposed concept.

2.1 Discussion on Problem One: Display Information Both Spatially and in Advance

In terms of placement limitations, classic text-based eHMIs show information in front of the car. However, there are instances in real life where bystanders pass by the car's side or back without seeing the information displayed in front of the vehicle (Fig. 2, upper left). Classic text-based eHMIs are firmly attached to vehicles, leading to two problems. First, the font size and text length cannot exceed vehicle size. Second, the LED display cannot show text separately from the self-driving vehicle. If the vehicle is not visible to the pedestrian, the message on the eHMI is invisible. For S1 (Fig. 2, upper right), a solution (see bottom) can be used to display the speech balloon with the phrase "A car is coming" in front of the pedestrian (the back of the self-driving vehicle).

Fig. 2. The problem and solution for S1 and S2

In scenario S2 (Fig. 2, bottom left), a self-driving vehicle detects a person emerging from the tunnel as it draws near. The pedestrian is unaware of the self-driving vehicle because the wall blocks it. The windshield or LED display cannot separate information from the car. Projectors can display text away from the car. However, projectors have two restrictions. 1) It requires a clear screen or plate to project; thus, the existing eHMI is limited to projection onto the ground. 2) It has strict lighting requirements. To date, experiments with projected eHMIs have been conducted in a dark environment. In our solution (Figure 2, bottom right), self-driving vehicle urges the pedestrian to halt in advance by floating the speech balloon far enough, providing a clearly understandable message, and making it visible to the pedestrian. The text in the speech balloon could be the descriptive word: "A car is coming from left."

2.2 Discussion on Problem Two: Information Asymmetry and the Need to Display Complex Information

In addition to using AR features to address the restrictions of the device (LED and projector), the features of speech balloons have valuable features for optimizing the typical problems in car-to-pedestrian interaction in two major aspects.

Fig. 3. The problem and solution for S3

First, a self-driving vehicle must convey complex information messages to address information asymmetry in a complex scenario.

Fig. 4. The solution for S4 (from left to right): The self-driving vehicle conveys its advisory, description, perception, commanding using a normal-shape, rectangular, cloud-like bubble and jagged-edged shape balloon respectively.

In S3 (Fig. 3, left), when a driving self-driving vehicle planning to go straight detects a wall blocking an approaching traditional car, the latter collides with a walking pedestrian. The pedestrian observes that the self-driving vehicle does not use the turning signal and thus believes that it is safe to proceed to the cross. However, displaying a simple "STOP" as some classic eHMIs may confuse pedestrians, letting them think the self-driving vehicle is malfunctioning. Therefore, the self-driving vehicle should display a command message with an explanation. However, it takes time to read long sentences. Studies have shown that speech balloons can break long sentences into short portions and communicate information more efficiently [5]. By implementing the speech balloon's properties, the self-driving vehicle can display the content in a multi-line format list in order of priority (Fig. 3, right). If time is short, the pedestrian can read only the prior part of them. At the same time, this can reduce the information gap between pedestrians and self-driving vehicle.

To illustrate the second aspect, we need to explain a concept called "message type." Dey et al. [3] discovered several types (intention, instruction, situational awareness, warming, and advisory) and their combinations. Colley and Rukzio [7] constructed a design space with two large groups (explicit and implicit), with each group containing eight message types. Based on this background, we propose an argument that shows only message content without indicating the message type, which is not sufficient for pedestrians. People who have experienced watching anime or manga have a common understanding that different balloon types can convey varying states or messages, such as a character's commands, thoughts, or even emotions. These features can help solve the message ambiguity in car-to-pedestrian interactions (Fig. 4).

2.3 Discussion on Problem Three: Pitch the Message to Specific Pedestrians

This section discusses the interactions between a self-driving vehicle and multiple pedestrians, multiple self-driving vehicles and one pedestrian, and multiple self-driving vehicles and multiple pedestrians (labeled 1-to-N, N-to-1, and N-to-N, respectively) [8].

Individual differences resulted in a 1-to-N situation. For example, there is no absolute "right," "left," "slightly," or "sharply" when the self-driving vehicle displays "I will turn left slightly." The relative position is altered as the pedestrian stands at a new location. In addition, pedestrians from different backgrounds (e.g., nations or ages) may have different preferences for how they want to be informed (i.e., personalization). This can improve communication effectiveness, as users understand and respond faster when reading their native language (such as movie subtitles). Users can set their data in the digital ID, and the smart glasses can display the information corresponding to the user's preferences, such as changing the language display (Fig. 5).

Fig. 5. The solution for S5

In the N-to-1 case, the pedestrian typically receives a filtered message. Using the classic eHMI, the self-driving vehicle in the IoT network can process the message and send it to the pedestrian. The drawback is that the displayed message has a high probability of being blocked when there is a cluster of self-driving vehicles or mixed traffic in the N-to-1 case. In the first section, we discuss a solution: a self-driving vehicle can float speech balloons where pedestrians can see them.

