

Eye Contact as a New Modality for Man-machine Interface

Syusuke Kobayashi¹, Pitoyo Hartono²

Graduate School of Engineering, Chukyo University, Nagoya, Japan¹
School of Engineering, Chukyo University, Nagoya, Japan²

Abstract—In daily life, people use many appliances, where different machines and tools should be operated with their specialized interfaces. These specialized interfaces are often not intuitive and thus require considerable time and effort to master. On the other hand, human communications are rich in modalities and mostly intuitive. One of them is eye contact. This study proposes eye contact for enriching modalities for human-machine interface. The proposed interface modality, based on a neural network for object detection, allows humans to initiate machine operations by looking at them. In this paper, the hardware framework for building this interface is elaborated and the results of usability assessment through users' experiments are reported.

Keywords—Eye contact; man-machine interface; non-verbal communication; object detection; neural network

I. INTRODUCTION

Machines and tools have been invented for centuries to support humans to efficiently work and to make life comfortable. However, most machines, for example, home appliances, require specialized interfaces that are not necessarily intuitive for humans. The situation is more serious in industrial settings where many types of machinery need to be operated under specialized rules. Sometimes it takes considerable time for humans to be proficient in the interface while misoperation may cause serious accidents. Therefore, it is desirable to develop interfaces that minimize human errors and allows users to instantly understand and be proficient in using them. The intuitiveness of an interface is influenced by various factors since the user's subjective senses are important. For example, in psychology, there is a concept of Organization of Memory [1, 2]. Since there is a limit to the capacity of human memory, it is important to represent information and skills in simple forms that can be easily remembered or executable by humans. One of the important aspects of the simplicity of the representation is familiarity. From these considerations, interfaces will be more intuitive if they are based on familiar experiences [3]. Hence, enclosing familiarity is a good strategy for building an intuitive human-machine interface.

While most machines need fixed and specialized interfaces, humans flexibly utilize various modalities for communicating with each other. In human interactions, verbal communication is the most frequently used interface. However, humans also utilize rich nonverbal modalities for communication, such as gestures, facial expressions, gaze, and eye contact.

Recently, interfaces based on voice and natural language recognition have been widely used in households. They enable humans to verbally interact with machines. While verbal interfaces are effective in household settings, they are not necessarily useful in other settings, such as factories, busy streets, public spaces, and public transportation. In such situations, humans complement verbal communication with nonverbal modalities [4] to seamlessly interact with each other. Hence, it is also beneficial for human-machine interfaces to complement verbal-based modality with nonverbal modalities, for example, eye contact.

In recent years, the advance in technology allows the proposal for many nonverbal interfaces between humans and machines. For example, gaze-based interactions between humans and computers have been developed to assist people who are unable to perform some physical movements due to spinal cord injury or other causes, but also for helping healthy users efficiently operate computers. For example, some studies have been conducted on the use of eye gaze for cursor manipulation in Graphical User Interface (GUI) [5, 6]. When operating a GUI using a mouse or touch screen, the user's gaze is directed to a button on the screen before making a selection. While it is natural and reasonable to use eye movement as a pointer, there is a so-called Midas Touch Problem [7], in which the system cannot determine the intention of a user, in that it is difficult to distinguish whether the user is looking at the screen or has an intention for clicking a button on the screen. Furthermore, it has been reported that clicking by staring or blinking has some latencies compared to clicking a mouse [8]. A system has also been developed to move a wheelchair in the direction of the user's gaze [9]. Here, the discomfort of having to look down the road when operating the wheelchair has been reported in the experiment in simply linking gaze input to a particular movement of a machine is not natural and not intuitive. Additionally, eye gaze is information that can be used to extract human unconscious interests and attention. One of the gaze-based systems [10] generates e-commerce recommendations based on gaze information. While conventional recommendation systems require past shopping characteristics of a user, to determine what to recommend, this system can estimate the user's preferences with high accuracy based on his/her gaze movements. In addition, there is also a study that detects drivers' distractions using gaze tracking as Advanced Drive Assistance System (ADAS) [11]. By dealing with gaze, a human can interact with machines intuitively, naturally, and efficiently. The intuitiveness of gaze information motivates this study.

This study attempts to propose a means for adding a new modality for nonverbal interaction between humans and machines. Here, the basic concept is to allow eye contact between humans and machines. Eye contact has four roles in human communication [12-14]. The one that is highly relevant to this study is the cognitive role of displaying attention to other people and conveying an intention for starting to communicate. In this study, eye contact is expanded to establish intuitive interactions between humans and machines.

