

RESEARCH

Open Access



Vision-sharing system for android avatars to enable remote eye contact

Kaoruko Shinkawa¹, Mizuki Nakajima² and Yoshihiro Nakata^{1*}

Abstract

Maintaining eye contact is a fundamental aspect of non-verbal communication, yet current remote communication tools, such as video calls, struggle to replicate the natural eye contact experienced in face-to-face interactions. This study presents a cutting-edge vision-sharing system for android avatars that enables remote eye contact by synchronizing the eye movements of the operator and the avatar. Our innovative system features an eyeball integrated with a wide-angle lens camera, meticulously designed to mimic human eyes in both appearance and functionality. This technology is seamlessly integrated into android avatars, enabling operators to perceive the avatar's surroundings as if physically present. It provides a stable and immersive visual experience through a head-mounted display synchronized with the avatar's gaze direction. Through rigorous experimental evaluations, we demonstrated the system's ability to faithfully replicate the operator's perspective through the avatar, making eye contact more intuitive and effortless compared with conventional methods. Subjective assessments further validated the system's capacity to reduce operational workload and enhance user experience. These compelling findings underscore the potential of our vision-sharing system to elevate the realism and efficacy of remote communication using android avatars.

Keywords Android avatar, Eyeball, Interface, Teleoperation, Wide-angle lens camera

Introduction

The increasing demand for remote communication tools has raised expectations for technologies that enable more natural face-to-face interactions. Video calls have become the predominant method of communication with individuals who are geographically distant. However, the challenge lies in creating a sense of physical presence during these calls, often leading to feelings of fatigue. One significant factor contributing to this is the inability to establish natural eye contact through video calls [1].

Eye contact is a significant aspect of nonverbal communication in human interactions. It occurs approximately 61% of the time during casual conversations between two people, with about half of these instances being mutual [2]. Humans rely on eye contact for various purposes, such as seeking information, signaling openness, concealment and exhibitionism, and establishing and recognizing social relationships, which are essential for effective communication [3]. Eye contact also possesses both approach and avoidance forces. When we sense that a certain social factor, such as physical proximity, is not suitable for the relationship we have with someone, we instinctively adjust our eye contact to maintain the appropriate social distance [3]. However, current remote communication tools do not allow for the natural exchange of eye contact that occurs during in-person interactions.

Consequently, researchers have turned to robotic avatars as a potential solution. Tanaka et al. [4] have reported that physical embodiment of interaction partners

*Correspondence:

Yoshihiro Nakata
ynakata@uec.ac.jp

¹ Department of Mechanical and Intelligent Systems Engineering, Graduate School of Informatics and Engineering, The University of Electro-Communications, Chofu 182-8585, Tokyo, Japan

² Department of Robotics and Mechatronics, School of Science and Technology for Future Life, Tokyo Denki University, Adachi-ku 120-8551, Tokyo, Japan



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

influences the perceived presence of the partner. This was demonstrated through a comparison of dialogues via robot avatars and various existing communication media. The utilization of robot avatars is anticipated to facilitate remote communication, creating a sense of being physically present with the other individual. Furthermore, Sacino et al. [5] suggested that a high level of anthropomorphism in appearance is effective in improving the presence of a robot. Considerable research has been conducted on androids, which are robots designed to have human-like appearances [6–10]. These robots exhibit human-like behavior and, in recent years, have even been equipped with the ability to perform eye movements.

Based on these advancements, we hypothesized that utilizing androids as avatars can enhance remote communication by enabling eye contact across long distances. To achieve this, avatars and teleoperation systems should be developed to replicate human-like appearance and eye movements while providing operators with a face-to-face view of the remote environment. Advancements in teleoperation technology have focused on enhancing the operator's sense of presence by incorporating haptic feedback and highly immersive visual experiences in VR headsets [11]. Furthermore, androids that mimic human behavior more closely are being developed, such that the person interacting with the avatar can feel the presence of the other person. Some of these robots can move their faces with multiple degrees of freedom, enabling them to not only replicate the movements of the operator's head and arms but also convey various facial expressions.

However, in remote communication, instances in which both the operator and individual interacting with the robot avatar effectively communicate are limited. This is because the appearance of the avatar often deviates from that of a human when equipped with cameras and sensors to relay information about the avatar's environment to the operator, which enhances functionality. Moreover, research and development on androids that aim to mimic humans often focus on achieving autonomous or pre-programmed behavior, rather than prioritizing operability when utilized as avatars.

Therefore, we propose a teleoperation system that focuses on visual interaction, enabling eye contact through an android avatar while providing a natural user experience, both in terms of operation and natural appearance. The main contributions of this study are as follows.

- Proposition of a vision-sharing system concept that enables eye contact through an android avatar.
- Development of an eyeball integrated with a wide-angle lens camera for android avatars, designed to closely resemble human eyeballs in appearance.

- Implementation of the proposed system and the developed eyeball on an android avatar.
- Evaluation of the system's accuracy through experiments in which participants operate the android avatar and gaze at specified points.
- Assessment of the impact of the proposed system on the ease of establishing eye contact through experiments involving pairs of individuals engaging in mutual eye contact through an android avatar

We have presented our vision-sharing system concept and the development of the eyeball integrated with a wide-angle lens camera for android avatars [12]. Additionally, a preliminary report on the implementation of this system on an android avatar has been documented [13]. However, evaluating the proposed system's ability to facilitate remote eye contact necessitates two specific investigations.

- To determine whether the operator can perceive their surroundings from the avatar's perspective.
- To assess whether the individual interacting with the avatar experience a sense of eye contact when the operator gazes at them through the avatar.

This study presents the findings of two experiments conducted to evaluate the accuracy and effect of the system on eye contact between two individuals through an avatar. These experiments delve into the aforementioned aspects and provide an in-depth discussion on the efficacy of the proposed system.

Concept of the vision-sharing system

We propose a vision-sharing system that enables synchronization of eye movements and sharing of the same visual perspective between operators and avatars. An overview is shown in Fig. 1. Notably, this system requires avatars to be equipped with cameras in both their left and right eyeballs, enabling them to mimic human-like eye movements in all directions. To achieve immersive avatar operation, the system was designed for use with a head-mounted display (HMD). During operation, the avatar's eye movements were synchronized with those of the operator based on the operator's gaze as detected by the HMD. Simultaneously, images captured by the avatar's eye cameras were presented to the operator. Images from the left-eye camera of the avatar were displayed on the left side of the HMD, whereas those from the right-eye camera were displayed on the right. Each image was projected onto a hemispherical virtual screen centered on the operator's viewpoint and displayed on the HMD. This provided a stereoscopic view of the avatar's surroundings.

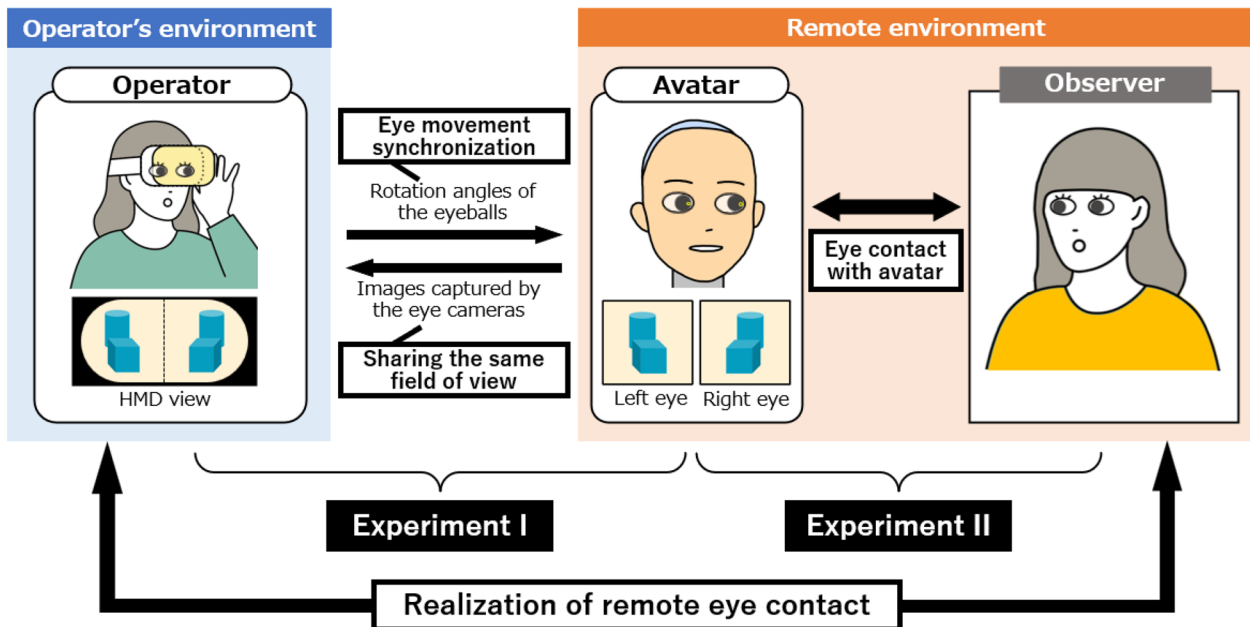


Fig. 1 Concept of the vision-sharing system for remote eye contact

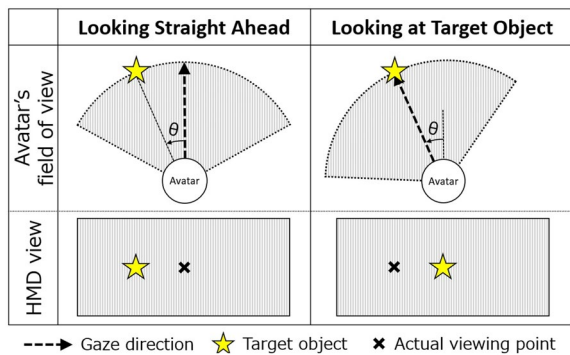


Fig. 2 Problem caused by eye movement synchronization when the virtual screens are fixed

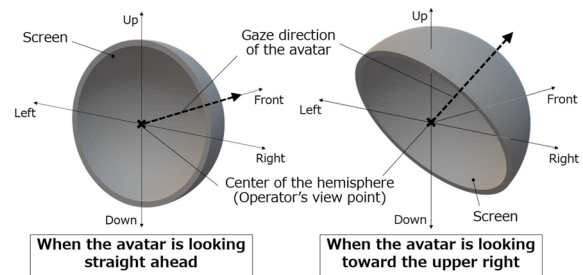


Fig. 3 Hemispherical screen rotation

Our previous study [14] partially validated the effects of eye movement synchronization. However, by synchronizing eye movements, when the operator shifted the focal point of the image displayed on the HMD to a desired position, the camera's capture area also shifted, resulting in a discrepancy where the desired visual content was not aligned with the focal point (see Fig. 2). This synchronization prevents eye contact with the individual interacting with the avatar. To address this, a system was developed to rotate the hemispherical screens based on the gaze direction of the avatar (see Fig. 3). Screen rotation was based on the rotation angle of the avatar's eyeballs, with the screen rotating

in the opposite direction to match the estimated rotation angle of the avatar's gaze at the time of projection. This ensures a stable, seamless view for the operator, aligning the avatar's gaze with that of the operator.

