Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Virtual Reflections on a Dynamic 2D Eye Model Improve Spatial Reference Identification

Matti Krüger, Yutaka Ooshima, Yu Fang

2026

Preprint:

This is an accepted article published in IEEE Transactions on Human-Machine Systems. The final authenticated version is available online at: https://doi.org/10.1109/THMS.2026.3651818 Copyright 2026 IEEE

This article has been accepted for publication in IEEE Transactions on Human-Machine Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/THMS.2026.3651818

Virtual Reflections on a Dynamic 2D Eye Model Improve Spatial Reference Identification

Matti Krüger, Yutaka Oshima, and Yu Fang, *Member, IEEE*

***Abstract*—The visible orientation of human eyes creates some transparency about people's spatial attention and other mental states. This leads to a dual role of the eyes as a means of sensing and communication. Accordingly, artificial eye models are being explored as communication media in human-machine interaction scenarios. One challenge in the use of eye models for communication consists of resolving spatial reference ambiguities, especially for screen-based models. To address this challenge, we introduce an approach that incorporates reflection-like features that are contingent on the movements of artificial eyes. We conducted a user study with 30 participants in which participants had to use spatial references provided by dynamic eye models to advance in a fast-paced group interaction task. Compared to a non-reflective eye model and a pure reflection mode, the superimposition of screen-based eyes with gaze-contingent virtual reflections resulted in a higher identification accuracy and user experience, suggesting a synergistic benefit.**

***Index Terms*—eye model, nonverbal behavior, spatial referencing, human-robot interaction**

I. Introduction

FIRST and foremost, the eyes are sensory organs that acquire visual information from the environment. Sharp visual perception through the eyes relies on the projection of light from a scene at a depth of interest onto light-sensitive tissue inside the eye. To achieve such a projection selectively, the eyes have evolved mechanisms that allow us to sample a scene through various actions that alter, for instance, eye orientation, pupil diameter, and lens curvature. One byproduct of this action-guided visual perception is that aspects of cognition become observable. For example, eye movements and the direction of the pupil reveal where a person is looking and, when further following the gaze direction, what a person is currently attending to. Thus, eyes may be understood not only as sensory organs but also as communication devices [1], [2]. Indeed, in support of their informative value, eyes have been found to be a particularly strong attractor for other people's gaze (see, e.g., [3]–[5]) and the exploitation of gaze direction transparency appears to even have evolved into reflexive gaze following behavior [6]. According to the cooperative eye hypothesis [1], [7], human eyes may have developed their salient features precisely to increase their social utility. The social role of the eyes is also emphasized by the participation of the eyes in the expression of and influence on emotions [8]–[10] as well as the particular sensitivity to eye contact [11]. Due to the perceptual salience of the eyes, communication abilities, and social role, models that mimic human eyes are regularly used as communication tools, especially in virtual characters and robots designed for human-machine interaction (e.g., [2], [12]–[17]). If implemented successfully, robotic eye models have been found to elicit similar psychophysiological responses in humans as the eyes of human interaction partners [18], [19]. Such findings suggest that eye models may be purposefully used in interactive and especially in cooperative [20]–[22] settings, e.g., to capture a person's attention intuitively, communicate directions of interest, or indicate specific mental states.

A. Challenges for communicative eye models

However, eye models on virtual characters and robots are often subject to constraints that may limit communication utility. Recreating human-like ocular behavior requires multiple actuators to be driven in a complex conjunction at low latency and with high accuracy (e.g., [23]). This means that design and construction are either complex or a compromise in terms of model performance. Artificial eyes can therefore easily fail to recreate realistic eye movements with human-like latency, accuracy, and context-dependent variability, due to limitations of the available hardware or the behavior-driving models [15]. In consequence, their movements can appear unnatural, scripted, or even disturbing [24], possibly leading to an ignorance of the eyes in the long term. Virtual eyes that are presented on displays (e.g., [15]–[17], [25], [26]) may overcome many limitations related to latency and available degrees of freedom but can have difficulties in unambiguously conveying spatial information due to their two-dimensional screen projection.

The ability to follow the gaze of another person is partially dependent on the spherical shape of the eyes, which produces visual cues that help estimate the orientation of each eye. Considering the relative orientations of both eyes can reveal not just the direction but the actual distance and even the point of attention in space because the point of convergence of both eyes' orientation (in extension) corresponds to the focus point [27]. Because the angle at the point of convergence becomes smaller with distance and, conversely, increases with proximity, this eye vergence angle (EVA) is a useful indicator of focus distance. The EVA is closely tied to the interpupillary distance (IPD), which varies slightly as a function of the EVA and thus encodes some depth information. In natural 3D settings, this information is reinforced by motion parallax, as small head movements produce dynamic changes in the observer's view that further disambiguate the target of another person's gaze.

Matti Krüger is with Honda Research Institute Europe GmbH, Offenbach am Main, Germany.

Matti Krüger, Yutaka Oshima, and Yu Fang are with Honda Research Institute Japan Co., Ltd., Wako-shi, Saitama, Japan.

Email: firstname.lastname@jp.honda-ri.com|@honda-ri.de

Manuscript received January **, 2025; revised Month* **, 2025.