There were existing interfaces that attempt to utilize eye contact. For example, the smart speaker “Tama” [15] can be activated using mutual gaze for starting an interaction. It is reported that the usability and the sense of dialogue improved. Other systems include the construction of an IoT system that combines eye gazing and gestures for human appliances [16]. It realized intuitive interaction between humans with home appliances through gaze and gesture via a so-called “Watch module”. Our study shares some similarities with these past studies, in that they realize natural and intuitive interaction by using eyes [15-17]. However, the proposed study differs in that it offers direct interactions with the objects without requiring any other intermediate media and hence increases the naturalness and intuitiveness of the interaction. The proposed system also established a flexible relationship that was not limited to smart speakers and home appliances, but also any type or number of objects.

In the past, a basic framework was developed for this system [18]. This paper reports on the preliminary experiments’ results on the performance of the proposed systems and the user’s assessments. It is important to mention that it is not our intention to compete with the existing systems’ efficiency. Here our objective is to investigate the usability of the proposed eye-contact system and to assess its potential for enriching the user interface modalities. In this paper, the proposed system’s characteristics are assessed through statistical tests on users’ experiment data.

The rest of the paper is organized as follows. Section II explains the hardware and software configurations of the proposed Gaze Switch. Section III explains the experiments, while the final section explains the conclusions and future work for this study.

II. OUTLINE OF GAZE SWITCH

Gaze Switch developed in this research is an interface that enables humans to activate or deactivate a machine by looking at it. This interaction is comparable to eye contact. Fig. 1 illustrates the process of establishing inter-human and human-machine through eye contact.

Eye contact between two humans starts when they gaze at each other and in the process, each party needs to perceive the gaze. In this study, for human machine-interaction, it is assumed that the machines are always gazing at humans, and so when a human visually perceives the machines, eye contact is established. Here, a neural network is utilized for determining the target object. Fig. 2 shows an overview of the Gaze Switch system developed in this research.

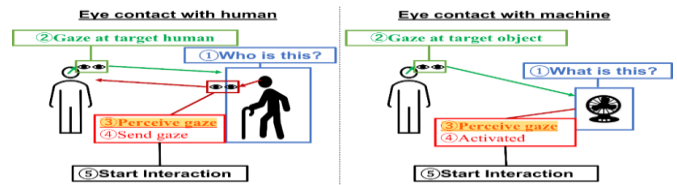


Fig. 1. Eye contact.

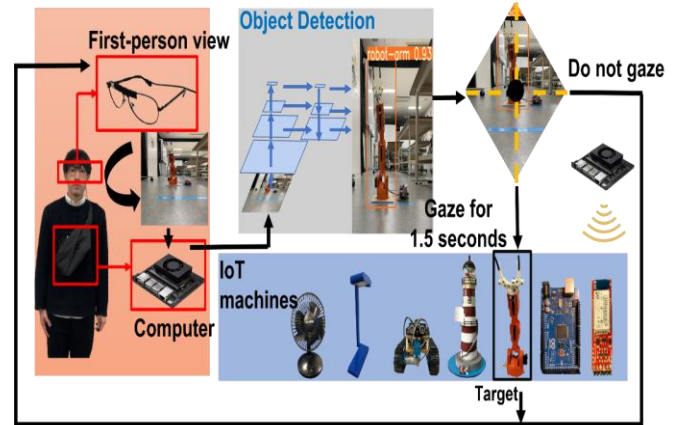


Fig. 2. Outline of the proposed system.

Here, a small camera is attached to the eyeglasses worn by a user. This camera captures and sends the image to a computer to be further processed by a neural network running on the computer for object detection. Here, the objects to be detected must be pre-specified for training the neural network, although the type and number of the objects are not constrained. YOLOv5 [19, 20] is utilized for the neural network’s easy implementation and fast response. Fig. 3 shows the five machines as targets in this study.

The input to YOLOv5 is the image perceived by a human through the attached camera, and the output is bounding boxes, the center normalized coordinate of objects in the images, their heights and widths, and their IDs, as shown in Fig. 4.

To train the neural network, 3427 images for training data and 773 images for test data consisting of the five machines in various postures and distances were generated and labeled. The learning results are presented in the next section.



Fig. 3. Appearances of five machines.

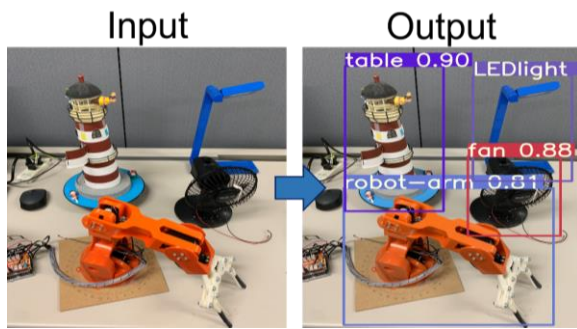


Fig. 4. The output of YOLOv5.