The proposed system was implemented and evaluated. To assess its effectiveness in enabling remote eye contact through an avatar, two key aspects needed to be evaluated:

1. **Accuracy of the implemented system:** Determining whether the operator can perceive the avatar's surroundings as if they were physically present at the avatar's location, as intended by the system's design.
2. **Impact on the subjective difficulty of establishing eye contact:** Assessing whether the system facilitates easier eye contact through the avatar.

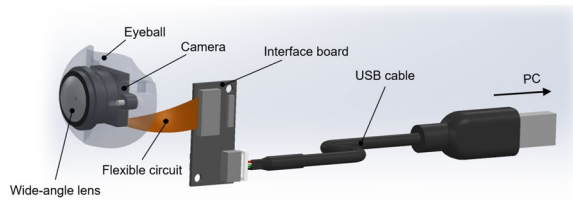


Fig. 4 Structure of the designed eyeball integrated with a wide-angle lens camera

Table 1 Component specifications

Component	Specifications
Wide-angle lens	200 degrees
Camera	OV5640, 5 megapixels
Eyeball	Molding with resin
Interface board	12.5mmx26.0mmx1.0mm
USB cable	Type A

We conducted evaluation experiments, referred to as Experiments I and II, to validate the aforementioned aspects. In each evaluation experiment, we compared the following two conditions:

- **Fixed screen condition (“Fixed condition”)**: A condition in which the virtual hemispherical screens are fixed (conventional system).
- **Synchronized screen condition (“Sync condition”)**: A condition in which the virtual hemispherical screens are rotated based on the avatar’s gaze direction (proposed system).

Implementation of the vision-sharing system on an android avatar

We developed both hardware and software components necessary for implementing the proposed vision-sharing system on an android avatar.

Hardware

The implementation of the proposed system necessitates the integration of a camera within each eyeball of the avatar to provide the android avatar’s operator with a visual field equivalent to that of a human. Moreover, the eyeballs must perform human-like movements and have a human-like appearance. To satisfy these requirements, we developed an eyeball integrated with a wide-angle lens camera specifically designed for android avatars.

The structure of the developed eyeball is shown in Fig. 4, with the specifications for each component shown in Table 1. A camera module featuring a 200°-wide-angle lens

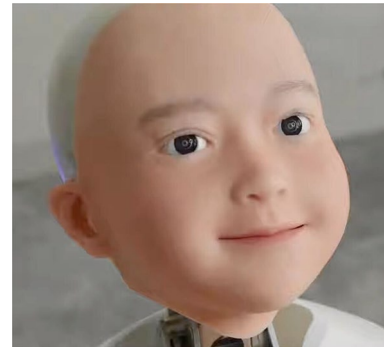


Fig. 5 Android avatar Yui

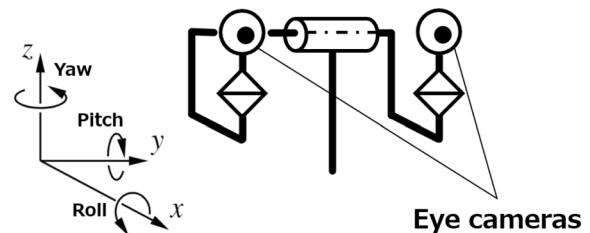


Fig. 6 Mechanical structure around the eyeballs

was employed to capture images from the android’s perspective. The design of the eyeball ensured that the wide-angle lens resembled the pupil and iris of a human eyeball. The wide-angle lens camera was connected to the control PC through a USB cable and an interface board, enabling the acquisition of captured images. It was enclosed in a resin component that mimics the appearance of the sclera of the human eye. The iris was designed with a slightly blurred edge to replicate the natural appearance of human eyes, achieving a highly human-like appearance.

Software

Our vision-sharing system was successfully implemented on the humanoid cybernetic avatar Yui, serving as an android avatar [6] shown in Fig. 5. The mechanism surrounding the eyeballs is shown in Fig. 6, with a photograph of the system shown in Fig. 7. The three motors enabled yaw axis rotation for each eyeball and pitch axis rotation for both eyeballs. This enables each eyeball to perform independent horizontal and linked vertical movements. The range of motion of the eyeball, with the lens facing forward, defined as 0°, is $\pm 35^\circ$ horizontally and 14° to 8° vertically. The eyeballs embedded within the android avatar are shown in Fig. 8. The developed eyeball closely resembled the human eye.

In this system, an HMD (MetaQuest Pro, Meta) and Unity were utilized to enable eye tracking, allowing for the simultaneous tracking of the operator’s gaze and

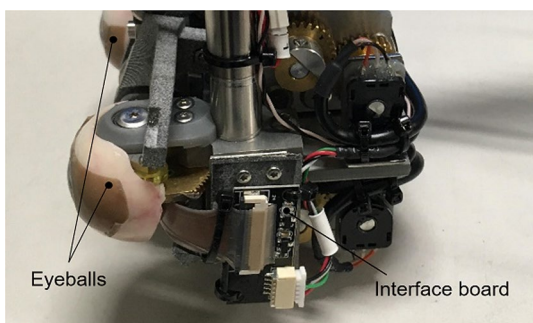


Fig. 7 Photograph of the mechanism around the eyeballs. The tape is attached to the eyeballs as the photograph was obtained during manufacturing

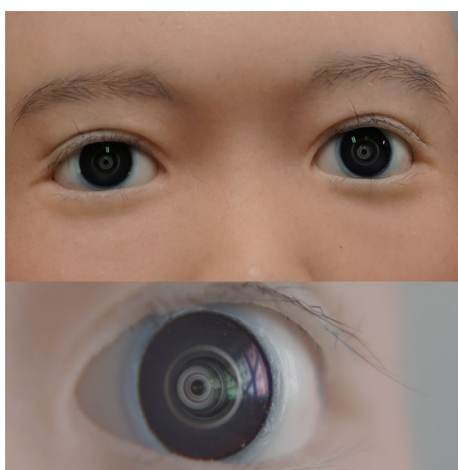


Fig. 8 Close-up views of an android avatar with eyeballs integrated with wide-angle lens cameras

presentation of images from the avatar’s perspective. ROS2 was employed to construct the system and facilitate data exchange between the avatar and operator. The configuration of the communication system is shown in Fig. 9. To ensure stable communication, all evaluation experiments were conducted with the PC for the operating interface and the PC for the avatar connected through a wired LAN.

The yaw and pitch eye angles of each operator were obtained from the HMD eye tracker, and the values corresponding to the respective motors were input as target values. For the pitch direction, the average pitch angle of each eye was used as the target value for both eyeballs.

During operation, the left and right displays of the HMD provided a virtual space. A camera was installed at the origin of each virtual space and the HMD displayed the image captured by the camera. The position and direction of the camera in each Unity space remained constant and the image displayed on the HMD was independent of the position and direction of the HMD. The hemispherical screens were strategically positioned such

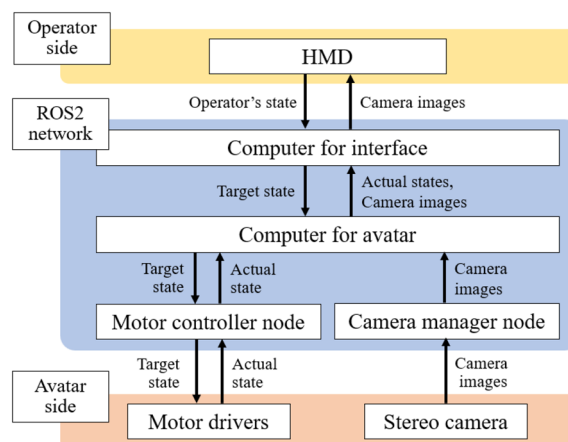


Fig. 9 Communication system configuration



Fig. 10 Captured images displayed on the screens

that the center of each hemisphere aligned with the origin in the virtual space, specifically with the camera position serving as the viewpoint. The eye camera captured 200° range images, with the 180° range image being projected onto a hemispherical screen. This projection enables the operator to observe the avatar’s environment on a scale equivalent to that of the real environment. Figure 10 shows the actual display of the image.

Throughout the operation, camera images and the actual rotation angles of each avatar’s eyeballs were recorded along with timestamps. By utilizing linear interpolation based on these timestamps, the system determined the direction of the avatar’s gaze at the time of image capture, subsequently adjusting the screens projecting the images accordingly. Each image was displayed on a screen that rotated to align the normal vector of the cross-section through the center of the screen hemisphere with the avatar’s gaze vector (see Fig. 3). This allowed the avatar’s eyeballs to rotate in sync with the operator’s movements, creating a lifelike interaction as if the operator were physically present at the avatar’s location. The avatar’s appearance captured by one of the authors while looking at a camera positioned slightly to

the right of the avatar's front is shown in Fig. 11. This was done in both the Fixed and Sync conditions. In contrast to the conventional system with fixed screens Fig. 11a, the implementation of the proposed system in Fig. 11b results in more pronounced eye rotations and a direct gaze toward the camera.

For stable control, the target update and data acquisition frequencies of the motor were set to 10 Hz during the evaluation experiments. To ensure stable communication, the acquisition frequency of the camera images was set to 30 Hz, with images being captured at a resolution of 800 x 600. A median filter was utilized to process input values related to the avatar's eyes, reducing unintended vision oscillations caused by communication delays, accuracy of avatar control, and eye tracking through the HMD. Furthermore, to reduce the possibility of causing VR sickness owing to frequent view changes, the weighted-average filter shown in Algorithm 1 was applied to the input values of the avatar eye and screen rotation angles to prevent view changes when the difference between the avatar eye and screen rotation angles was minimal. The avatar eye camera experienced occasional noise issues owing to interface board problems. However, thorough evaluations validated that these noises did not significantly impact the results.