© 2026 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Human-Machine Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/THMS.2026.3651818

Estimating another person's eye orientation and focus direction becomes more challenging when the eyes are distant or presented in a 2D format, such as on screens. In these cases, contextual cues, such as head orientation and body posture, can support inferences about another person's attention region [4]. However, the lack of depth information can lead to misperceptions of gaze direction and distance, particularly in ambiguous situations like group interactions. While depth may be simulated on 2D displays through artificial motion parallax, e.g., via head-tracking contingent image changes [28], [29], such approaches are limited as they rely on observer movement and do not engage stereopsis. Moreover, adapting depth cues to individual perspectives becomes problematic in multi-person interactions, limiting its utility to one-to-one interaction scenarios.

To overcome the loss of depth information while preserving the benefits of virtual eyes and 3D facial features, researchers have developed retro-projected robot heads [30]. These project images onto a translucent head model, creating a 3D display that conveys spatial information. However, this approach is susceptible to ambient illumination [30], which can cause projected elements to be obscured and create unnatural lighting effects, such as inverted shadows. Retro-projected head and eye models may hence be restricted to environments with sufficiently controlled illumination.

All discussed approaches for creating eye models for communication purposes come with issues that constrain their utility in various ways. The following section aims to introduce an approach that may circumvent some of these issues and extend the social utility of artificial eye models.

B. One-step spatial references

In human interactions, understanding another person's focus of attention involves two steps: looking at their eyes and mapping their gaze onto its target in space. Uncertainty in ocular cues can make this spatial inference unreliable. However, artificial or virtual eyes are not bound by biological constraints but can be visually modified to reduce communicative uncertainty. This raises the question of how their appearance might be adapted to intuitively convey spatial references without interfering with other communicative functions.

Human eyes partially reflect their surroundings like convex mirror, although the small size of the reflection makes it difficult to recognize. In contrast, artificial eyes are not limited by size or physical reflection properties. Reflections on such eyes may be more recognizable, especially because the use of rendered eye models also allows for virtual reflections and projections that can be altered to enhance visibility. Consequently, here we propose to augment artificial eye models intended for human-machine interaction, such as eyes shown on screens, by including virtual mirror images of the machine's current focus of attention in the eye models (see Figure 1). In reference to that feature, we refer to such eye models as *Mirror Eyes*.

Mirror Eyes simultaneously display the direction and content of attention. For an observer, this can turn the described two-step process of understanding spatial references into a single step and remove the need for unambiguously interpretable gaze vectors, substantially facilitating the design of effectively communicating eyes. Although the added detail displayed on such Mirror Eyes could increase information processing demands for an observer, we expect the perception of virtual reflections to be accompanied by a gain in gaze target identification accuracy, especially in situations with potential reference ambiguity, such as group interactions. We further expect the experience of such references to be enhanced in comparison to classical eye models and to pure dynamic virtual reflections that lack eye features, such as pupil and iris, which typically indicate spatial attention.

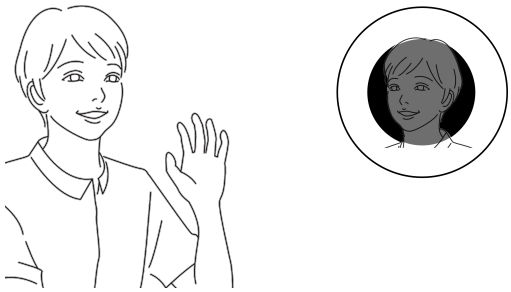


Fig. 1. Illustration of an eye model that displays a mirror image of an attended area as a reflection-like overlay on its pupil and iris.

To formalize the preceding claims, we propose the following research hypotheses. The use of Mirror Eyes ...

H1: improves people's identification accuracy of spatial references compared to eye models that lack mirror augmentation.

H2: reduces the speed of positive spatial reference identification compared to eye models that lack mirror augmentation.

H3: as a spatial reference tool in a group setting improves the user experience compared to eye models that lack either mirror-like augmentation or pupil-like features.

To evaluate these hypotheses, we developed a functional prototype of an eye model with Mirror Eye capabilities and conducted a user study using this model. The remainder of this paper presents both the model and the empirical investigation in detail.

II. Methodology - Prototype

We created a functional prototype that implements the Mirror Eye concept for utilization on a wide range of hardware, such as personal computers with webcams. It relies on the following interconnected modules as depicted in Figure 2:

A. Dynamic eye model

The first module consists of a dynamic 2D eye model that allows real-time movement of pupil and iris images relative to the sclera to create an illusion of rotation. To align these movements with locations of interest in space such that the pupil appears to point towards a target location, a mapping from a camera image space to eye coordinates is applied (see Figure 3 and Equations 3-4). The camera image is mapped from the perspective of the location of the eye model on a screen. To simulate sensitivity to depth despite the flat camera

© 2026 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Human-Machine Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/THMS.2026.3651818

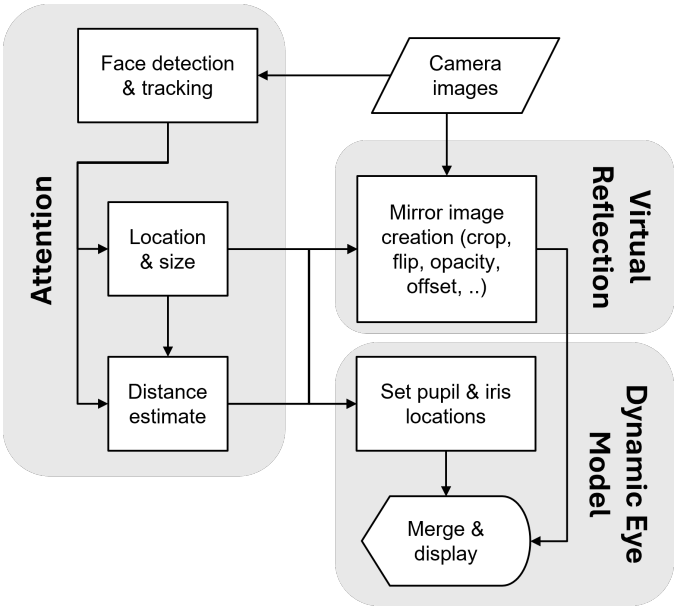


Fig. 2. Connected modules of the Mirror Eye prototype.