After detecting an object with YOLOv5, the system checks whether the human gaze is focused on that object. Here, in establishing eye contact, it is assumed that the human always put the intended object at the center of his/her field of view. Thus, the center of the coordinate of the obtained images is treated as the focus of the gaze. The system checks that the eye contact object is within the gaze focus for 1.5 seconds. Subsequently, the system refers to the object ID and sends a signal to the target machine. Here, each target machine is connected to a control PC wirelessly via Bluetooth.

By executing the above process in real time, the proposed Gaze Switch system is realized.

III. EXPERIMENTS

The viability of the proposed Gaze Switch is assessed through user experiments. Before the user experiments, some preliminary experiments were run for verifying the basic operability of the proposed system.

The preliminary experiments were run on Windows 10 Pro Intel® Core™ i7-9700k CPU @3.60GHz to 4.90GHz 16.00GB and Nvidia® GTX750Ti GPU @1020MHz to 1085MHz, while the user experiments were run on Jetson Xavier NX [21], operated by Jetson Pack 4.5.1 for improving the system's compactness and processing speed. The neural networks for both experiments were the same.

A. Preliminary Experiments

In the preliminary experiment, the detection range of the neural network was assessed. The necessity of this assessment is due to the existence of a natural range of eye contact and whether the proposed system adheres to this natural range [22]. In particular, the mean Average Precision (mAP) against validation data was evaluated as shown in Fig. 5.

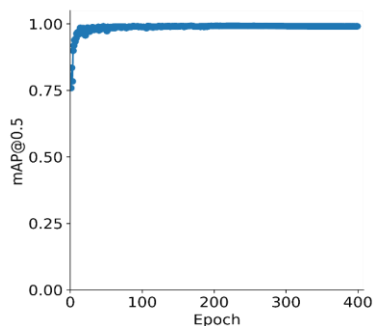


Fig. 5. Learning precision.

From Fig. 5, it can be observed that the mAP exceeds 0.9 after the training process, indicating that the neural network can learn the object detection task.

Next, validation data were created for evaluating the neural network's detection accuracy. Here, 700 labeled images of the seven objects at various distances are generated and checked for their detection accuracies. Fig. 6 shows the results of the accuracy test regarding the distance of the object (with the example of the robotic arm).

This figure shows that the accuracy does not significantly decrease until 5 meters. This indicates that the operating range of the system is around 5 meters which is similar to the human's natural range for eye contact.

The preliminary experiments indicate that the proposed system is viable for establishing eye contact intuitively and naturally.

After assessing the neural network's learning and detection capabilities, two experiments with nine human subjects were conducted. The main purpose was to verify the basic usability and operability of the Gaze Switch. Before the experiments, the subjects were explained the objective of the proposed interface.

After that, the subjects practiced using the interface for about five minutes. In the experiment, a monitor in front of the subject randomly showed an object that the subject must operate. Here, the subject should try to operate the specified machine by looking at it. The interface was evaluated in whether the human subject could operate the object within a specific time range. In the experiment, the subjects were instructed ten times in random order to operate each target twice.

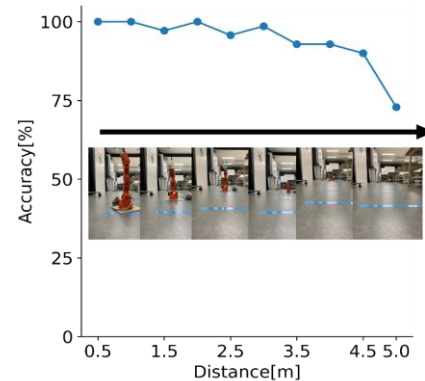


Fig. 6. Accuracy across distances.

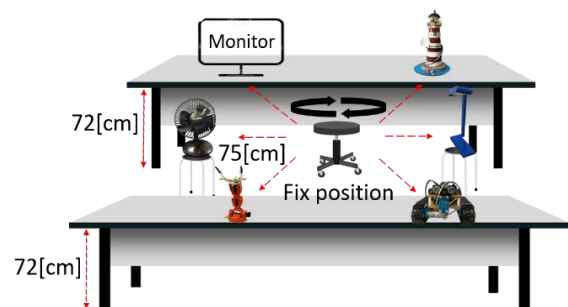


Fig. 7. Settings for preliminary experiment 1.

In this experiment, the target objects were randomly positioned but fixed to their respective position. During the experiment, the subjects were instructed to sit in a fixed position and use a swivel chair to turn their bodies to establish eye contact with the objects (see Fig. 7).