Algorithm 1 Algorithm of the weighted-average filter

```

1:  $k$ : weighting factor
2:  $RAW \leftarrow$  calculated angle
3:  $PREV \leftarrow$  previous input angle
4: if  $|RAW - PREV| > 5$  then
5:    $k \leftarrow 0.5$ 
6: else if  $|RAW - PREV| > 1$  then
7:    $k \leftarrow 0.1$ 
8: else
9:    $k \leftarrow 0$ 
10: end if
    return  $k \times RAW + (1 - k) \times PREV$ 

```

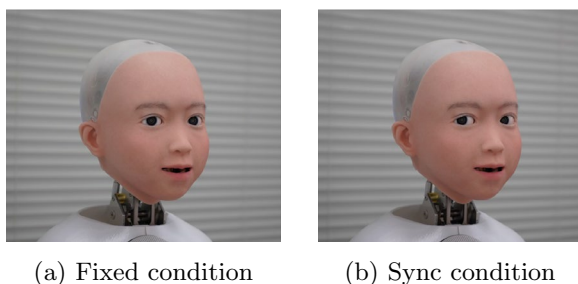


Fig. 11 Appearance of the avatar when the operator is looking at the camera in each condition

Experiment I: accuracy evaluation

An initial experiment was conducted to evaluate the accuracy of the system, ensuring that operators could effectively view their surroundings from the avatar's perspective within the developed system. The experimental hypotheses are as follows:

- The operator's gaze direction when focusing on a target point through the avatar closely aligns with the gaze direction from the avatar's position toward that point in the Sync condition.
- The operator experiences a lower workload in the Sync condition owing to improved view stability.

Method

In this experiment, participants were instructed to gaze at a specific point using an avatar, and the researchers examined how closely their gaze direction matched the avatar's viewpoint when looking at the target point. The participants, equipment, materials utilized, and the experimental procedure are examined below.

Participants

Eleven students from the Nakata Laboratory at the University of Electro-Communications participated in the study, each receiving 1000 JPY in cash vouchers. Data from 10 participants (nine males and one n/a, aged 21–24) were included in the analysis, as one participant was unable to complete the task properly owing to improper use of the HMD.

Materials

The object shown in Fig. 12 was used to display the target position for gazing at the participants. This object, a part of a hollow sphere with an inner radius of 0.5 m, featured holes of diameter 12 mm at 10° intervals from $\theta = 0^\circ$ to 30° and at 30° intervals from $\phi = 0^\circ$ to 330° , as shown in the figure, from the center of the surface. This object was placed in front of the avatar, such that the center of the sphere aligned with the center of the avatar's two eyeballs (see Fig. 13). LEDs were placed in each hole and illuminated to indicate the target point to the participant.

The workload was evaluated using the Japanese version [15] of the NASA-TLX [16], a subjective method for evaluating workload. The participants rated each of the six NASA-TLX endpoints on a scale of 0–100. To omit the weighting process, Raw TLX (RTLX), which is the simple average of the six NASA-TLX scores, was utilized to estimate the overall workload [17].

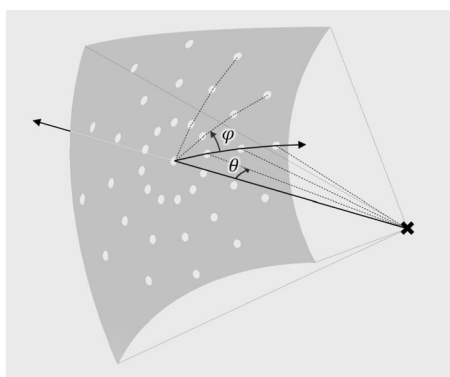


Fig. 12 Experimental object to display target points

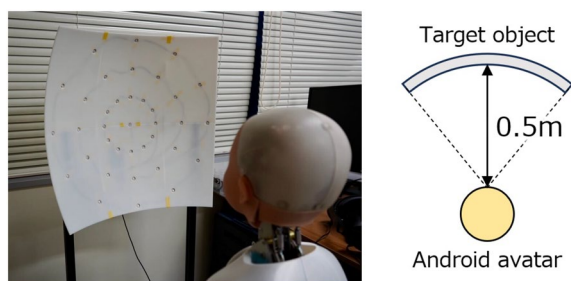


Fig. 13 Experimental setup for the avatar side

Procedure

The design, task details, and experimental procedures used in this study are outlined below.

Design

The experiment was conducted using a between-subject study design under Fixed and Sync conditions. Six participants performed the task in the Fixed condition, whereas the remaining five participants completed the task in the Sync condition. Data from five participants in the Fixed condition (five males, aged 22–23) and five participants in the Sync condition (four males, one n/a, aged 21–24) were analyzed, with the exception of one participant in the Fixed condition who had a problem.

Task

We designed a task for the participants to gaze at an indicated position on an object placed in front of the android avatar. The participants wore HMD and manipulated only their avatar's eyes by moving their own eye-balls. Each participant held a button in each hand to manipulate the avatar's gaze.

During the experiment, an LED illuminated one of the holes on the object in front of the avatar (Fig. 13). Participants focused on the illuminated point for approximately 5 s and then pressed the button in their right hand. Subsequently, they pressed the button in their left hand to

move the illuminated point, repeating the process. The LEDs were activated in counterclockwise order from $\phi = 0$, starting from the center ($\theta = 0^\circ$) and from the circle of $\theta = 10^\circ$ to $\theta = 30^\circ$. The task was terminated when the aforementioned trials were completed for all target points. Participants were instructed to re-press the button if they blinked or if any noise disrupted the camera image while pressing the button and looking at the target point before pressing the button in their left hand.

The direction of the participant's gaze was recorded as a measure of the accuracy of the system when the participant pressed the button while gazing at each target point.

Experimental Steps

Participants were informed that the study aimed to explore the functionality of a teleoperated robot and gather their feedback on it

First, the participants were instructed to perform eye calibration while wearing the HMD. A tool officially provided by Meta, the distributor of the HMD, was used for eye calibration.

Before starting the task, an operational test was conducted on android avatars. With the avatar's eye movements synchronized, the participants were instructed to move their eyes in different directions to ensure that the avatar's movements and camera image were functioning correctly.

Once the operation test was completed, the experimental task outlined in section 4.1.3 was conducted.

All images displayed on the HMD when the participant pressed the buttons were recorded to ensure that noise in the camera images did not significantly influence the experimental results.

Ethical considerations

The study protocol was approved by the ethics committee of the University of Electro-Communications (No. H23034), and all participants provided written informed consent before the experiments.

Results

All of the camera images recorded during the task were noise-free, indicating proper data acquisition.

Using the gaze direction data, the focal point on the object when viewed from the avatar's perspective (referred to as "actual focal point" hereafter) was calculated. Furthermore, the target gaze direction was determined by the position of each target point in relation to the avatar's position. The deviation angle was then calculated as the angle between the target and actual gaze direction vectors.

The actual focal points of the participants under each condition are shown in Fig. 14, showcasing the deviation angle from the target gaze direction as the

median for the five data points at each target point. Data for all participants are shown in Fig. 26 in Appendix A.

Additionally, the deviation angles between the target and gaze directions at each target point are shown in Fig. 14 as box plots. The results of the Mann–Whitney U test ($\alpha = 0.05$, one-sided) revealed that the deviation angle was significantly smaller for the Sync condition than for the Fixed condition ($p < 0.05$) for all target points except $(\theta, \phi) = (0^\circ, 0^\circ), (10^\circ, 0^\circ), (30^\circ, 90^\circ), (30^\circ, 210^\circ)$. Moreover, for $(\theta, \phi) = (30^\circ, 210^\circ)$, the Sync condition exhibited a significant trend for the deviation angle to be smaller than that of the Fixed condition ($p < 0.1$). The medians and test results are shown in Table 3 in Appendix A.

The workload results for each condition are shown in Fig. 15. The result of the Mann–Whitney U test ($\alpha = 0.05$, one-sided) indicated no significant difference in the RTLX scores between the Fixed condition ($Mdn = 43.3$) and Sync condition ($Mdn = 40.5$) ($U = 10, p = 0.345$).

Experiment II: subjective evaluation

Second, an experiment for a subjective evaluation was conducted to investigate whether the developed system effectively enhanced eye contact through an android avatar. The hypotheses of this experiment are as follows:

- When the operator gazes at the observer through the avatar, the observer perceives a stronger mutual eye contact with the avatar in the Sync condition.
- The disparity between the Fixed and Sync conditions intensifies as the observer's position shifts further away from directly facing the avatar.
- The operator's workload is lower in the Sync condition.

Method

In this experiment, we recruited two pairs of participants. During the experiment, one participant was designated as the operator, whereas the other was assigned the role of observer. Our study aimed to determine whether the observer perceived the avatar as making eye contact with them, as well as to assess the workload experienced by both the operator and observer when the operator directed the avatar to gaze at the observer at a specific location. The participants, equipment, materials used, and the experimental procedure are outlined below.

Participants

Twenty-two students from the University of Electro-Communications participated in the experiment in pairs, with each participant receiving cash vouchers worth 1,000 JPY as a token of appreciation. Data from nine of the 11 pairs of participants (12 males and six females, aged 19–24) were analyzed, excluding two pairs who encountered technical difficulties or issues with the experimental equipment.

Materials

The experimental setup is shown in Fig. 16. Eight height-adjustable chairs were placed in front of the avatar as seats for the observer, as shown in Fig. 17a. Because the distance from the avatar could influence the perception of the gaze direction, they were placed in two rows at distances of approximately 1.2 m and 2.1 m from the avatar (see Fig. 17b). These distances were chosen to represent personal distance (used mainly in conversations involving individuals with close relationships) and social distance (utilized primarily in formal conversations, such as in business settings) defined by Hall [18]. The seating arrangement was carefully planned to ensure that the avatar's eye movements were within a natural range when making eye contact with the seated participants. Additionally, the positioning of the chairs was designed to prevent any interference from the participants during the experiment. A display was positioned next to the avatars to guide the observer to their designated seat. The display shows an image similar to that in Fig. 18, which visually indicates the sitting positions of participants (Fig. 19).

To assess the subjective difficulty of maintaining eye contact, a 5-point Likert scale questionnaire was utilized. Participants were asked to rate the level of difficulty they experienced when making eye contact with the other person through the robot, with 1 point indicating “not difficult” and 5 points indicating “very difficult.” As in Experiment I, the RTLX of the Japanese version of NASA-TLX was utilized to evaluate the workload.

Procedure

The design, task content, and experimental procedures employed in this study are outlined below.