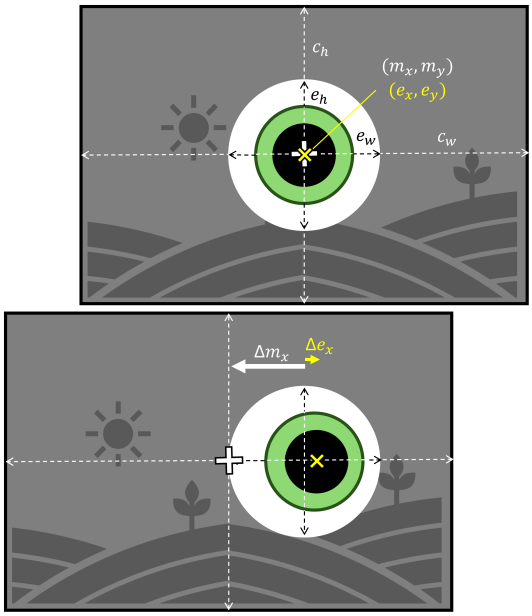


Fig. 3. Illustration of the spatial shift for pupil and mirror image location as described by Equations 1-4. Top: All centers are aligned. Bottom: A small rightward shift of the pupil is accompanied by a larger leftward shift of the mirror image.

image, we have created a module that estimates distances to objects in the scene based on provided object sizes and camera specifications for focal length and sensor size. The depth estimates are then used to adjust horizontal offsets in the case of the simultaneous application of multiple eye models. In accordance with typical vertebrate anatomy, we apply two eye models in conjunction and utilize the depth estimates to create individual horizontal offsets for each eye's pupil and iris such that interpupillary distance varies with focus distance, resulting in an effect of binocular disparity. This allows an observer to infer depth information from alignment differences between the eyes and thus may facilitate gaze target recognition.

B. Attention

To make the eye modules come to life by moving across a scene in a meaningful way, an attention module is required. Simply put, the task of such a module is to define and continuously update target coordinates in camera space that are then mapped to pupil and iris position as described above. Because we were interested in the application of eye models as referencing tools in social settings, we chose to let the eyes look at and follow people in a scene. For this purpose, we implemented a face-tracking module that can detect and simultaneously track multiple faces present in a camera feed. For the purpose of the present work, we based the face-tracking on the spatiotemporal proximity between individual face detections. With this method, faces detected at subsequent time steps near previous detections in image space are assigned the same identifier. This enables smooth tracking without the need for individual face learning or recognition for scenarios in which the minimal distance between multiple people can be well controlled. Target coordinates are then defined relative to one or multiple recent coordinates of face identifiers, allowing the eyes to look at, avoid, follow, and switch between faces.

C. Virtual reflection

Having established dynamic eye models that appear to look at and follow faces in a scene, the next module adds the virtual reflection capability consisting of the following steps: For each eye, an overlay of a horizontally flipped camera feed is created at a chosen transparency level. To indicate what part of the scene is being attended, the pupil should be located at or near the region of interest on the respective camera image. However, only moving the full eye model across the large image space to point to objects in the scene would result in a loss of the previously described eye-scene alignment. To overcome this issue, we apply a spatial shift and coordinate translation that moves the modified camera image opposite to the direction of eye movement as depicted in Figure 3 and described by Equations 1-4 as follows.

Given a camera image with size c_w x c_h, an eye model with size e_w x e_h, the spatial shift of the mirror image (m_x, m_y) for each eye with spatial offset (o_x, o_y) is given by

$$m_y = (\frac{c_y}{c_h} + o_y) \cdot (c_h - e_h) + \frac{e_h}{2} \tag{1}$$

and

$$m_x = (1 - (\frac{c_x}{c_w} + o_x)) \cdot (c_w - e_w) + \frac{e_w}{2} \tag{2}$$

The location of the pupil and iris on the eye (e_x, e_y) moving in the opposite direction is then given by

$$e_y = c_y - m_y + \frac{e_h}{2} \tag{3}$$

$$e_x = c_x - n_x(c_w - e_w) \tag{4}$$

This dynamic reflection technique has the added benefit of mapping potentially large camera images to relatively small eye models without loss of resolution through the coordinate

© 2026 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Human-Machine Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/THMS.2026.3651818

translations while contributing to an appearance of the eyes as three-dimensional reflectors for which the rotation magnitude is defined by the coordinate translation magnitude.

D. Software

We implemented all modules using the Python programming language [31] with the libraries OpenCV [32] for image processing and MediaPipe [33] for face detection.