The human subjects were instructed to operate an object that randomly appears on the monitor within 10 seconds, and then return his/her gaze to the monitor. If the subject fails to operate the object within 10 seconds, the experiment continues, but the task is considered a failure. Fig. 8 shows the average accuracy for operating the instructed object.

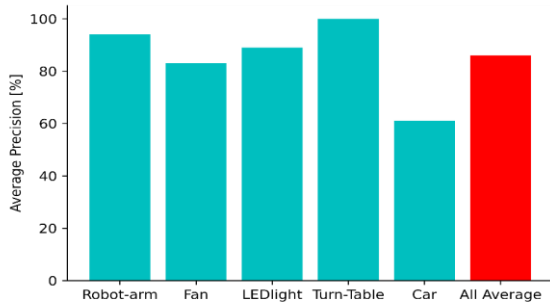


Fig. 8. The result of preliminary experiment 1.

The overall average accuracy was 88% which indicates the subjects were able to operate the specified targets. However, the accuracy rate for the object “Car” is low. This low accuracy is due to “Car” moving away from its fixed position. This indicates that it may be difficult for a human to operate moving objects from a fixed position using this interface.

Next, Preliminary Experiment 2 was conducted with the setup shown in Fig. 9.

The same task was applied to this experiment as in Preliminary Experiment 1. However, in this experiment the subject was allowed to freely move rather than operate from a fixed position. This experiment aims to evaluate the operability of this system in an environment that is more similar to a general living space. Fig. 10 shows the average accuracy in Preliminary Experiment 2.

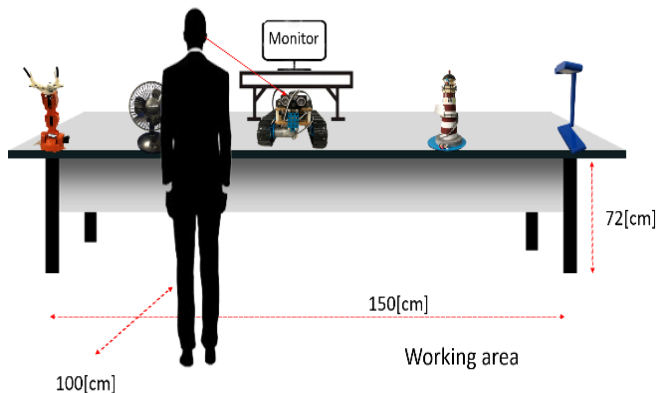


Fig. 9. Settings for preliminary experiment 2.

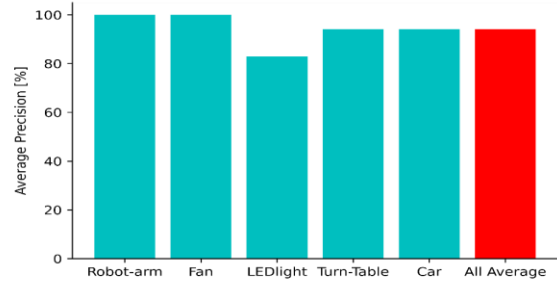


Fig. 10. The result of preliminary experiment 2.

The results in Preliminary Experiment 2 are better than those in Preliminary Experiment 1. In particular, the average accuracy for the object Car is significantly improved. This accuracy is because each subject can move his/her body in a way that makes it easier to follow the object when necessary.

From these experiments, the operability range and usage of the proposed system can be learned. The insights gained from the preliminary experiments are then utilized for users’ assessment tests.

B. User Assessment Experiment

User assessment experiments were conducted to investigate the usability of the proposed eye-contact system and to assess its potential for its usage of interface modality. As an evaluation index, usability defined in ISO 9240-11 [23] was used. This criterion encompasses efficiency, effectiveness, and user satisfaction when a user executes a specific task to achieve a goal. This section explains the experiment method and the analytical results. A short demo movie for this experiment can be accessed from <https://youtu.be/rKBBP2aLcxY>.

A conventional remote-control switch was utilized as a benchmark against the proposed Gaze Switch. The outline of a remote-control switch is shown in Fig. 11.

Here, a target machine can be activated/deactivated by pushing a correlated button in the remote controller (RC) like selecting a TV channel.

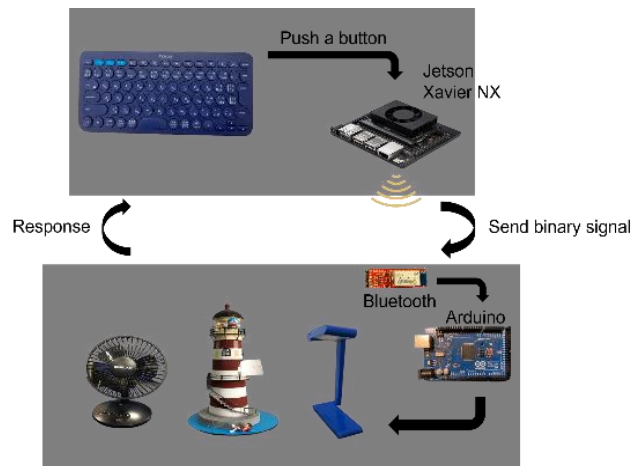


Fig. 11. Outline of the remote-control switch.