Design

The experiment was conducted using a within-subjects study design, comparing the Fixed and Sync conditions. Six pairs completed the task in the Fixed to Sync condition, whereas the remaining five pairs completed the task in the Sync to Fixed condition. Data from five participants in the Fixed to Sync condition were included in the analysis, with one participant from the Fixed condition

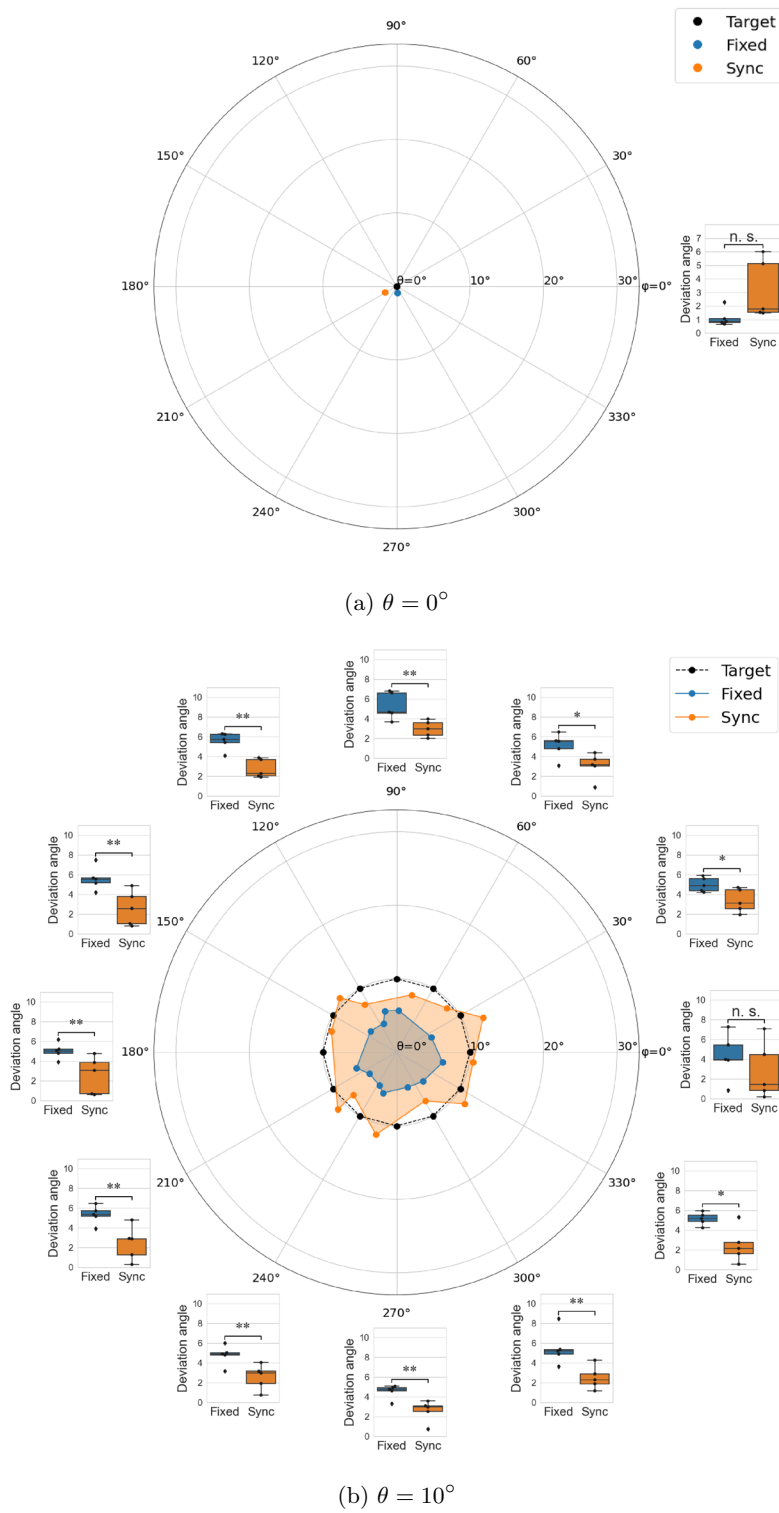
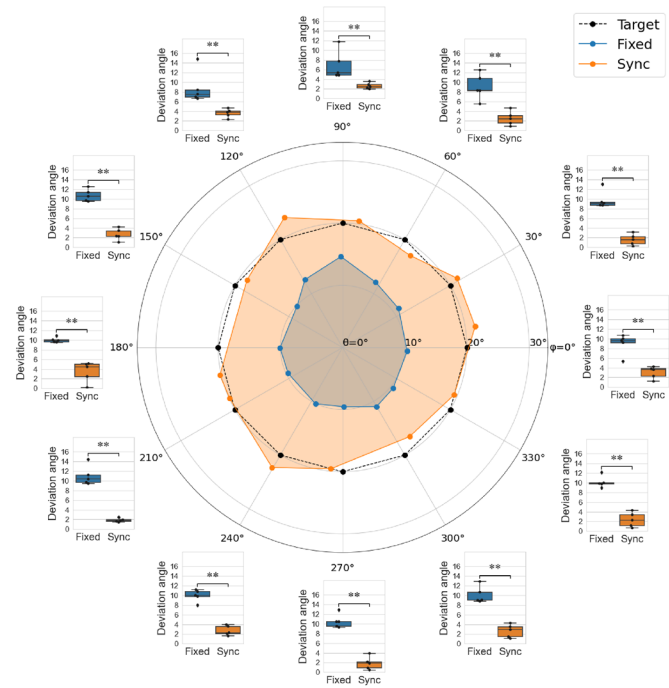
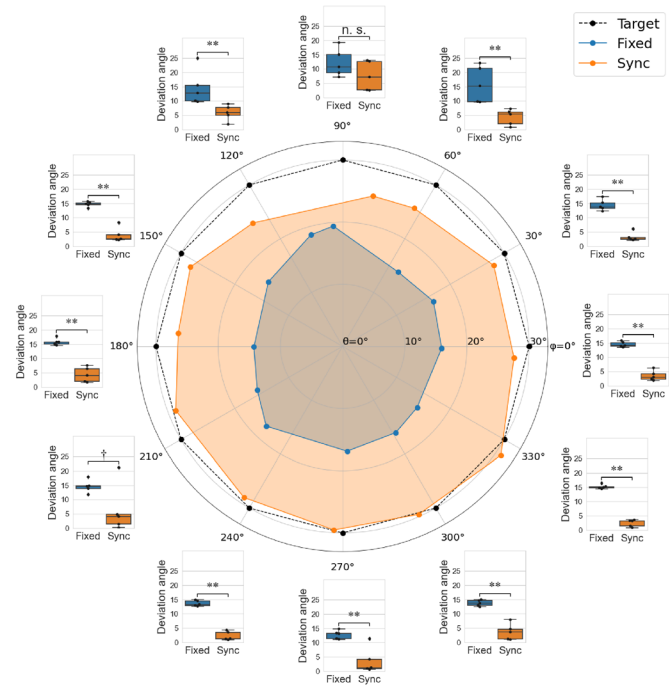


Fig. 14 Median of the participants' actual focal point data and deviation angles between target and actual focal points at each target point ($\dagger p < 0.1, *p < 0.05, **p < 0.01$)



(c) $\theta = 20^\circ$



(d) $\theta = 30^\circ$

Fig. 14 continued

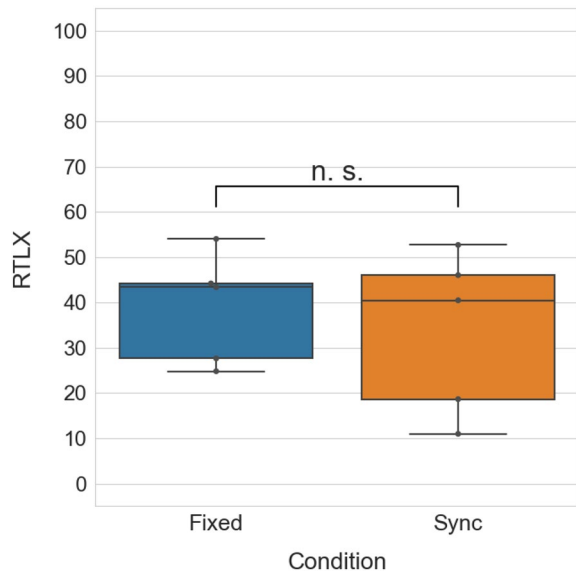


Fig. 15 RTLX scores

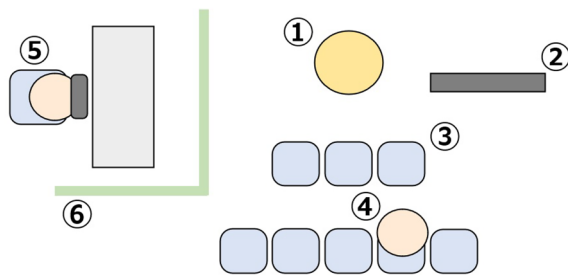


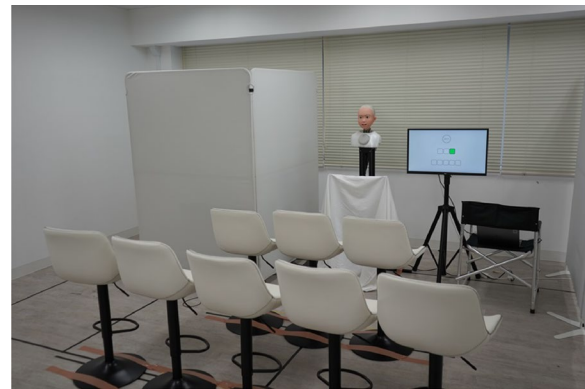
Fig. 16 Layout of the experiment site: (1) android avatar “Yui”; (2) display; (3) height-adjustable chairs; (4) observer; (5) operator; (6) partitions

excluded owing to an issue. The gender and age distributions of the participants, categorized according to the order of conditions and roles, are listed in Table 2.