III. Methodology - Experiment

To test the hypotheses specified in Section I-B, we designed an experiment that required participants within a group to quickly and accurately identify who in the group was being attended by an eye model to proceed in a fast-paced game-like task. In each experiment, we recorded the accuracy and speed of gaze reference identification. The purpose of the secondary game-like task was to create conditions that would allow a quick and structured recording of many samples while distracting the participants cognitively from the primary stimulus-response task.

A. Participant task

As a secondary game-like task, we modified a traditional Japanese word game known as *Shiritori*. The original game consists of taking turns saying words that start with the last letter of the word uttered by the previous person. For example: If person 1 says “Robot” then person 2 will need to say a word that starts with the letter “T” such as “trouble”, then it is again person 1’s turn to come up with a word that starts with “E” such as “ensures” followed by person 2 with a word starting with “S” and so on.

For this study, we altered the rules of the game in the following way:

1. **Number of players:** The game is played with three people at a time.
2. **Unknown order:** The speaking order is unknown to participants for every word.
3. **Machine selection:** A machine mediator pseudo-randomly decides whose turn it is for each new word and then indicates its choice by looking at that person.
4. **Response button:** Participants who think that they have been selected need to press a button before being allowed to say the next word. They should press the button as soon as they feel that they are being selected and should think about the next word only after the button press.
5. **Time limit:** There is a time limit of 3 seconds for pressing the button. After pressing the button there is a second time limit of 5 seconds for saying the next word. If a person does not respond in time, the mediator will look at the next person.
6. **Target succession rule:** The person who uttered the previous word cannot be immediately re-selected by the machine mediator unless the selected person fails to respond within one of the time limits.

The introduction of a third player and the machine mediator, instead of a turn-taking rule, ensured that participants were always uncertain about whose turn it would be next and had to make a decision based on the mediator’s gaze to proceed with the game. The use of the button press was two-fold: It facilitated identifying and recording whether a participant had understood the gaze reference correctly and allowed us to measure how quickly true positive and false positive reference identifications took place. We considered the absence of a button press within the 3-second time limit as a negative response. The short time limit for the button press urged participants to react as quickly as possible, thereby reducing the potential influence of confounding factors and enabling rapid data collection. We defined each sequence of gazing, button press, and word utterance as one trial. Thus, a single trial took a maximum of 8 seconds.

B. Experiment structure

We utilized three different display conditions of the eye models as depicted in Figure 4. The first condition (Eyes-Only) consisted of pure eye models as described in Section II-A without any reflection-like properties. The second condition (Mirrors-Only) consisted of a pure dynamic mirroring mode that was centered on the current face of interest as described in Section II-C but contained no typical eye features such as pupil and iris. The third condition (Mirror Eyes) consisted of the combination of the first two conditions, i.e., the overlay of dynamically updated mirror-like features on the pupil and iris. The use of these three conditions allowed us to test whether any potential effects of the Mirror Eyes might be attributed to one of its two major constituents or whether these would arise only from their combination.

Fig. 4. Eye model display conditions. 4a: Basic eye model with no reflection of an attended face. 4b: dynamic mirror that tracks an attended face. 4c: Combination of eye model and dynamic mirror of attended face on pupil and iris (Mirror Eyes). The faces in Figures 4b and 4c have been blurred for publication purposes to protect privacy. No face blur was present in the experiment.

C. Procedure

Figure 5 illustrates the procedure and individual components of the experiment. Before starting the experiment, all participants received detailed instructions and practiced for 10 trials in each eye condition.

The subsequent experiment consisted of two parts that differed in how trials of the different eye model display conditions were grouped into blocks. In the first part, each display condition was used for a block of 30 trials, resulting in three blocks. Blocks of this type are further referred to as ”single condition block”. The order of single condition

© 2026 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Human-Machine Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/THMS.2026.3651818

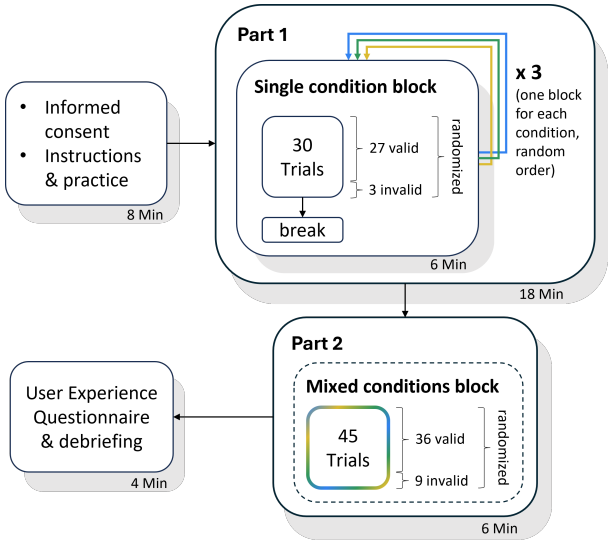


Fig. 5. Experiment procedure with duration estimates for individual steps.