Naturally, as many subjects have already been familiar with RC for a long time, it is easy to predict that RC yields higher usability measures. Therefore, the objective of this experiment is not to directly compare the usability measure of the proposed Gaze Switch with RC but to compare the improvement of the usability over some repeated experiments. Here, it is enough to argue that if the users experience improvements in their familiarity and operability of the Gaze Switch over some repeated experiments, then the validity of the Gaze Switch as a new modality can be confirmed. It is also important to show that the improvements' signatures are similar to those of RC. Here, the number of repeated experiments was four with an interval of one week before the next experiment.

In the experiment, the task is based on a scenario where a user activates a household appliance while reading a book at home. This scenario is applied based on the results of the preliminary experiments where the fixed position of users yields worse results so that the rigorousness of the test is guaranteed. Fig. 12 illustrates the experimental environment.



Fig. 12. The setting for the user assessment experiments.

The procedure was as follows.

Step 1. The subject sits and reads a book.

Step 2. A control PC instructs the subject to activate or deactivate one of three designated machines: Fan, LED Light, or Turn Table, in a random manner.

Step 3. The subject stops reading and activates/deactivates the machine with an interface at hand.

Step 4. The subject resumes reading upon confirming the activation or deactivation.

Go to Step 1 until the terminal condition is met.

During the experiment, the reaction time from the instruction to the activation/deactivation of the specified machine and the accuracy of the interaction were evaluated. Additionally, the users were asked to complete the questionnaire for measuring their satisfaction based on System Usability Scale (SUS) [24] on a 5-point Likert scale as shown in Fig. 13. The result of SUS is standardized in the range of 0 to 100. A null hypothesis test regarding the significant difference between the interfaces' characteristics and the familiarity factor due to the repeated usage of these usability-quantitative data was run. Here, the significance of the difference is assessed through a p-value with a 0.05 threshold. Here, the null hypothesis is that the evaluated factors are identical regarding the usage of RC and Gaze Switch.

Questionnaire Item	strongly agree	agree	uncertain	disagree	strongly disagree
1. I thought that I would like to use this system frequently.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. I found the system unnecessarily complex.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. I thought the system was easy to use.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. I thought that I would need the support of a technical person to be able to use this system.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. I found the various functions in this system were well integrated.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. I thought there was too much inconsistency in this system.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. I would imagine that most people would learn to use this system very quickly.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. I found the system very cumbersome to use.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. I felt very confident using the system.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. I needed to learn a lot of things before I could get going with this system.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 13. Question items of the system usability scale.

Eq. (1) shows the improvement rate, $IR_{i,j}$ of the j -th factor in the i -th repeated test, where $R_{i,j}$ is the score of factor j in the i -th experiment where $j \in \{\text{Reaction time, Accuracy, SUS score}\}$ and $i \in \{\text{Second, Third, Fourth}\}$. This experiment was conducted with 12 subjects.

$$IR_{i,j} = \frac{R_{i,j} - R_{First,j}}{R_{First,j}} \times 100\% \quad (1)$$

First, the average reaction time needed for operating the specified object with RC and Gaze Switch for the respective test is shown in Fig. 14(a) and the IRs are shown in Fig. 14(b).

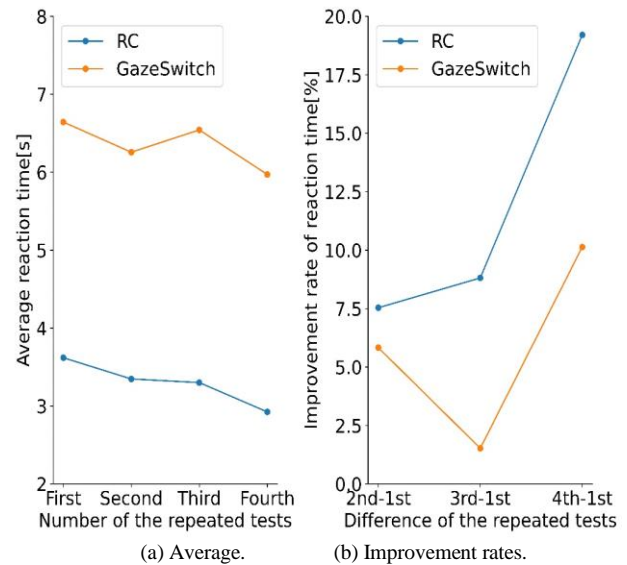


Fig. 14. Reaction time.