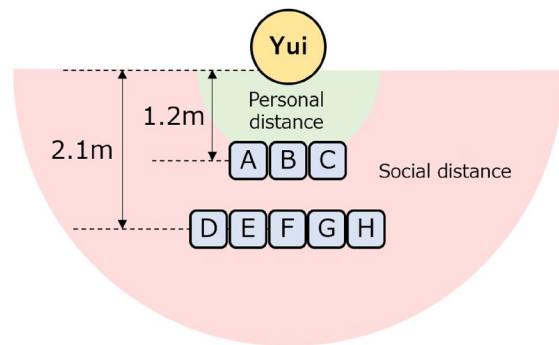
Task

Two experimental participants—an operator and a face-to-face participant—were instructed to make eye contact through the android avatar. The participants wore HMD and manipulated only their avatar’s eyes by moving their eyeballs. Participants performed the task using a button held in each hand. Throughout the experiment, participants wore headphones and listened to white noise

The operator was instructed to make eye contact with the observer and press a button on their hand while the observer was seated in the designated position. The operator maintained eye contact with the observer until the final sound emanated from the headphones. Following the conclusion of the sound, the observer was



(a) Eight chairs placed in front of the avatar



(b) Distance between the avatar and chairs

Fig. 17 Layout of eight height-adjustable chairs

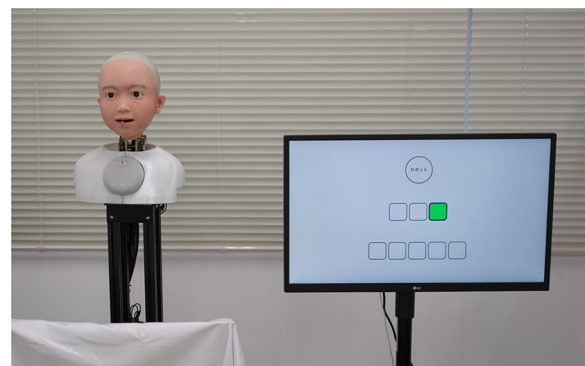


Fig. 18 Display indicating sitting positions

instructed to relocate to a different seat and the process was repeated.

The observer was directed to move to the designated seat as indicated on the display, facing the avatar. Once seated, a 3-s countdown sound was played from headphones worn by the observer. During the time between the end of the countdown and final sound, the observer was prompted to press a button if they felt the avatar was

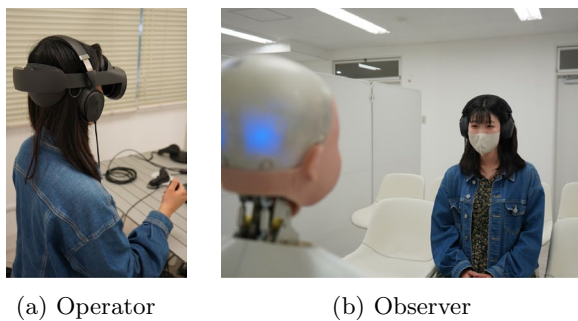


Fig. 19 Photograph of the reproduction of participants during the experiment. These photographs show an author wearing the HMD in place of the participants to recreate the scene during the experiment

gazing at them. If they did not sense the avatar's gaze, they were to refrain from pressing the button and await the final sound. After the final sound was heard, participants checked the display, moved to the next seat, and repeated the same process.

A countdown sound was played through the observer's headphones when the operator pressed a button. The final sound was played simultaneously on both the operator's and observer's sides, 7 s after the countdown ended in the observer's headphones. This interval was established to allow ample time for the observer to press the button again if necessary, ensuring the accuracy of their intuitive judgment. A click sound would indicate when the button was successfully pressed. Observers were instructed to press the button upon feeling the avatar's gaze, to press it again if the click sound was not heard, and to understand that the outcome would not be influenced by the number of button presses until the final sound was heard. The displayed image was switched simultaneously with the final sound.

The operator and observer repeated the aforementioned trial twice for each seat, totaling 16 times per condition. Regarding the order of seats designated for the observers, four sets were created by randomly arranging eight locations in advance. The order of the first and second tasks was determined by connecting the two sets for each task. Care was taken to ensure that the same seat did not appear consecutively when combining random sets. Regardless of the condition conducted first, the order of the seats for the first and second tasks remained consistent for all pairs.

To evaluate the ease of eye contact with the avatar, we recorded whether the observer pressed the button in each trial. The results of gaze perception were analyzed by seat, and for each condition, eye contact was considered possible if the button was pressed both times for the same seat.

Experimental Steps

Participants were informed that the study aimed to investigate the operation of the teleoperated robot and the impressions of the operator. Before the experiment, the experimenter randomly assigned the roles of the operator and observer. In the event that one of the two participants is wearing glasses, the individual donning the eyewear will be designated as the observer. This was to accommodate participants whose glasses may hinder wearing the HMD. The roles of the operator and observer remained consistent for both tasks.

First, the participant was instructed to adjust the chair height to align its eye level with that of the avatar. For height adjustment, a portable object with a viewing hole at eye level of the avatar was used. The observer was instructed to adjust the chair for all eight seats to ensure that the avatar's eyes were visible through the viewing hole while seated.

The operator was instructed to perform eye calibration while wearing an HMD. Following calibration, an operation test was conducted under the conditions to be performed first. When the avatar's eye movements were synchronized, the operators moved their eyes in all directions to validate that the avatar was moving properly and that the camera image was seen appropriately. Additionally, the operator was informed that the avatar's head remained fixed and they would not be able to alter their view by moving their own head.

Following the operation test, the operator and observer were both provided with buttons and headphones that played white noise, and a test trial of the experimental task was conducted once. The operator was instructed to press the button on hand upon the observer taking their seat. The observer was instructed to sit in seat B in front of the avatar (Fig. 17b) and press the button following the countdown sound. Upon completion of the final sound, the operator validated its successful playback. Simultaneously, the observer validated the seamless execution of the countdown sound, click sound, and final sound.

The main experiment commenced after the trial run was completed. Once the initial seat was displayed on the screen, the observer moved on to the experimenter's signal, marking the start of the task. Upon task completion, both the operator and observer completed a questionnaire regarding the difficulty of maintaining eye contact and workload.

Following the questionnaire, the process from the test trial to the questionnaire was repeated under all other conditions.

During the task, the images displayed on the operator's HMD were saved from the end of the countdown until the final sound was played, that is, while the

Table 2 Details of participants in Experiment 2

All participants	Gender	Age
Fixed → Sync	7 males, 1 female	19–24
Sync → Fixed	5 males, 5 females	19–23
All	12 males, 6 females	19–24
Operators	Gender	Age
Fixed → Sync	4 males, 0 female	19–24
Sync → Fixed	3 males, 2 females	19–23
All	7 males, 2 females	19–24
Observers	Gender	Age
Fixed → Sync	3 males, 1 female	19–23
Sync → Fixed	2 males, 3 females	19–23
All	5 males, 4 females	19–23

observer judged the gaze direction. This process was crucial to ensure that any noise present in the camera images did not have a significant impact on the results. Considering the PC load, images were saved every 0.5 s. Additionally, to allow for proper analysis in case of any discrepancies, the behaviors of the avatar and observers were recorded with a video camera.

Ethical considerations

The research protocol employed in this study was approved by the ethics committee of the University of Electro-Communications (No. H23034(2)). All participants provided written informed consent prior to the experiment.

Results

Among the images displayed on the HMD, instances of noise were limited to one or fewer per trial, leading to the conclusion that noise did not significantly influence the outcomes of the experimental outcomes.

The results of the observer's gaze direction judgment for each condition are shown in Fig. 20. The success rate for establishing eye contact was calculated for each seat based on the percentage of observers who pressed the button in both trials. The outcomes of the McNemar test ($\alpha = 0.05$, one-sided) indicated minimal differences in success rates between the Fixed and Sync conditions for any seat. A contingency table of the number of successful and failed trials for each seat and the test results are shown in Table 4 and Table 5 in Appendix B.

The eye contact difficulty ratings for both the operator and observer under each condition are shown in Fig. 21. The outcomes of the Wilcoxon signed-rank test

($\alpha = 0.05$, one-sided) indicated minimal differences in the difficulty ratings for the operator between the Fixed ($Mdn = 3$) and Sync conditions ($Mdn = 2$) ($W = 8.5$, $p = 0.609$). However, for the observers, a significant trend was observed, indicating a decrease in difficulty ratings from the Fixed ($Mdn = 4$) to the Sync condition ($Mdn = 1$) ($W = 3.5$, $p < 0.1$).

To assess the workload, the mean value of each NASA-TLX item was calculated to determine the RTLX score. The RTLX scores for each condition were then compared, as shown in Fig. 22. The results of the Wilcoxon signed-rank test ($\alpha = 0.05$, one-sided) showed no significant differences in the operator's RTLX scores between the Fixed ($Mdn = 44.8$) and Sync conditions ($Mdn = 28.5$) ($W = 18.0$, $p = 0.326$). Similarly, no significant differences were observed in the observer's workload between the Fixed ($Mdn = 25.7$) and Sync conditions ($Mdn = 9.5$) ($W = 22.0$, $p = 0.5$).

Discussion

Findings from experiment I

The findings of Experiment I indicated that in the Sync condition, the gaze direction of the operator when viewing an object through the avatar closely matched the gaze direction when viewing the object from the avatar's location, compared with the Fixed condition. In the Fixed condition, the deviation angle was approximately half of θ of the target point, whereas in the Sync condition, except for the three points at the target coordinates ($\theta, \phi = (30^\circ, 60^\circ), (30^\circ, 90^\circ), (30^\circ, 120^\circ)$), all deviation angles remained below 5° . One possible explanation for the remaining deviation angle could be attributed to the accuracy of mapping the scale of the camera images in the implemented system, of the captured images with a

