Design of a User State Estimation to Enhance Human-friendly Services

Yuka Sone*, Jinseok Woo**,†[0000−0002−6118−6617], and Yasuhiro Ohyama**

*Sustainable Engineering Program, Graduate School of Engineering,
**Department of Mechanical Engineering, School of Engineering,
*,**Tokyo University of Technology,
1404-1 Katakuracho, Hachioji Tokyo 192-0982, Japan,
†Corresponding author
g512402736@edu.teu.ac.jp, woojsk@stf.teu.ac.jp, ohyama@stf.teu.ac.jp

Abstract. In recent years, the mechanization and automation of the service industry have advanced rapidly. Additionally, smart home appliance systems utilizing IoT technology are becoming increasingly widespread in general households. These systems are standardized based on their services and are operated through devices such as remote controls, executing commands specified by the user. However, users are required to configure the system in advance to align with the service system. In contrast, human-provided services are not limited to predetermined commands but adapt their support patterns based on the recipient's emotions and atmosphere. Replacing humans with robotic systems poses significant challenges in providing services tailored to the user's emotional state or intentions. In light of this, the present study proposes a system that analyzes user gaze during the operation of smart home appliances to construct a service system that is more responsive to human needs. The experimental system is developed within a Mixed Reality(MR) environment, where gaze information is analyzed as users interact with appliances of interest. By incorporating support from a virtual robot partner, the system evaluates how closely the assistance aligns with the user's intent. Through the measurement of gaze movements, the system was able to provide the information users were seeking. It is anticipated that by inferring users' impressions of target objects based on their gaze data, the system will be capable of delivering more advanced and personalized services.

Keywords: Human estimation · Gaze analysis · Human-centric system · Human system interaction.

1 Introduction

In recent years, technologies such as robotics and AI have rapidly advanced, leading to significant automation of services that were traditionally performed by humans. These service systems have become increasingly integrated into everyday life and can be found in a variety of environments, including hotels, restaurants, retail stores, and public transportation [1, 2].

Many of services are implemented through mechanisms like remote controllers, which require customers to follow precise usage instructions tailored to the form of the service. This has led to the issue where some individuals are unable to utilize these services due to their complexity. Moreover, there is a challenge in that robots find it difficult to provide personalized services based on emotional information inferred from human behavior and facial expressions, as humans can do. As an initial step towards providing personalized services tailored to individuals, we propose a system that offers services based on gaze information. For robots to deliver personalized services similar to those provided by humans, it is necessary to measure human movements and infer emotions [3, 4].

Among various expressions of human emotions, eye gaze is a particularly important factor, and research has demonstrated a close relationship between the direction of gaze and the emotions expressed on the face [5–7]. To investigate the relationship between gaze and emotions, we developed a smart home system utilizing mixed reality devices capable of tracking eye movements. Moreover, various studies have been conducted on eye gaze in head-mounted XR devices, including MR and virtual reality systems [8].Using this system, we aim to confirm whether providing services based on gaze information leaves a positive impression on users.

The structure of this paper is as follows. Section 2 describes the design of the virtual space using Mixed Reality (MR) devices and the smart home control system using gaze data. Section 4 discusses the implementation and effectiveness verification of smart home control on MR devices. Section 5 presents the conclusions and future challenges.

2 Human Interface Design based on Mixed Reality Device

Considering future real-world control scenarios, we created a virtual space based on an actual environment, as depicted in Fig. 1. The virtual space primarily consists of five types of objects: windows, lightings, interactive robots, a refrigerator, and tables. This virtual space was developed using Unity Real-Time Development Platform [9], and users can experience it using Mixed Reality (MR) devices such as HoloLens 2. Fig. 2 illustrates the overall system by using HoloLens 2 to control smart home appliances in a real environment.

Using HoloLens 2, we simulate the virtual space and enable control of real-world appliances within this space by utilizing user information. HoloLens 2 tracks the user's gaze and hand movements, toggling the visibility of buttons for appliance control based on gaze movement and enabling interaction by selecting buttons with hand gestures. The appliances that can be controlled in this system include curtains, lighting, and a refrigerator.

Virtual buttons are utilized to control each household appliance. These buttons are categorized into three types: (a) for the refrigerator, (b) for lighting, and (c) for curtains as shown in Fig. 3.