RC is superior to Gaze Switch with regard to the reaction time as can be learned from Fig. 14(a). To check whether the difference is significant, a Wilcoxon signed rank test was conducted that showed a significant difference ($p < 0.001$). In addition, a Friedman test on the familiarity factor of RC and Gaze Switch was conducted. The result showed significant differences between the repeated tests on RC ($p = 0.002$) but no significant differences on Gaze Switch ($p = 0.376$). Fig. 14(b) shows that the gap in the reaction time between RC and Gaze

Switch does not decrease with the number of tests, and this trend can be predicted to stay true. Hence, it can be argued that the difference regarding the reaction time does not depend on familiarity due to the repeated usage of the two interfaces but depends on the basic characteristics of the two interfaces. Table I shows the average reaction time of each machine in the experiment on the respective test.

During the experiments, Gaze Switch allows the subjects to interact with the machines by perceiving them for 1.5 seconds but the latency compared to RC's reaction time was about 3.0 seconds. This latency is due to a discrepancy between the human's field of view and that of the camera, but not the system's operating range.

Next, the average accuracies (the correctness of activating/deactivating the instructed object) using RC and Gaze Switch for the respective test are shown in Fig. 15(a) and IRs are shown in Fig. 15(b).

It is obvious from Fig. 15(a) that the subjects did not need a long time to get familiar to use the Gaze Switch, indicating its good intuitiveness. The intuitiveness of the Gaze Switch is further emphasized in Fig. 15(b). Regarding the difference, the result of a Wilcoxon signed rank test on the interface factor showed a significant difference ($p=0.008$) between the accuracy of Gaze Switch and RC. In addition, a Friedman test on the familiarity factor in RC was conducted with no significant differences in RC ($p=0.137$). The significance test results show that RC is not necessarily stable in its intuitiveness. This is because the users occasionally misoperate

the machines with RC due to the failure to memorize the relation between the buttons in the RC and the machines. By contrast, with the proposed Gaze Switch, the user can operate an intended machine by looking at it, so it does not need any memorization. From this experiment, it can be argued that the intuitiveness of the proposed interface contributes to its accuracy. Table II shows the accuracy of the respective machines.

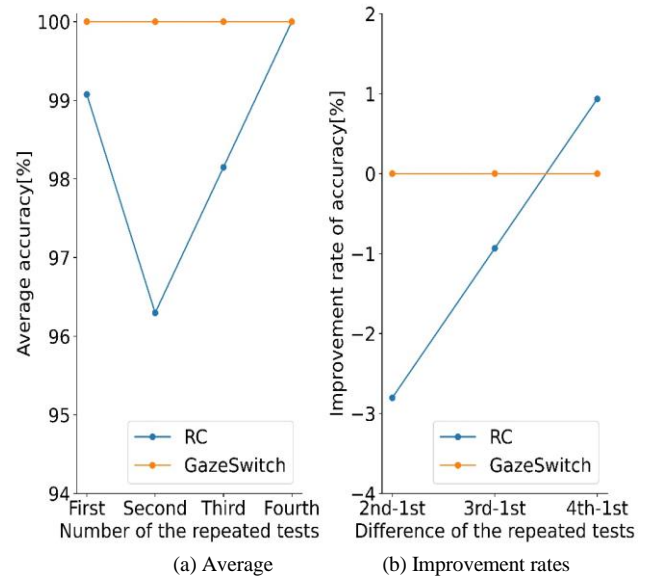


Fig. 15. Accuracy.

TABLE I. AVERAGE REACTION TIME

Trials	RC Average Reaction Time (s)			Gaze Switch Average Reaction Time (s)		
	<i>Fan</i>	<i>TurnTable</i>	<i>LEDlight</i>	<i>Fan</i>	<i>TurnTable</i>	<i>LEDlight</i>
First	3.81	3.54	3.51	6.39	6.93	6.61
Second	3.44	3.27	3.33	5.73	6.70	6.34
Third	3.33	3.54	3.03	6.44	6.89	6.30
Fourth	2.93	2.81	3.03	6.04	5.94	5.93
All	3.38	3.29	3.29	6.15	6.62	6.30