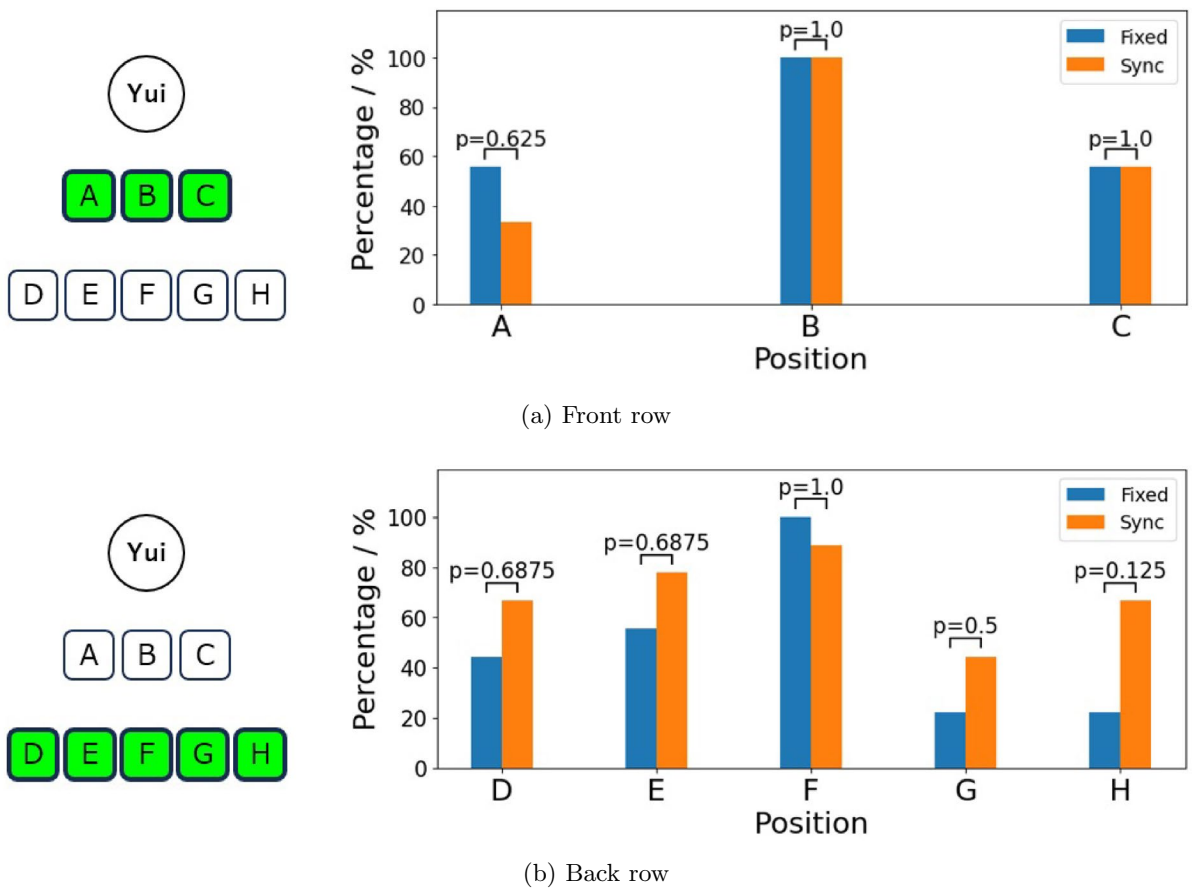


Fig. 20 Success rate of eye contact at each seat

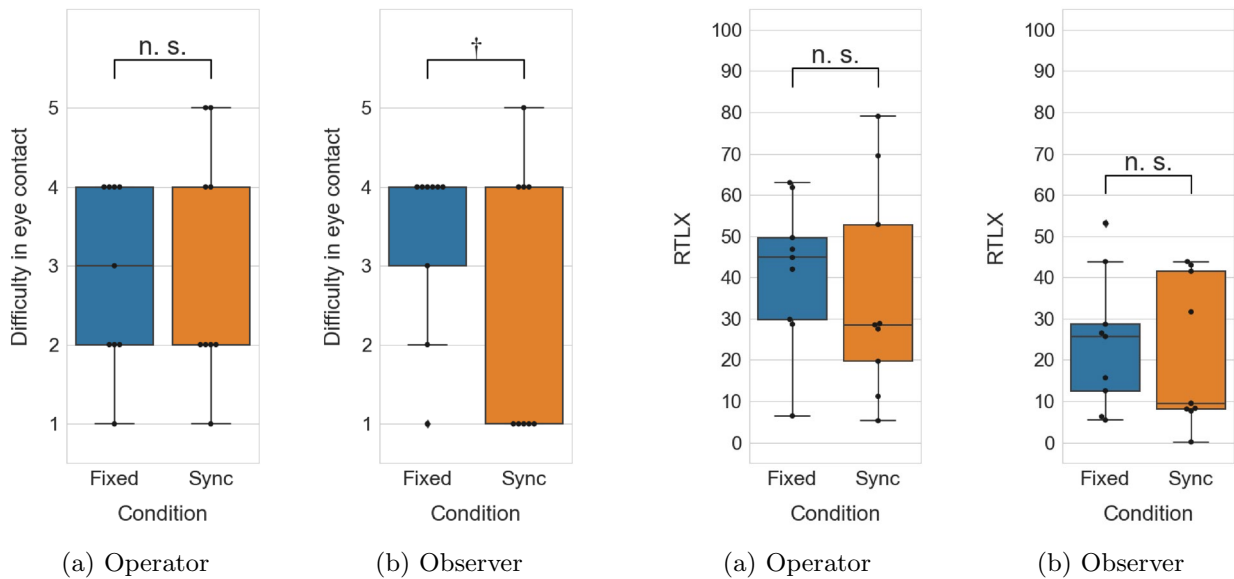


Fig. 21 Difficulty of eye contact ($tp < 0.1$)

Fig. 22 RTLX scores

horizontal angle of view of 200° , those with a range of 180° were pasted onto a hemispherical screen. However, uncertainty in the horizontal angle of view of the camera may have impacted the mapping accuracy, resulting in the images not being displayed at a scale equal to that of the real world. The utilization of a weighted-average filter when inputting screen rotation angles may also be a contributing factor.

From $\phi = 60^\circ$ to 120° at $\theta = 30^\circ$, an increase in the deviation angle was observed in the Sync condition. This could be attributed to the positioning of the eyeball slightly above the center of the display when wearing the HMD utilized in this study. Consequently, for some participants, these target points may have been situated near the upper edge of the HMD's field of view, making them difficult to focus on in the Sync condition, where the view remained static with eye movement. Free responses in the questionnaire included comments from participants in the Sync condition indicating that "looking upwards was difficult and tiring." Additionally, for the aforementioned reasons, eye calibration was not accurately performed in some participants.

Moreover, the absence of significant differences in workload evaluation between conditions could be attributed to the psychological burden experienced by participants in the Sync condition, stemming from the challenge of focusing on upper target points. Conversely, in the free responses, three out of five participants in the Fixed condition expressed frustration caused by screen vibrations: "It was difficult to gaze continually at a single point because the screen was shaking," "I felt some stress when my gaze did not go where I wanted it to," and "While keeping my gaze fixed, the camera moved and the light shifted from my line of sight." Under the Fixed condition, the view changed with eye movements, causing the screen to vibrate when focusing on a single point. In contrast, the Sync condition compensated for changes in view resulting from eye movements, enabling a stable focus on a single point. This underscores the superiority of the proposed system over the conventional system for utilizing avatars as communication tools.

Findings from experiment II

From the results of Experiment II, although no significant differences were observed in eye contact success rates between conditions for any seat, the Sync condition achieved higher success rates for all seats in the back row, except for the central seat F. In the front row, over 50% of the participants in the Fixed condition believed they could make eye contact with the avatar. The seating arrangement of this experiment positioned the seats at the ends of the front row approximately $\pm 22^\circ$ from the forward direction of the avatar. Because

the deviation angle from the target gaze direction in the Fixed condition was approximately half of the target angle, the theoretical difference in the gaze angle between the Fixed and Sync conditions was approximately 11° . This difference may not have been sufficiently large to influence the perception of eye contact in some participants.

Additionally, a bias in the success rate was observed between the left and right side seats in both conditions in the back row. Analysis of the avatar's gaze angles revealed a consistent bias of approximately 3° toward seat H from the central seats for all participants in both the front and back rows. Furthermore, the amplitude of the gaze angle at seat C in the front row was approximately 4° larger than that at seat A, and the amplitude at seat H in the back row was approximately 4° larger than that at seat D. This suggests that the avatar's head was likely oriented approximately 2° to 3° toward seats A and D from the central seats, leading to lower success rates for seats G and H than for seats D and E. Conversely, at seat H, which had the largest amplitude of gaze angle, the Sync condition demonstrated a more pronounced increase in success rate compared to the Fixed condition, highlighting the effectiveness of the proposed system in areas with larger gaze angle amplitudes.

Furthermore, a significant trend toward lower subjective difficulty in making eye contact was observed in the Sync condition. In the free responses from the questionnaire completed immediately after performing the task in the Fixed condition, eight out of nine observers encountered challenges in maintaining eye contact through the avatar, with comments such as "it was difficult to align my gaze." No positive response was observed. In contrast, in the free responses from the questionnaire completed after the Sync condition, while four out of nine observers similarly mentioned challenges in establishing eye contact, three others left positive remarks, such as "I was surprised that my gaze hardly deviated," and "It was much easier to make eye contact than I expected, it felt almost human." The subjective difficulty in making eye contact varied among participants, with five participants stating it was "not difficult at all," whereas four participants found it "very difficult" or "somewhat difficult," indicating considerable individual differences. However, for some participants, the Sync condition likely contributed significantly to reducing the difficulty of making eye contact.

Although the results from Experiment I suggested a potential reduction in the operator's workload, the present experiment similarly showed no significant difference in the operator's workload. Among the operators who commenced the experimental task with the Sync condition, four out of five rated the workload higher in the Fixed condition during the second task. In contrast,

for those who started with the Fixed condition, only two out of four operators rated the workload lower in the Sync condition during the second task compared with the Fixed condition. One possible explanation is that in the Fixed condition, the ability to change the view with eye movements may have caused operators who started with this condition to perceive a loss of control over the view when performing the Sync condition, thereby experiencing an increase in workload. However, in the questionnaires completed immediately after performing the task in the Sync condition by one of the operators who began with the Fixed condition, a statement was made that “my eyes were tired after the first task, but they were less tired after the second task,” indicating that the proposed system is effective in reducing workload during operation.

Additionally, in the free responses from the questionnaire completed after performing the task in the Sync condition, a comment was made that “I usually find it somewhat uncomfortable to make eye contact when facing someone, but I did not feel that way much during this experiment.” The avatar’s ability to maintain a greater physical distance from the counterpart than face-to-face interactions may reduce the discomfort associated with engaging in close eye contact. The use of an android avatar to establish eye contact could potentially make communication involving eye contact easier than in a face-to-face environment.

Limitations

The results of Experiment II showed that the success rate for eye contact was below 80% for all seats, except for the central seats, regardless of the condition. In this experiment, participants were tasked with determining whether they felt as though they were making eye contact with an avatar solely based on the direction of its gaze. However, other factors beyond gaze direction can impact the perception of being observed. Based on previous research, head orientation has an important influence on gaze direction perception [19]. Further, eyebrow movements and the degree of eyelid opening and closing may influence the perception of eye contact. Furthermore, comments from the free responses in the questionnaire included statements, such as, “I lost the sense of what it

feels like to make eye contact.” Eye contact is often perceived intuitively. However, in the experiment conducted, participants were given an extended judgment time of 7 s, which may have led to overthinking and potentially diminished their intuitive sense of eye contact. An experiment that can evaluate the perception of eye contact should be performed, as eye contact occurs naturally in interactions.