The combination of these two cases (1-to-N, N-to-1) results in an N-to-N condition. We explain this by showing a future self-driving vehicle taxi service application in S6 (Fig. 6). Railway stations, airports, and hotels have designated exits for taxi services. A disorganized taxi service may lead to pandemonium when many customers stream into a parking lot. The IoT system can allocate taxis according to the amount of a customer's luggage, the order in which they arrive at the location, and where they are standing in the parking lot. The vehicles then communicate with the designated customers for the most efficient diversion. Each consumer will only see the speech balloon of the assigned self-driving vehicle to reduce confusion with other vehicles.

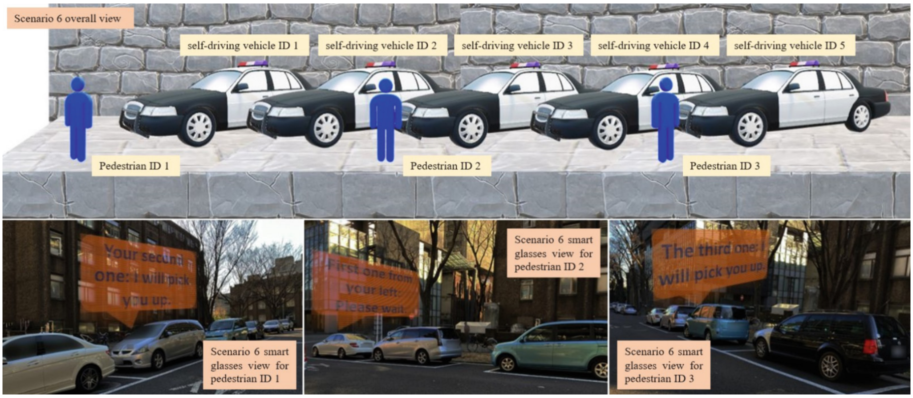

Fig. 6. The solution for S6

3 Conclusion

This paper explores the application scenario in a distant setting with fully developed IoT systems, self-driving vehicles, and AR technology. We began with a text-based eHMI, introduced the concept of an AR speech balloon for self-driving vehicles communicating with pedestrians, and created a holographic-style mockup. We identified three problems with the classic text-based eHMI and constructed six scenarios to illustrate how AR speech balloons can alleviate these problems.

All of the discussions in this paper assume that smart glasses are as ubiquitous as mobile phones. Thus far, there has been little demand from public users for smart glasses. For example, Magic Leap, a famous smart glasses manufacturer, saw only 6,000 sales in its first year of availability, much below the "at least" 1 million units anticipated [9]. In addition to being used by the industry, smart glasses have few customers because of their high price and complicated operation. Nevertheless, we believe that with the development of technology, people's acceptance and demand for new things (AR glasses and autonomous driving) will increase accordingly, such as the recently popular Quest3. We anticipate that AR glasses for traffic might be first adopted by professional road users, such as taxi drivers and delivery drivers (cars, motorbikes, and bicycles). Professional road users spend a long time on the road, and safety and efficiency are critical for their jobs. AR glasses that facilitate communication with autonomous vehicles may be an attractive tool for them.

References

1. Lee, J., Park, W., Lee, S.: Discovering the design challenges of autonomous vehicles through exploring scenari via an immersive design workshop. In: Designing Interactive Systems Conference 2021, DIS 2021. pp. 322–338. Association for Computing Machinery, New York (2021). https://doi.org/10.1145/3461778.3462119
2. Mahadevan, K., Somanath, S., Sharlin, E.: Communicating awareness and intent in autonomous vehicle-pedestrian interaction. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, pp. 1–12. Association for Computing Machinery, New York (2018). https://doi.org/10.1145/3173574.3174003
3. Dey, D., et al.: Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. Trans. Res. Interdiscipl. Perspect. **7**, 100174 (2020)
4. Yamanishi, R., Tanaka, H., Nishihara, Y., Fukumoto, J.: Speech-balloon shapes estimation for emotional text communication. Inform. Eng. Express **3**(2), 1–10 (2017)
5. Kurahashi, T., Sakuma, R., Zempo, K., Mizutani, K., Wakatsuki, N.: Retrospective speech balloons on speech-visible ar via head-mounted display. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct, ISMAR-Adjunct 2018, pp. 423–424. Institute of Electrical and Electronics Engineers (IEEE), Piscataway, NJ, USA (2018). https://doi.org/10.1109/ISMAR-Adjunct.2018.00127
6. Anime Outline. https://www.animeoutline.com/how-to-draw-manga-speech-bubbles/, (Accessed 14 Feb 2024)
7. Colley, M., Rukzio, E.: A design space for external communication of autonomous vehicles. In: 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Automotive. UI 2020, pp. 212–222. Association for Computing Machinery, New York (2020). https://doi.org/10.1145/3409120.3410646
8. Jensen, K.B., Helles, R.: Speaking into the system: social media and many-to-one communication. Eur. J. Commun. **32**(1), 16–25 (2017)
9. VR Times. https://virtualrealitytimes.com/2019/12/06/report-magic-leap-sold-6000-ar-headsets-in-6-months/, (Accessed 14 Feb 2024)