TABLE II. AVERAGE ACCURACY

Trials	RC Average Accuracy (%)			Gaze Switch Average Accuracy (%)		
	<i>Fan</i>	<i>TurnTable</i>	<i>LEDlight</i>	<i>Fan</i>	<i>TurnTable</i>	<i>LEDlight</i>
First	100	97.2	100	100	100	100
Second	97.2	94.4	97.2	100	100	100
Third	97.2	97.2	100	100	100	100
Fourth	100	100	100	100	100	100
All	98.6	97.2	99.3	100	100	100

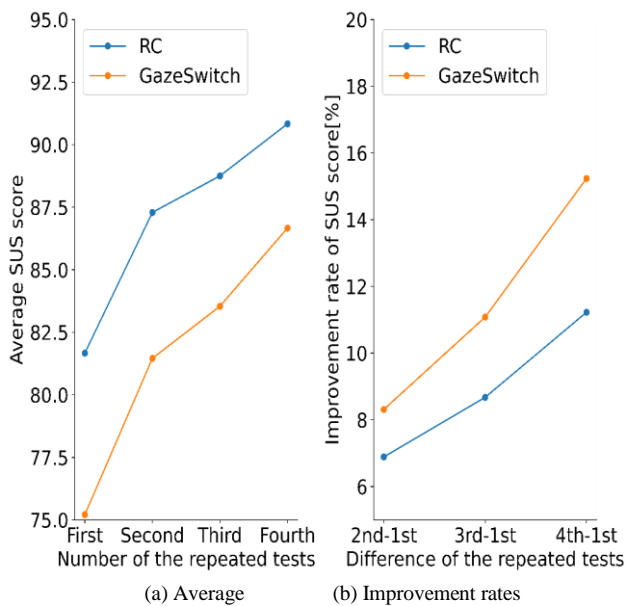


Fig. 16. SUS score.

Finally, the average SUS scores of RC and Gaze Switch are shown in Fig. 16(a) and the IRs are shown in Fig. 16(b).

Regarding the SUS score, RC is inferior to Gaze Switch as indicated in Fig. 16(a). The value for Cronbach's Alpha ($0 \leq \alpha \leq 1$) to measure whether each item of the questionnaire is reliably able to measure the identical concept (here, satisfaction) by confirming the average covariance between pairs of the items, and the variance of the total score, was $\alpha = 0.83$. These results indicate the result of the questionnaire is reliable because the value is 0.8 or more. Meanwhile, a Wilcoxon signed rank test to the result on the interface factor showed a significant difference ($p=0.014$). Likewise, A Friedman test on the familiarity factor in each RC and Gaze Switch showed significant differences in RC ($p=0.027$) and Gaze Switch ($p<0.001$). A Wilcoxon signed-rank test on the familiarity factor in RC and Gaze Switch was also conducted. The result showed a significant difference between the First and Fourth of the repeated tests in Gaze Switch ($p=0.004$ after Bonferroni correction for the multiple comparisons problems). Fig. 16(b) shows that the difference in questionnaire scores and the number of repeated tests is getting wider.

It is restressed here that the objective of the experiments is not to directly compare the performance of the RC against the proposed Gaze Switch. The primary objective is to investigate the characteristics of the Gaze Switch in its usage as an interface modality using RC as a baseline. The experiments indicate that the Gaze Switch shows good user intuitiveness. Regarding the reaction time and the user SUS, the Gaze Switch shows good familiarity growth, meaning repeated usage will yield better experiences. The growth trends are also similar to that of a more established interface of RC, which shows the appropriateness of the proposed Gaze Switch as a new modality in the human-machine interface.

IV. CONCLUSION

In this study, we developed a hardware framework for expanding an intuitive and familiar communication modality, eye contact, for human-machine interaction. The proposed system allows humans to intuitively operate machines through eye contact. Unlike the existing gaze interfaces that often depend on specialized tools, the system allows direct interaction with various machines, thus offering better flexibility and intuitiveness. The users' assessment tests in this study demonstrate that familiarity with eye contact in human daily communications translates into intuitiveness and robustness of the system. Through this study, it can be argued that eye contact is a reasonable modality in the human-machine interface.

The authors are aware of some technical drawbacks of the proposed system. For example, the relatively long reaction time decreases the usability of the proposed system. This is due to the discrepancy between the human field of view and the field of view captured by the camera. In the near future, this problem can be alleviated by better calibration of the camera or using a multi-camera system to align the view better.

Like the rich modalities in human interactions, in the future, the proposed Gaze Switch is not intended for single usage but in combination with other modalities, for example, verbal and nonverbal interfaces. The combinations of various interfaces will improve the precision of human-machine interactions and remove the difference between inter-human interactions and human-machine interactions. The seamless integration of machines into human interactions in daily life is one of the most important aspects in the coming era of AI technology, Metaverse, and XR, and hence the proposed eye-contact system has good potential for enriching the existing modalities for human-machine interfaces.