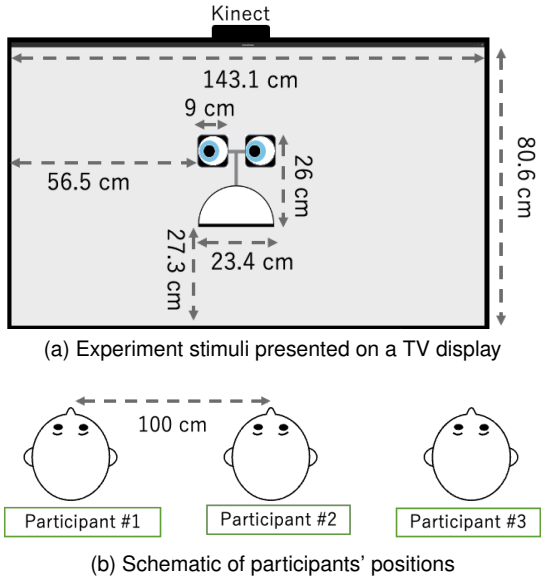


Fig. 6. Schematic of the experimental setup: 6a: Arrangement and dimensions of the visual stimuli consisting of one of the three dynamic eye models rendered into the eyes of a mock-up robot head shown on a gray display. 6b: Distancing between the three participants. The participants were free to move their heads and bodies while in their positions.

blocks within the first part of the experiment was pseudo-randomized between participant triplets. The second part of the experiment consisted of a single "mixed condition block". In a mixed condition block the eye model display conditions switched pseudo-randomly between individual trials, such that the eyes might, for instance, look at one person with Mirror Eyes but focus on the next with non-reflective eyes. At least 15 trials of each condition were distributed across a mixed condition block. While the single condition blocks in the first part allowed participants to anticipate the nature of the cue they would receive from the mediator, the mixing of conditions in the second part added cue uncertainty, making tuning to a specific set of features disadvantageous. Differences between the two parts could thus point towards a possible dependence or independence between identification performance and cue type predictability.

Because participants could *steal* individual trials from each other by wrongly pressing the button when another participant was attended, we created a real-time trial balancing procedure that would either swap or append trials for involved participants in such cases depending on compatibility with the target succession rule (see Section III-A). False responses by participants could therefore extend the total number of trials in a block up to a predefined total time limit of 4 minutes for the first part and 6 minutes for the second part.

Assuming a medium effect size of $\eta^2 = 0.06$, a power analysis for a repeated measures design with three within-subject conditions, 30 participants, and an alpha level of 0.05 estimated the statistical power to be 84.9%.

D. Dependent measures

1) Reference identification performance: To test H1 and H2, we used the reference identification accuracy and reaction times as dependent measures. Both are derived from the participants' button press behavior. The reference identification accuracy for each participant is defined as binary classification accuracy, i.e., as the ratio of the sum of true positive (TP) and true negative (TN) identifications over the sum of all trials involving that participant.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Therefore, the accuracy does not only take the true positive identification of a reference into account but is also affected by the participants' ability to identify when they are not being attended. To provide opportunities for the identification of such non-references, we programmed the eye model to "make mistakes" in a subset of trials by not attending to any particular participant but instead focusing on areas located between participants. For such trials, the lack of a button press within the time limit would then count as a true negative identification. To ensure exposure of each participant to at least one error case for each block and condition while minimizing error predictability, the rate of machine mistakes was set to 10% (3 trials) for each condition block in experiment part 1 and to 20% (3x3 trials) for part 2, which contained all three conditions.

We chose to use only a single response button per participant to avoid any coordination-dependent confounders. This button indicated a positive response. Negative responses were given implicitly by not responding within the 3-second time limit. Therefore, reaction times were only available for true positive and false positive responses as indicated by Hypothesis 2.

E. User experience

In this experiment, the 8-item version of the User Experience Questionnaire (UEQ-S; [34]) was used to assess participant experience with the virtual eyes. After completing all experimental blocks, participants filled out the UEQ-S, which consists of eight items, four representing a pragmatic scale and

© 2026 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Human-Machine Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/THMS.2026.3651818

the other four representing a hedonic scale. Each item is rated on a 7-point semantic differential scale, ranging from -3 (most negative evaluation) to +3 (most positive evaluation). Among psychometrically validated tools for measuring user experience (e.g., [34]–[36]), we selected the UEQ-S because it efficiently captures two core user experience dimensions, the pragmatic and hedonic quality [37], while minimizing participant burden through its brief completion time.

F. Participants

The experiment involved 30 participants (4 females and 26 males; aged 21 to 56 years; mean age: 32.8 ± 8.52 years) with normal or corrected-to-normal vision, who participated in groups of three. Participants had no prior exposure to the experimental setup or eye model. All participants signed consent forms, and the experiment was reviewed and approved by the Bioethics Committee in Honda's R&D (100HM-064H). One participant was excluded from accuracy and reaction time analysis due to a lack of recorded button presses.

G. Setup

Figure 6 shows the schematic of the experimental setup. The experiments were conducted in an office environment in which three participants stood in front of a 65-inch display (Bravia KJ-65X80J; Sony, Japan). Participants stood approximately 2 meters from the display, with a 1-meter distance between one another (Fig. 6b). The positions were selected to provide a comfortable social distance while still keeping participants close enough to each other and far enough from the display to create some gaze direction ambiguity. A camera (Azure Kinect DK; Microsoft, USA) was mounted directly above the center of the screen to capture the scene containing the participants, i.e., the eyes' attention targets. Each participant was equipped with a key button (FS1-P; Edikun, Japan) on their hand, connected to the experimental computer via USB. A computer (Magnus One ECM73070C; Zotac, Hong Kong) managed both the visual stimulus presentation and data collection. The experimental stimulus, an image of a generic robot head equipped with the dynamic eye model of the respective condition, was displayed at the center of the screen on a gray background with a height of 26 cm (Figure 6a). Each eye was represented in a 9 × 9 cm square. The diameter of the sclera of each eye was 8.8 cm, the diameter of the iris (outer circle) was 6.4 cm, and the diameter of the pupil (inner circle) was 3.5 cm. The reflection opacity in the Mirrors-Only and Mirror Eyes conditions was set to 60%. The display measured 80.6 cm in height and 143.1 cm in width and was positioned on a stand that was about 99.5 cm tall. This positioning ensured that the eye models were at a height of approximately 1.5 meters, which allowed participants to observe them comfortably from a standing position.