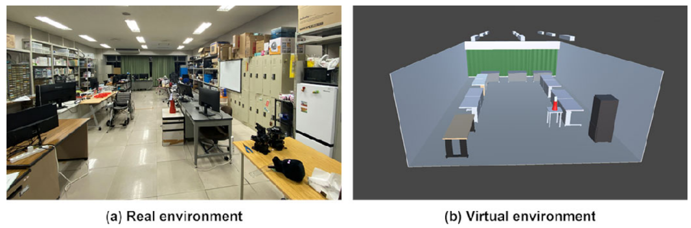

Fig. 1: The Example of Real and Virtual Environment

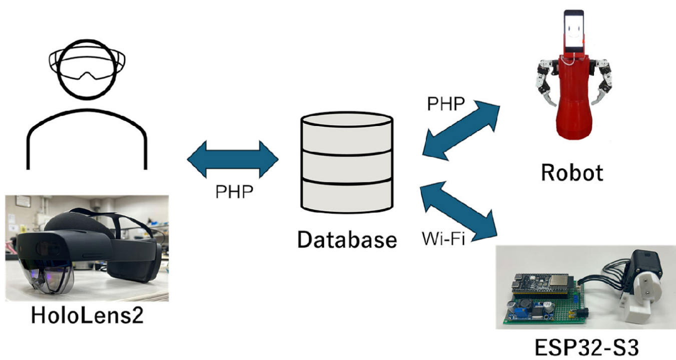

Fig. 2: Overview of the Smart Home System

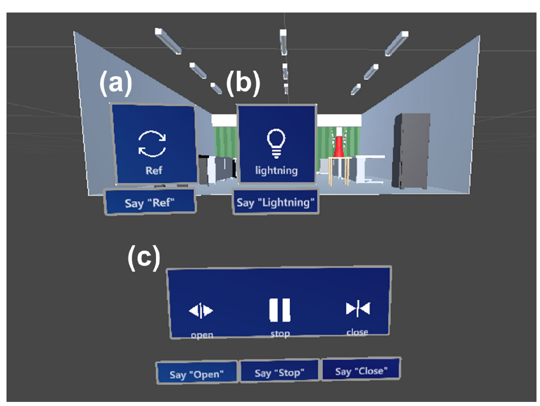

Fig. 3: Buttons for Controlling Appliances in the Virtual Space: (a) Refrigerator, (b) Lighting, (c) Curtain

Each button performs specific operations as described below. First, pressing the button for (a) the refrigerator, switches the scene in the virtual space to display only the refrigerator. In this state, directing gaze to the desired part of the refrigerator opens the door virtually, allowing inspection of its contents.

Next, pressing the button for (b) lighting toggles the ON/OFF state of the lights, as depicted in Fig. 4. Lastly, the button for (c) curtains, is designed with three linked buttons: for open, stop, and close functions. Pressing the open or close buttons causes the curtains to continue moving until the stop button is pressed or they reach fully opened or closed positions. In the virtual space, the operation of the curtains can be observed as shown in Fig. 4

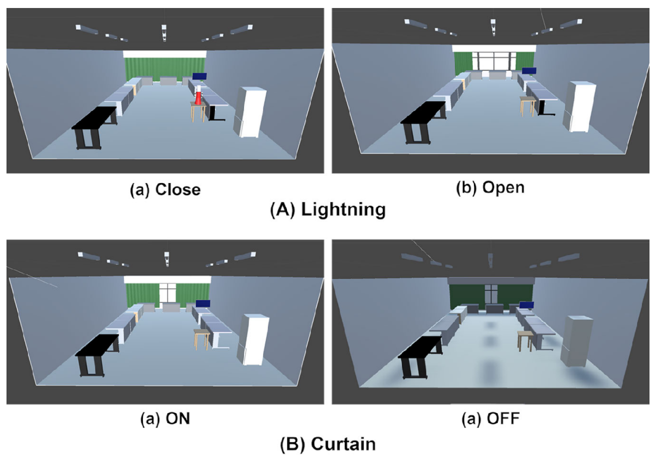

Fig. 4: The example of Lighting Equipment Control and Curtain Control

These buttons are controlled by gaze information. The regions where buttons for operating each appliance appear based on gaze are referred to as gaze detection regions. Initially, buttons do not appear in the virtual space as shown in Figs. **??** and **??**, they appear when gaze is directed towards the gaze detection region of the targeted appliance. Furthermore, buttons are set to disappear again if gaze is not directed towards the detection region or the button for more than 3 seconds. This toggle of button visibility is designed to make it clear where the buttons for desired appliances are located, especially as the number of controllable appliances may increase in the future.