Conclusion

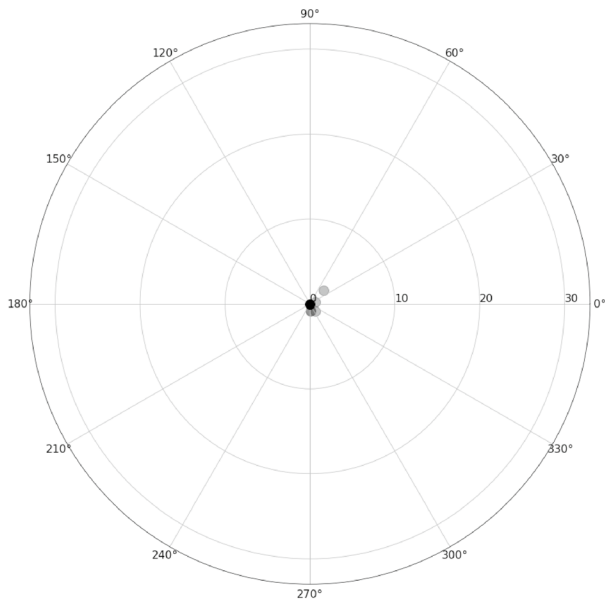
We developed an eyeball integrated with a wide-angle lens camera designed for android avatars. This technology, along with a vision-sharing system, aims to enhance the operation and appearance of the avatar to closely mimic human gaze behavior. We conducted experiments to thoroughly evaluate the effectiveness of this system. Experiments demonstrated that the proposed vision-sharing system enabled operators to perceive their surroundings as if they were actually in the avatar’s location. Additionally, the proposed system demonstrated significant potential in facilitating eye contact through the android avatar.

In the experiments conducted for this study, issues, such as the low resolution of images presented to the operator and the necessity of applying filters to the input rotation angles of the avatar’s eyeball and screen likely influenced the participants’ subjective evaluations owing to communication problems. Addressing these challenges through alternative communication methods could provide a clearer assessment of our system’s effectiveness.

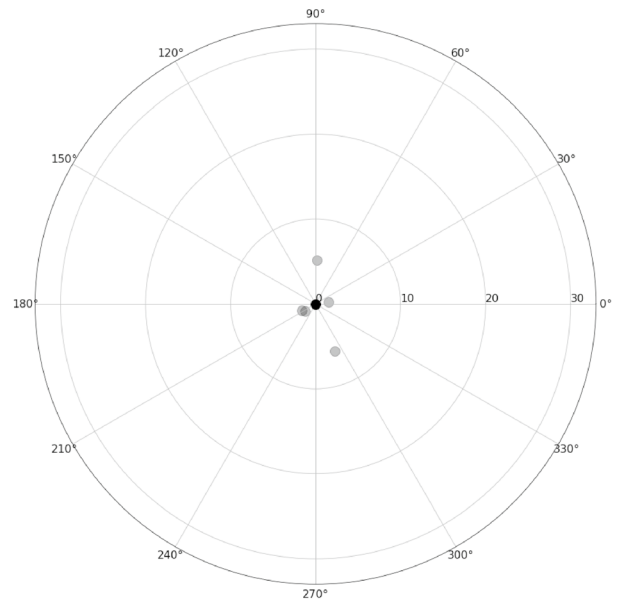
Future research will involve experiments in which communication is performed through the avatar to evaluate the impact of the proposed system on eye contact during communication. Additionally, we will explore facial expression elements beyond eye movements to develop systems that facilitate natural non-verbal communication between the operator and the observer.

Appendix A Experiment I

See Figs. 23, 24, 25, 26 and Table 3.

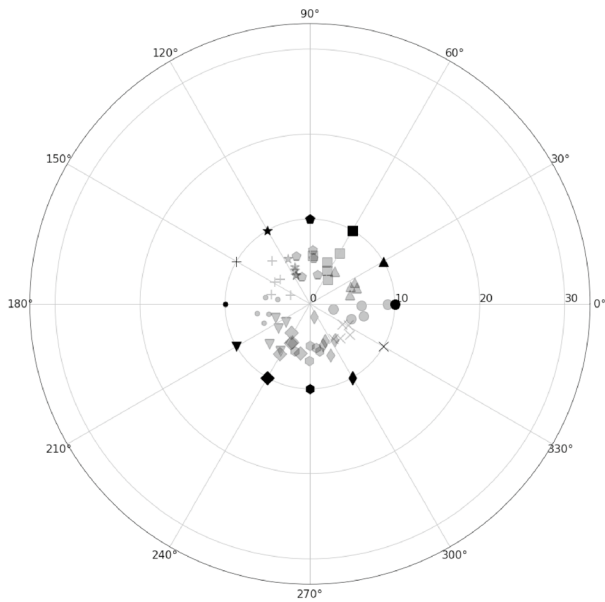


(a) Fixed condition

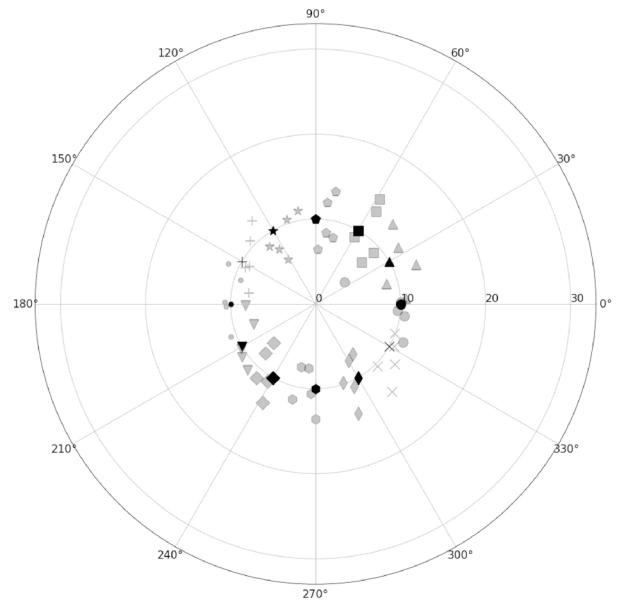


(b) Sync condition

Fig. 23 Actual focal point data at $\theta = 0^\circ$. The black and gray dots indicate target points and experimental data, respectively



(a) Fixed condition



(b) Sync condition

Fig. 24 Actual focal point data at $\theta = 10^\circ$. The black and gray dots indicate target points and experimental data, respectively

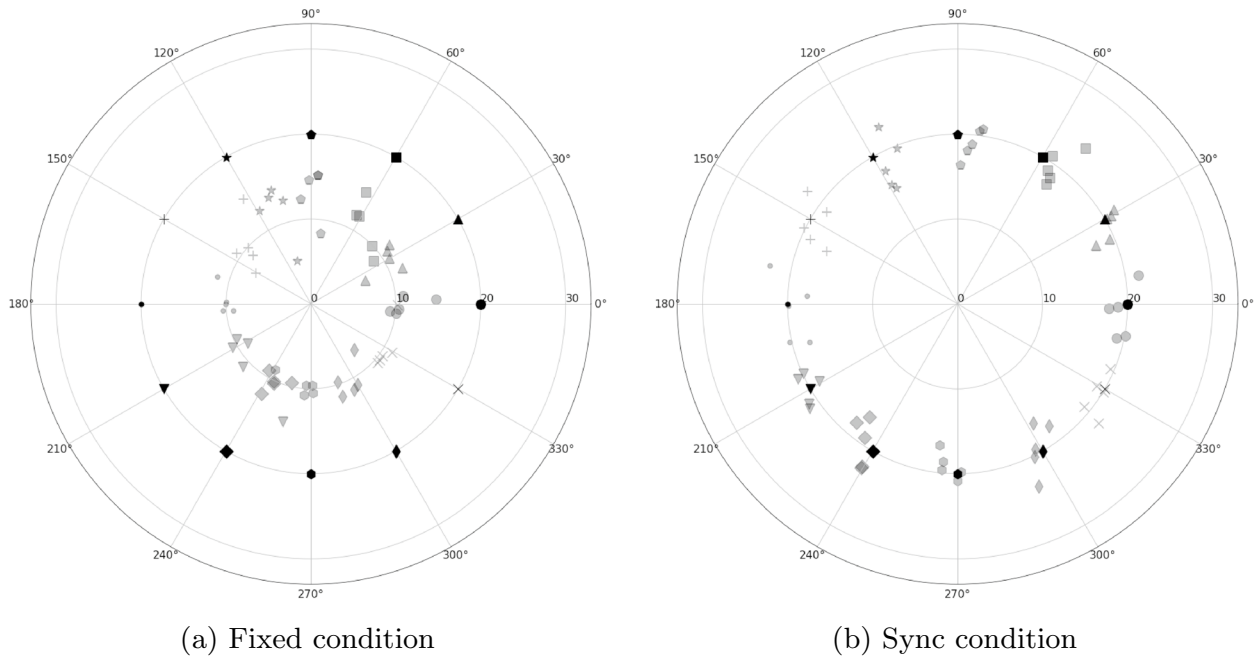


Fig. 25 Actual focal point data at $\theta = 20^\circ$. The black dots indicate target points, whereas the gray dots indicate experimental data