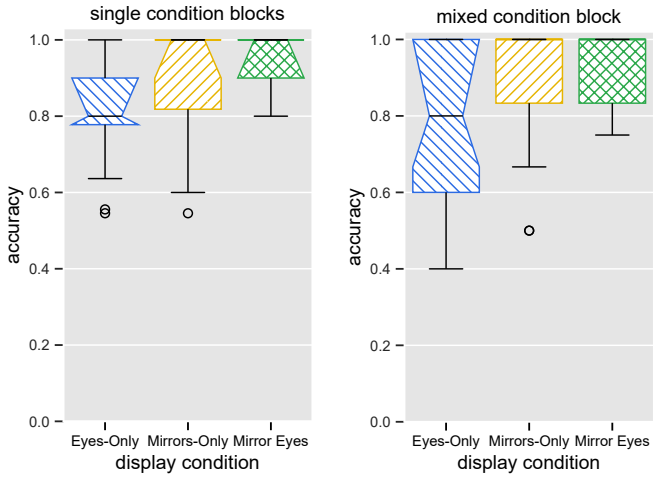
H. Software

We implemented the experiment using the Python programming language [31]. The subsequent data analysis was conducted in Python using the libraries Pandas [38] for data structuring, Matplotlib [39] and seaborn [40] for visualizations, and statsmodels [41], scikit-posthocs [42], and pingouin [43] for statistical analysis.

IV. Results

TABLE I
SUMMARY OF TWO-WAY ANOVA RESULTS FOR ACCURACY VALUES.

	df	sum sq	mean sq	F	$PR(>F)$
Display condition	2	0.5430	0.2715	14.9880	$<$ **0.0001**
Block type	1	0.0009	0.0009	0.0512	0.8212
Interaction	2	0.0032	0.0016	0.0893	0.9146
Residual	168	3.0434	0.0181		



(a) Single condition blocks (part 1) (b) Mixed condition block (part 2)

Fig. 7. Box plots showing the distributions of accuracy scores across conditions for single (left) and mixed condition blocks (right).

A. Reference identification performance

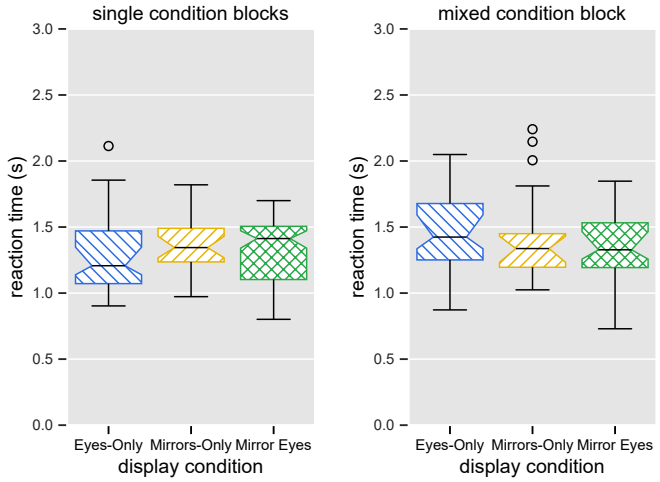
1) Accuracy: Figure 7 shows the distribution of accuracy values across conditions in both parts of the experiment. The accuracy was high across all conditions for the single condition blocks (Eyes-Only: $\mu = 0.82, sd = 0.13$; Mirrors-Only: $\mu = 0.91, sd = 0.12$; Mirror Eyes: $\mu = 0.94, sd = 0.07$). Also in the mixed condition block the accuracy values were high across conditions (Eyes-Only: $\mu = 0.81, sd = 0.20$; Mirrors-Only: $\mu = 0.92, sd = 0.15$; Mirror Eyes: $\mu = 0.94, sd = 0.09$). A two-way ANOVA (see Table I) revealed a significant main effect for display condition ($F(2, 168) = 14.99, p < 0.001$) but no effect for block type ($F(1, 168) = 0.05, p = 0.82$), or interaction between display condition and block type ($F(2, 168) = 0.09, p = 0.91$). Post-hoc comparisons with Holm-Bonferroni adjusted p-values revealed a significant difference between the Eyes-Only and Mirrors-Only ($p < 0.01$) and between the Eyes-Only and the Mirror Eyes condition ($p < 0.001$). These results support our first hypothesis and additionally indicate an accuracy advantage of a dynamic mirroring mode over pure eye-based referencing. There are no indications of an influence of cue type predictability on the average accuracy.