3 A Robot Service System Considering the User's Condition

In addition to smart home appliances and furniture, a robot partner (the red robot) is installed in the virtual space as shown in Fig. 6 [10]. This robot is

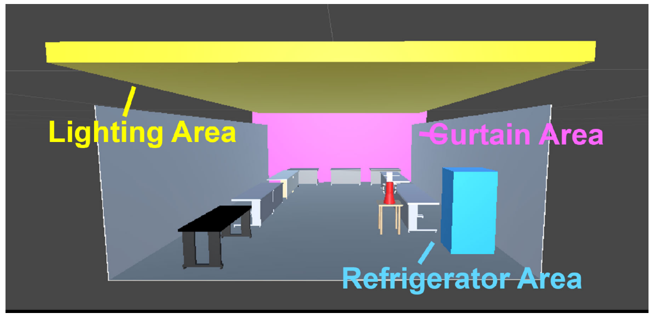

Fig. 5: Gaze Detection Area

introduced into the system as a tool to provide information to users through voice and arm movements. Based on gaze data acquired from HoloLens 2, the robot is programmed to speak about specific topics depending on the direction of the user's gaze.

The robot in the virtual space is modeled after real robots, mimicking movements based on actions that real robots can perform. Furthermore, to emphasize that the robot is interacting verbally, speech bubbles are displayed above the robot at appropriate times, showing the content of what the robot is saying.

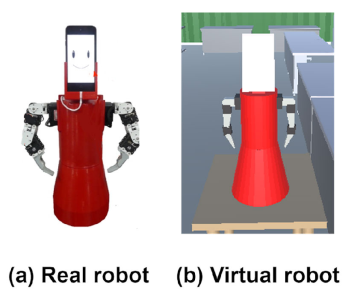

Fig. 6: The Real and Virtual Robot Partners

There are two patterns in which the robot engages with the user: one where it speaks unrelated topics and another where it tailors its speech based on the user's gaze direction. Fig. 7 shows the flowchart for determining the content when the robot engages in unrelated topics. The robot begins speaking 30 seconds after the start of the experiment. This delay aims to prevent participants from involuntarily following the robot's instructions immediately after the experiment begins, allowing them time to decide which appliances to interact with.

Next, Fig. 8 depicts the flowchart for determining the content when the robot engages in conversation tailored to the user's gaze direction.

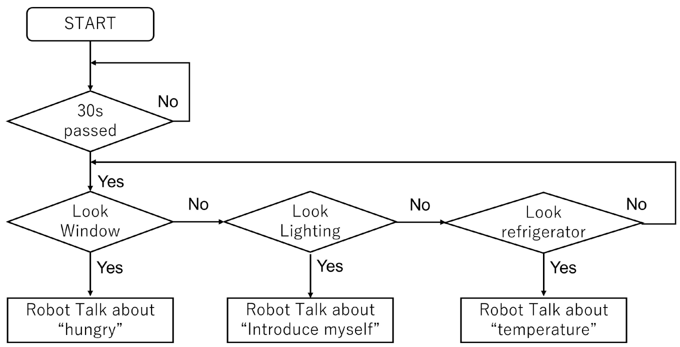

Fig. 7: Flowchart for Scenarios Not Aligned with the Direction of Gaze

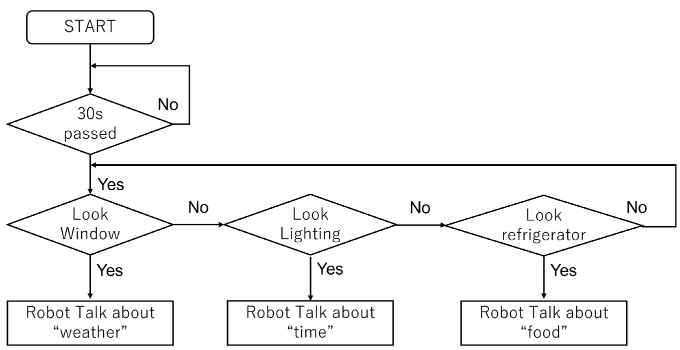

Fig. 8: Flowchart for Scenarios Aligned with the Direction of Gaze

4 Experiment

This chapter will cover experiments using the smart home system and the analysis of experimental results. The experiments were conducted with students at our university and participants were 7 persons in their 20s.