REFERENCES

- [1] G. Mandler, "Organization and Memory," *Psychology of Learning and Motivation*, vol. 1, pp. 327-372, 1967.
- [2] S. M. Polyn, K. A. Norman, and M. J. Kahana, "Task context and organization in free recall," *Neuropsychologia*, vol. 47, no. 11, pp. 2158-2163, 2009.
- [3] M. Hirokawa, K. Inoue, T. Iwaki, and T. Kashima, "Fundamental Study on an Intuitive Interface Design Theory -Approach from Viewpoint of Organization and Familiarity," *Transaction of Japan Society of Kansei Engineering*, vol. 13, no. 5, pp.543-554, 2014.
- [4] A. Mehrabian, "Communication without words," *Psychological Today*, vol. 2, pp. 53-55, 1968.
- [5] M. Choi, D. Sakamoto, T. Ono, "Bubble Gaze Cursor + Bubble Gaze Lens: Applying Are Cursor Technique to Eye-Gaze Interface," in *Proceedings of the 2020 ACM Symposium on Eye Tracking Research and Applications Full Papers*, no. 11, pp. 1-10, 2020.
- [6] M. Choe, Y. Choi, J. Park, "Comparison of Gaze Cursor Input Methods for Virtual Reality Devices," *International Journal of Human-Computer Interaction*, vol. 35, issue. 7, pp. 620-629, 2018.
- [7] R. J. K. Jacob, "The use of eye movements in human-computer interaction what you get," *ACM Transactions on Information Systems*, vol. 9, no. 2, pp.152-169, 1991.
- [8] C. Lutteroth, M. Penkar, G. Weber, "Gaze vs. Mouse: A Fast and Accurate Gaze-Only Click Alternative," in *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* pp. 385-394, 2015.

- [9] J. M. Araujo, G. Zhang, J. P. P. Hansen, S. Puthursserypaddy, "Exploring Eye-Gaze Wheelchair Control," ACM Symposium on Eye Tracking Research and Applications, no. 16, pp. 1-8, 2020.
- [10] S. Jaiswal, S. Virmani, V. Sethi, K. De, P. P. Roy, "An intelligent recommendation system using gaze and emotion detection," Multimedia Tools and Applications, vol. 78, pp.14231-14250, 2018.
- [11] A. Ledezma, V. Zamora, O. Sipele, M. P. Sesmero, A. Sanchis, "Implementing a Gaze Tracking Algorithm for Improving Advanced Driver Assistance Systems," Electronics, vol. 10, no. 12, 1480, 2021.
- [12] A. Kendon, "Some functions of gaze-direction in social interaction," Acta Psychologica, 26:22-23, 1967.
- [13] A. Senju, and M.H. Johnson, "The eye contact effect: mechanisms and development," Trends in Cognitive Sciences, vol. 13, no. 3, pp. 127-134, 2009.
- [14] M. S. Cary, "The Role of Gaze in the Initiation of Conversation," Social Psychology, vol. 41, no. 3, pp. 269-271, 1978.
- [15] I. Kawaguchi, H.Kuzuoka, and D. McMillan, "The Effect of Interaction Using Gaze Input/Output on Smart Speaker," The Transactions of Human Interface Society, vol. 21, no. 3, 2019.
- [16] J. Kim, S. Choi, and J. Jeong, "Watch & Do: A smart IoT inter-action system with object detection and gaze estimation," IEEE Transactions on Consumer Electronics", vol. 65, no. 2, pp. 195-204, 2019.
- [17] F. Putze, D. Weiß, L. -M. Vortmann, T. Schultz, "Augmented Reality Interface for Smart Home Control using SSVEP-BCI and Eye Gaze," 2019 IEEE International Conference on Systems Man and Cybernetics (SMC), pp.2812-2817, 2019.
- [18] K. Syusuke, and P. Hartono, "Development of Eye Gaze Switch," Proceedings of the Forum on Information Technology 2021, J-007, Vol. 3, pp. 255-257, 2021.
- [19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.779-788, 2016.
- [20] G. Jocher, YOLOv5 by Ultralytics (Version 7.0). <https://doi.org/10.5281/zenodo.3908559>, 2020.
- [21] Nvidia, Inc.: Jetson Xavier NX; <https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-xavier-nx/>.
- [22] S. Daniel, "Eye contact perception at distances up to six meters," Portland State University Dissertations and Theses, Paper 3693, 1985.
- [23] International Organization for Standardization, "ISO9241-210:2020 Ergonomics of human system interaction Part 210: Human-centred design for interactive systems," 2010.
- [24] J. Brooke, "SUS-A quick and dirty usability scale," in Usability evaluation in industry, pp. 189-194, 1996.