© 2026 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Human-Machine Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/THMS.2026.3651818

TABLE II
SUMMARY OF TWO-WAY ANOVA RESULTS FOR REACTION TIMES.

	df	sum sq	mean sq	F	$PR(>F)$
Display condition	2	0.0534	0.0267	0.3484	0.7063
Block type	1	0.1876	0.1876	2.4489	0.1195
Interaction	2	0.2440	0.1220	1.5921	0.2066
Residual	168	12.8724	0.0766		



(a) Single condition blocks (part 1) (b) Mixed condition block (part 2)

Fig. 8. Box plots showing the distributions of reaction times across conditions for single (left) blocks and the mixed condition block (right).

2) Reaction times: Figure 8 shows the distribution of reaction times across conditions in experiment parts 1 and 2. Average reaction times were similar across all conditions for the single condition blocks of experiment part 1 (Eyes-Only: $\mu = 1.29, sd = 0.29$; Mirrors-Only: $\mu = 1.38, sd = 0.20$; Mirror Eyes: $\mu = 1.34, sd = 0.24$). Reaction times in the mixed condition block of experiment part 2 were comparable to those of the first part, especially for the two mirror conditions (Eyes-Only: $\mu = 1.46, sd = 0.32$; Mirrors-Only: $\mu = 1.40, sd = 0.32$; Mirror Eyes: $\mu = 1.35, sd = 0.27$). A two-way ANOVA (see Table II) did not reveal any significant effects for display condition ($F(2, 168) = 0.35, p = 0.70$), block type ($F(1, 168) = 2.45, p = 0.12$), or interaction between display condition and block type ($F(2, 168) = 1.59, p = 0.20$).

These results are not indicative of any reaction time disadvantage of the Mirrors-Only and Mirror Eyes modes compared to the Eyes-Only mode and hence do not support the second hypothesis. There are no clear indications of an influence of cue type predictability on the average reaction time.

B. User experience

Figure 9 presents the results of the UEQ-S ratings of 30 participants who evaluated the user experience across three different display conditions of the eye models, which were referred to as "Eyes", "Mirrors", and "Mirror Eyes" during the study. Table III shows the scale means for the UEQ-S ratings. Standard deviations are shown in parentheses.

For the Eyes-Only condition, the mean score was 0.67 for pragmatic quality and -0.26 for hedonic quality. These scores

TABLE III
SCORE MEANS AND STANDARD DEVIATIONS OF THE UEQ-S (N = 30)

Display condition	Pragmatic quality	Hedonic quality	Overall
Eyes-Only	0.67 (1.27)	-0.26 (1.34)	0.20 (1.12)
Mirrors-Only	0.71 (1.59)	-0.22 (1.46)	0.25 (1.12)
Mirror Eyes	1.29 (1.38)	**1.30 (0.92)**	**1.30 (0.98)**

Bold values denote statistically significant differences ($p < 0.01$).

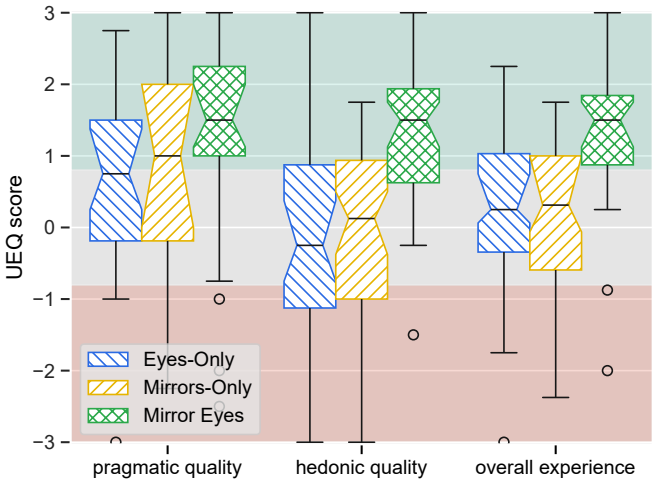


Fig. 9. Box plots for UEQ-S rating of pragmatic and hedonic quality, and overall user experience across the three different conditions of the eye models. Different background colors represent evaluation ranges: values between -0.8 and 0.8 indicate neutral ratings (grey), values above 0.8 signify positive ratings (green), and values below -0.8 indicate negative ratings (red).

are similar to those in the Mirrors-Only condition for which the means were 0.71 for pragmatic quality and -0.22 for hedonic quality. In contrast, the Mirror Eyes condition yielded the highest rating among the three conditions, with mean scores of 1.29 for pragmatic quality and 1.3 for hedonic quality. A one-way ANOVA revealed a significant difference in hedonic quality across the three conditions ($F(2, 87) = 14.86, p < 0.001$). Post-hoc multiple comparisons with Holm-Bonferroni adjusted p-values showed a significant difference between the Mirror Eyes and the other two conditions (both $p < 0.001$). For pragmatic quality, while the Mirror Eyes condition had a higher mean (1.29, indicating a positive rating in the green zone), no significant differences were found between the three conditions ($F(2, 87) = 1.81, p = 0.17$).

For the overall user experience evaluation, the Mirror Eyes condition received a mean score of 1.30, while the Eyes-Only and Mirrors-Only conditions scored 0.20 and 0.25, respectively. ANOVA results indicated a significant difference in overall user experience across the three conditions ($F(2, 87) = 9.93, p < 0.001$). Post-hoc multiple comparisons with Holm-Bonferroni adjusted p-values confirmed that the Mirror Eyes condition differed significantly from the other two conditions (both $p < 0.001$). The internal consistency of the pragmatic quality and hedonic quality scales for the Mirror Eyes condition was sufficiently high, with Cronbach Alpha values of 0.88 (pragmatic quality) and 0.81 (hedonic quality).