4.1 The Experimental Procedure

As shown in Fig. 9, the experiment procedure is as follows:

1. Explain the usage of HoloLens 2 and the operational procedures of the experiment using a tutorial created specifically for the experiment.
2. Start the experiment and allow participants to freely interact. End the experiment after the experiment was constructed for 1 minute by the least.
3. Conduct a questionnaire after the experiment ends.
4. Conduct the experiment again.
5. Conduct another questionnaire after the second experiment. The first experiment is referred to as Experiment 1 (ex.1), and the second experiment is referred to as Experiment 2(ex.2).

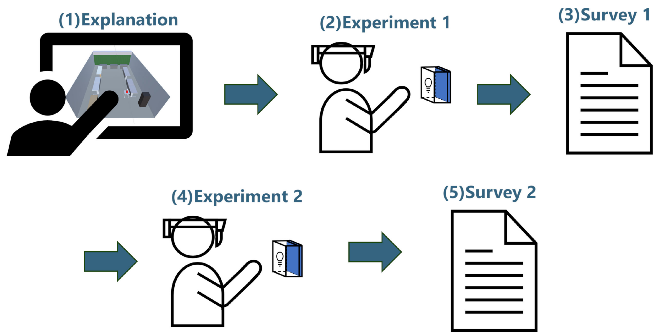

Fig. 9: Experiment Procedure

The difference between Fig. 9 (2) Experiment 1 and Fig. 9 (4) Experiment 2 lies in the content of interactions initiated by the robot in the virtual space during the experiment. In Experiment 1, the robot engages in unrelated topics regardless of the user's gaze direction, while in Experiment 2, the robot is programmed to tailor its conversation based on the user's gaze direction.

4.2 Questionnaire Survey After The Experiment

To survey participants' impressions of each object in the virtual space, a survey was conducted. Participants were asked to rank the five types of objects in order of their interest and allocate points accordingly to each object based on their ranking. Ratings were conducted on a scale of 0 to 4, where an object ranked as most interesting received 4 points and an object ranked as least interesting received 0 points. Higher points indicate a more positive impression, while lower points indicate a more negative impression. The objects included in the survey, as shown in Fig. 10, are: (a) Lightings, (b) Curtain, (c) Robot, (d) Refrigerator, and (e) Tables.

The experiment was conducted with 7 participants. The results of Experiment 1's survey are shown in Fig. 11. The object that received the highest points was (b) Curtain, with an average score of 3.0. The object that received the lowest points was (e) Table, with an average score of 0.4.

Next, the results of Experiment 2 are shown in Fig. 12 .The objects that received the highest points were (b) Curtain and (c) Robot, both with an average score of 2.7. The object that received the lowest points was (e) Table, with an average score of 0.6.

4.3 Analysis of User's Eye Gaze

We compared the gaze data of the subject who gave the most distinctive responses with the subject who gave the most average responses. The subject with