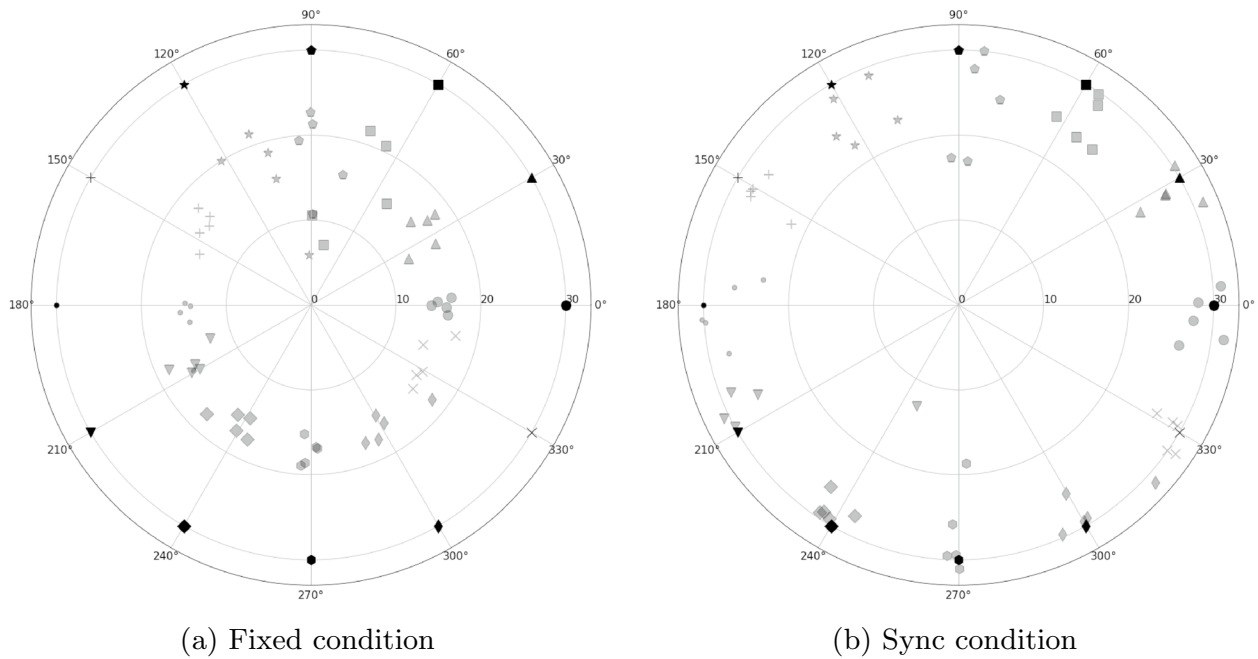


Fig. 26 Actual focal point data at $\theta = 30^\circ$. The black dots indicate target points, whereas gray dots indicate experimental data

Table 3 Results of the Mann–Whitney U-test for median deviation angle at each target point ($\dagger p < 0.1$, $*p < 0.05$, $**p < 0.01$)

θ	ϕ	U	p	Fixed Mdn	Sync Mdn
0	0	3.0	0.984	0.85	1.77
1	0	17.0	0.210	3.97	1.44
1	30	21.0	0.048 (*)	4.91	3.11
1	60	22.0	0.028 (*)	5.56	3.17
1	90	24.0	0.008 (**)	4.66	2.99
1	120	25.0	0.004 (**)	5.74	2.28
1	150	24.0	0.008 (**)	5.57	2.59
1	180	24.0	0.008 (**)	5.06	3.07
1	210	24.0	0.008 (**)	5.37	2.89
1	240	24.0	0.008 (**)	4.91	2.98
1	270	24.0	0.008 (**)	4.78	2.97
1	300	24.0	0.008 (**)	5.21	2.30
1	330	22.0	0.028 (*)	5.18	2.15
2	0	25.0	0.004 (**)	9.62	3.62
2	30	25.0	0.004 (**)	9.13	1.59
2	60	25.0	0.004 (**)	8.28	2.46
2	90	25.0	0.004 (**)	5.36	2.49
2	120	25.0	0.004 (**)	7.52	3.77
2	150	25.0	0.004 (**)	10.60	2.31
2	180	25.0	0.004 (**)	9.94	4.46
2	210	25.0	0.004 (**)	10.40	1.79
2	240	25.0	0.004 (**)	10.00	2.29
2	270	25.0	0.004 (**)	10.50	1.84
2	300	25.0	0.004 (**)	9.05	2.93
2	330	25.0	0.004 (**)	9.84	2.28
3	0	25.0	0.004 (**)	14.10	3.06
3	30	25.0	0.004 (**)	13.80	2.63
3	60	25.0	0.004 (**)	15.30	5.42
3	90	19.0	0.111	10.70	7.24
3	120	25.0	0.004 (**)	12.80	6.01
3	150	25.0	0.004 (**)	14.90	2.66
3	180	25.0	0.004 (**)	15.70	4.15
3	210	20.0	0.075 (\dagger)	14.70	4.14
3	240	25.0	0.004 (**)	13.30	1.39
3	270	24.0	0.008 (**)	13.10	1.29
3	300	25.0	0.004 (**)	13.70	3.70
3	330	25.0	0.004 (**)	15.00	3.24

Appendix B Experiment II

See Tables 4 and 5.

Table 4 Contingency table of the number of successful and failed trials in each seat

Seat A	Fixed: success	Fixed: failed
Sync: success	2	1
Sync: failed	3	3
Seat B	Fixed: success	Fixed: failed
Sync: success	9	0
Sync: failed	0	0
Seat C	Fixed: success	Fixed: failed
Sync: success	3	2
Sync: failed	2	2
Seat D	Fixed: success	Fixed: failed
Sync: success	2	4
Sync: failed	2	1
Seat E	Fixed: success	Fixed: failed
Sync: success	3	4
Sync: failed	2	0
Seat F	Fixed: success	Fixed: failed
Sync: success	8	0
Sync: failed	1	0
Seat G	Fixed: success	Fixed: failed
Sync: success	2	2
Sync: failed	0	5
Seat H	Fixed: success	Fixed: failed
Sync: success	2	4
Sync: failed	0	3

Table 5 Results of McNemar's test for success rates in each seat

Seat	χ^2	<i>p</i>	Fixed success rate	Sync success rate
A	1.00	0.625	55.6%	33.3%
B	0.00	1.000	100.0%	100.0%
C	2.00	1.000	55.6%	55.6%
D	2.00	0.688	44.4%	66.7%
E	2.00	0.688	55.6%	77.8%
F	0.00	1.000	100.0%	88.9%
G	0.00	0.500	22.2%	44.4%
H	0.00	0.125	22.2%	66.7%

Abbreviation

HMD Head-mounted display

Acknowledgements

This work was supported by JST Moonshot R&D Grant Number JPMJMS2011.

Author contributions

Conceptualization: K.S., M.N., Y.N.; methodology: K.S., M.N., Y.N.; software: K.S., M.N.; hardware: M.N., Y.N.; data Curation: K.S.; writing—original draft preparation: K.S.; formal analysis: K.S.; project administration: Y.N.; Supervision: Y.N.; All authors: writing—review & editing.

Funding

JST Moonshot R&D Grant Number JPMJMS2011

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations**Ethics approval and consent to participate**

The research protocols employed in this study was approved by the ethics committee of the University of Electro-Communications (No. H23034 and No. H23034(2)). All participants provided written informed consent prior to the experiment.

Competing interest

The authors declare that they have no competing interests.

Received: 9 August 2024 Accepted: 28 September 2024

Published online: 05 October 2024

References

- Bailenson JN (2021) Nonverbal overload: a theoretical argument for the causes of zoom fatigue. *Technol, Mind, Behav.* <https://doi.org/10.1037/tmb0000030>
- Argyle M, Cook M (1976) *Gaze and mutual gaze*. Cambridge U Press, Oxford, p 210
- Argyle M, Dean J (1965) Eye-contact, distance and affiliation. *Sociometry* 28(3):289–304. <https://doi.org/10.2307/2786027>
- Tanaka K, Nakanishi H, Ishiguro H (2014) Comparing video, avatar, and robot mediated communication: pros and cons of embodiment. In: Yuizono T, Zurita G, Baloiian N, Inoue T, Ogata H (eds) *Collaboration Technologies and Social Computing*. CollabTech 2014. Communications in Computer and Information Science, vol460. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44651-5_9

- Sacino A, Cocchella F, De Vita G, Bracco F, Rea F, Sciutti A, Andrighetto L (2022) Human- or object-like? cognitive anthropomorphism of humanoid robots. *PLOS ONE* 17(7):1–19. <https://doi.org/10.1371/journal.pone.0270787>
- Nakajima M, Shinkawa K, Nakata Y (2024) Development of the lifelike head unit for a humanoid cybernetic avatar 'Yui' and its operation interface. *IEEE Access* 12:23930–23942. <https://doi.org/10.1109/ACCESS.2024.3365723>
- Glas DF, Minato T, Ishi CT, Kawahara T, Ishiguro H (2016) Erica: the erato intelligent conversational android. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, NY, USA, 2016, pp. 22–29. <https://doi.org/10.1109/ROMAN.2016.7745086>
- Ishihara H, Yoshikawa Y, Asada M (2011) Realistic child robot "Affetto" for understanding the caregiver-child attachment relationship that guides the child development. In: 2011 IEEE International Conference on Development and Learning (ICDL), vol. 2, pp. 1–5. <https://doi.org/10.1109/DEVLRN.2011.6037346>
- Engineered Arts (2024) Ameca. <https://www.engineeredarts.co.uk/robot/ameca/>. Retrieved on June 22,
- Nakata Y, Yagi S, Yu S, Wang Y, Ise N, Nakamura Y, Ishiguro H (2022) Development of 'ibuki' an electrically actuated childlike android with mobility and its potential in the future society. *Robotica* 40(4):933–950. <https://doi.org/10.1017/S0263574721000898>
- Darvish K, Penco L, Ramos J, Cisneros R, Pratt J, Yoshida E, Ivaldi S, Pucci D (2023) Teleoperation of humanoid robots: a survey. *IEEE Trans Robot.* <https://doi.org/10.1109/TRO.2023.3236952>
- Shinkawa K, Hirayama T, Nakajima M, Nakata Y (2023) Development of an eyeball integrated with a wide-angle lens camera and vision-sharing system for android avatars. In: *ROBOMECH, 1A2-C17*. (in Japanese)
- Shinkawa K, Nakajima M, Nakata Y (2024) Vision-sharing system for android avatars: Enabling operator eye movement synchronization and immersive presentation of avatar sight. *ICRA2024 Workshop: Workshop on 'Society of Avatar-Symbiosis through Social Field Experiments'*
- Shinkawa K, Nakata Y (2023) Gaze movement operability and sense of spatial presence assessment while operating a robot avatar. In: *IEEE/SICE International Symposium on System Integration (SII)*, Atlanta, GA, USA, 2023, pp. 1–7. <https://doi.org/10.1109/SII55687.2023.10039342>
- Haga S, Mizukami N (1996) Japanese version of NASA task load Index sensitivity of its workload score to difficulty of three different laboratory tasks. *Jpn J Ergonom* 32(2):71–79. <https://doi.org/10.5100/jje.32.71>
- Hart SG, Staveland LE (1988) Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv Psychol* 52:139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hart SG (2006) NASA-task load index (NASA-TLX); 20 years later. *Proc Human Factors Ergonom Soc Ann Meet* 50(9):904–908. <https://doi.org/10.1177/154193120605000909>
- Hall ET (1966) *The hidden dimension*. Doubleday & Company Inc., New York City, USA
- Gamer M, Hecht H (2007) Are you looking at me? Measuring the cone of gaze. *J Exp Psychol: Human Percept Perform* 33(3):705–715. <https://doi.org/10.1037/0096-1523.33.3.705>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.