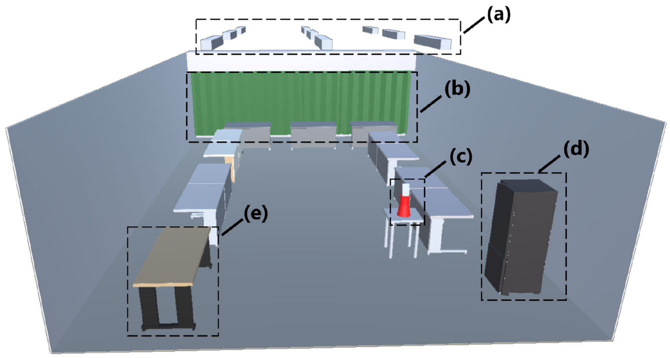

Fig. 10: The Names of Objects Used in the Survey

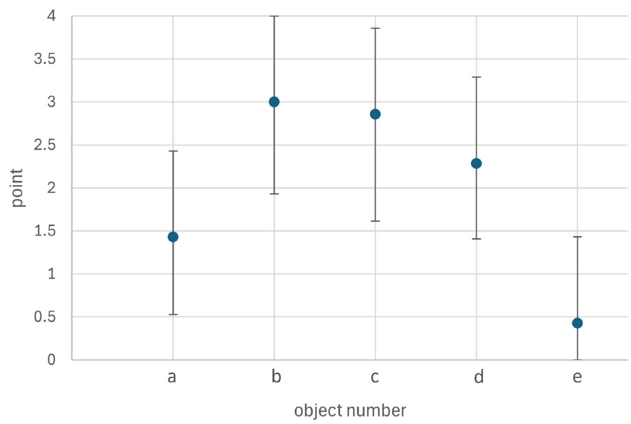

Fig. 11: The results of Experiment 1 Survey

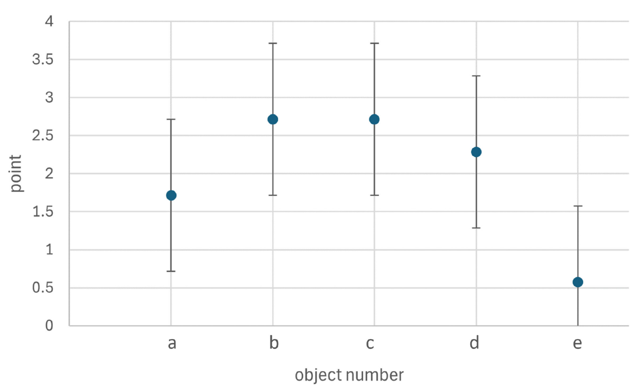

Fig. 12: The results of Experiment 2 Survey

the most distinctive responses is referred to as participant A, and the participant with the most average responses is referred to as participant B.

First, we analyzed the gaze data of participants A and B. Initially, we used the elbow method to determine the number of clusters for clustering. The results of applying the elbow method to the gaze data of participants A and B are shown in the following Fig. 13. Based on these results, the number of clusters (k) was determined as follows: $k = 2$ for participant A in Experiment 1, $k = 3$ for participant B in Experiment 1, $k = 3$ for participant A in Experiment 2, and $k = 3$ for participant B in Experiment 2.

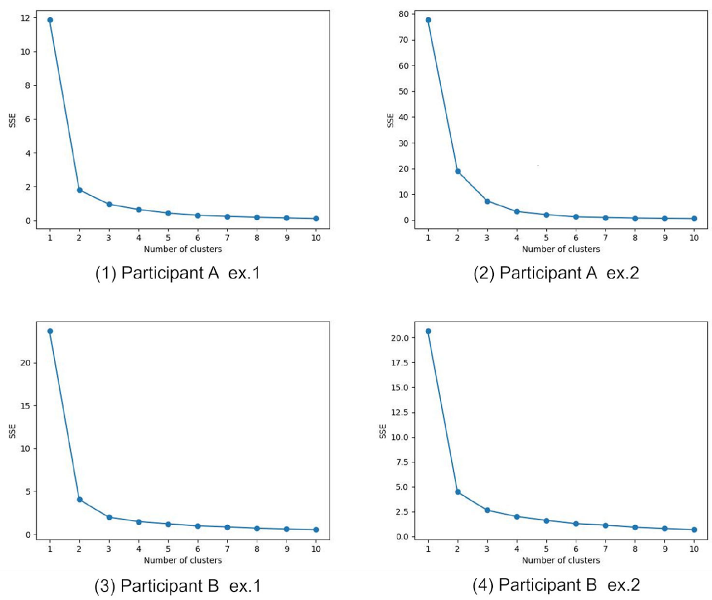

Fig. 13: Elbow Method Results of Experiment 1 (ex.1) and 2 (ex.2)

Using the results of the elbow method, we applied the K-means method to the gaze data of participants A and B. The results of this application are represented on the virtual space used in the experiment as shown in Fig. 14.

Tables 1 and 2 shows the number of times participant A and participant B directed their gaze towards specific objects during the experiments. Participant A indicated interest in objects (c) robot and (e) table in Experiment 1. However, upon examining the distribution of gazes, it becomes apparent that curtain received more gazes than robot or table. In Experiment 2, participant A expressed interest in objects (d) refrigerator and (e) table, but their gazes were not focused much on either, similar to Experiment 1, with curtain attracting more gazes.

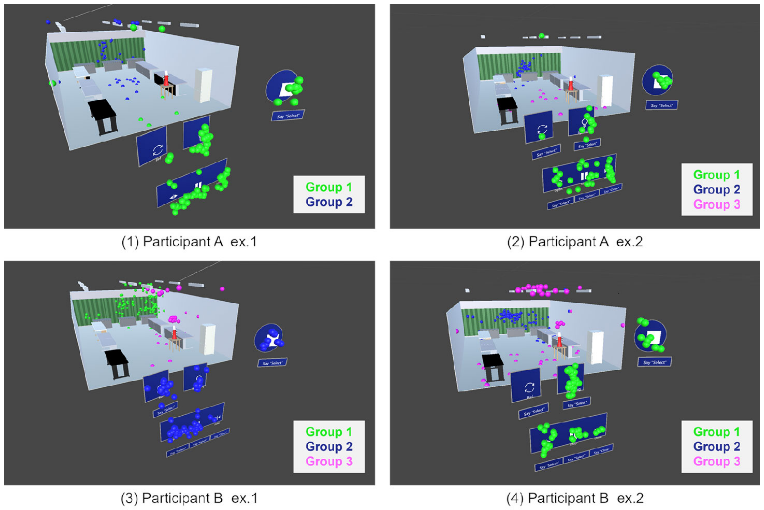

Fig. 14: Visualization of k-means clustering results in virtual space

Participant B expressed interest in objects (b) curtain and (c) robot in Experiment 1. Examining the gaze distribution reveals that, consistent with the survey results, participant B directed their gaze significantly towards curtain and robot. In Experiment 2, participant B continued to show interest in curtain and robot, with a noticeable concentration of gazes around these objects, similar to Experiment 1.

Comparing the gaze distributions of participant A and participant B, it is observed that participant A predominantly focused their gaze on the central part of the x-axis in the virtual space, whereas participant B's gaze was concentrated from the central to right part of the x-axis in the virtual space. Additionally, participant A focused more on the floor area within the virtual space compared to participant B.

Table 1: The Number of Times The Participant's Gaze Stopped

Experiment 1	Participant A	Participant B
lighting	8	12
Curtain	30	102
Robot	0	18
Refrigerator	0	0
Table	0	0
Floor	22	13

Table 2: The Number of Times The Participant's Gaze Stopped

Experiment 2	Participant A	Participant B
lighting	1	20
Curtain	38	95
Robot	0	6
Refrigerator	0	0
Table	0	0
Floor	26	45

5 Conclusion and Future Works

This paper describes the development of a smart home system using MR technology. In order to provide services tailored to the user, it is necessary to measure the user's condition. We used eye gaze as a method for this purpose, and provided information that matches the state of the person's eye gaze. To confirm this effect, we conducted a comparison experiment with a system that provides information that does not match the user's gaze, and confirmed that users prefer the system that provides information that matches their gaze. By measuring eye movement, the system was able to provide the information that the user was seeking. We believe that we can provide more advanced services by inferring the user's impression of an object from the user's information.

In the future, we believe that by measuring the user's facial expressions, hand movements, and other gestures in addition to eye gaze measurement, it will be possible to provide information that corresponds to more detailed changes in the user's emotions.

Acknowledgements This work was supported by JSPS KAKENHI (Grant-in-Aid for Early-Career Scientists) Grant Number 23K17261. The authors would like to thank the students for their help in testing our system and conducting questionnaires.

References

1. Wirtz, J., Kunz, W., Paluch, S.: The service revolution, intelligent automation and service robots. European Business Review 29(5), 909 (2021)
2. Aloysius, J.A., Hoehle, H., Goodarzi, S., Venkatesh, V.: Big data initiatives in retail environments: Linking service process perceptions to shopping outcomes. Annals of operations research 270(1), 25–51 (2018)
3. Chuah, S.H.W., Yu, J.: The future of service: The power of emotion in human-robot interaction. Journal of Retailing and Consumer Services 61, 102551 (2021)
4. Song, C.S., Kim, Y.K.: The role of the human-robot interaction in consumers' acceptance of humanoid retail service robots. Journal of Business Research 146, 489–503 (2022)

5. Ajanki, A., Hardoon, D.R., Kaski, S., Puolamäki, K., Shawe-Taylor, J.: Can eyes reveal interest? implicit queries from gaze patterns. User Modeling and User-Adapted Interaction 19, 307–339 (2009)
6. Bayliss, A.P., Frischen, A., Fenske, M.J., Tipper, S.P.: Affective evaluations of objects are influenced by observed gaze direction and emotional expression. Cognition 104(3), 644–653 (2007)
7. Schrammel, F., Pannasch, S., Graupner, S.T., Mojzisch, A., Velichkovsky, B.M.: Virtual friend or threat? the effects of facial expression and gaze interaction on psychophysiological responses and emotional experience. Psychophysiology 46(5), 922–931 (2009)
8. Plopski, A., Hirzle, T., Norouzi, N., Qian, L., Bruder, G., Langlotz, T.: The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality. ACM Computing Surveys (CSUR) 55(3), 1–39 (2022)
9. Haas, J.K.: A history of the unity game engine (2014)
10. Woo, J., Sato, T., Ohyama, Y.: Development of a human-centric system using an iot-based socially embedded robot partner. Journal of Robotics and Mechatronics 35(3), 859–866 (2023)