In sum, the results of the UEQ-S ratings indicate a more

© 2026 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Human-Machine Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/THMS.2026.3651818

positive user experience for the Mirror Eyes condition than for the other two conditions and thus support our third hypothesis.

V. Discussion

Here, we introduced an approach for overcoming spatial referencing challenges of 2D eye models as communication tools in human-machine interaction. By augmenting a screen-based eye model with a dynamic reflection-like overlay that is contingent on eye movements, pupil-based pointing is complemented by direct scene context to improve spatial reference identification for observers. We conducted a user study with 30 participants who were asked to carry out a group interaction task that relied on an identification of spatial references produced by screen-based eye models to compare this "Mirror Eyes" model to a pure eye model without mirroring capabilities and a pure dynamic mirror model without eye-like features.

A. Reference identification accuracy and speed

Although accuracy was high for all tested eye models, the two eye models with mirror-like features led to significantly higher reference identification accuracy than a pure non-reflective eye model. Despite the added contextual information, the observers of augmented eye models needed to process, the study did not reveal any significant impact of mirroring on reaction times. The lack of noticeable impact on reaction times stands in contrast to our second hypothesis, which assumed a speed reduction for reference identification as an expression of a potential trade-off for the additional information that needs to be processed. The absence of a mirroring-related speed trade-off might in part be explained by the practice trials prior to the experiment. These could have given participants an opportunity to identify and focus on the most informative features. Future work should facilitate the separation of dependent measures from early and later trials of each experimental condition to identify potential learning effects.

Another possible explanation for the lack of an apparent trade-off lies in the reference target class used in our study: The eye models always focused on faces. Human brains have evolved highly dedicated face processing abilities [44], likely due to their social role, allowing us to recognize target faces after as little as 260 ms [45]. Furthermore, the seen gaze direction has been found to modulate the activity of regions associated with face processing [46], which could suggest that a linkage between the two, as introduced here, might be exploiting fast visual processing networks. In addition, occasionally seeing their own faces might have facilitated the processing of other participants' faces [47], which could have reduced the risk of false positive identifications.

However, the potential links to neural face processing and associated findings also highlight one limitation of the present study: Because faces were the only spatial reference targets, we gain no insights about the utility of the approach for conveying references to other object categories that are likely to have fewer dedicated neural processing resources. Future investigations should therefore expand the set of gaze target object categories.

B. User experience

Our findings regarding accuracy and response speed are consistent with user experience ratings obtained via the UEQ-S questionnaire. The UEQ-S results indicate that the Mirror Eyes condition was attributed significantly higher hedonic qualities than the other conditions, supplying the means for the step from a neutral to a positive user experience. Overall, the statistical analysis revealed that the user experience with the Mirror Eyes condition was significantly better than with the Eyes-Only or Mirrors-Only conditions. The participants hence appear to have largely appreciated the support given by the Mirror Eyes condition.

Notably, a single participant differed in this regard by rating both mirror conditions lower than the pure eye model. This participant explained afterwards, that seeing oneself in the eyes evoked a strong sense of being observed that made them feel uncomfortable. Although this response came from a 1/30 minority of the participants, the potential for such responses may require additional consideration in future interaction designs involving similar features. Furthermore, the male-skewed participant sample may have constrained the range of experiences and behaviors represented in our findings.

C. Mental state transparency

Differential emotional responses to seeing oneself in the mirror have been described before [48]–[50] and could account for some variability in user experience. Another factor that may contribute to an adverse sense of being observed is an attribution of judgment or malevolence to the observing entity. Because we were mainly interested in the Mirror Eyes' utility for improving spatial ocular referencing, we kept the expression of the eye model neutral. However, we cannot rule out the possibility that individual users may attribute diverse mental states or personalities [51] to the agent driving an eye model. Close associations between eyes and mental states may impair some interaction scenarios, but also highlight further potential in the utilization of augmented eye models such as the Mirror Eyes.

Often described as windows to the mind, the eyes can reflect a larger set of mental states than just spatial attention. For instance, carefully designed gaze patterns, especially when integrated with contextual information in avatars or virtual agents, have been used in VR to subtly convey cognitive states and enhance social presence [52], [53]. Moreover, when combined with non-verbal cues such as gestures and facial expressions, gaze inference models show great promise in human-computer interaction applications [54], [55], and can also be used to express personality traits [54] or emotional states like fear and anxiety [56]. Nevertheless, close associations between eyes and mental states may also limit possible applications.

D. Cue saliency and attention

Eyes by themselves are strong attention attractors [3]–[5]. Reflection-like overlays may increase the salience of eye models even further, especially for reflections of other strong

© 2026 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Human-Machine Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/THMS.2026.3651818

attractors like faces [57], [58]. The salience of any machine expression deserves consideration in interaction design because it affects the balance between observability and distraction. Typically, a high expression salience makes an observer more likely to look at a face and less likely to look at elements in the environment. But because the virtual reflections are references to elements in the machine's environment, their effects on the distribution of an observer's attention could be twofold: They could draw attention to the eyes themselves, and they could draw attention to the reflected element in the environment. Mirror Eyes could therefore serve as a tool to establish joint attention between machine agents and human observers. On top of their original role as a general purpose tool for spatial references, an ability to establish and modulate joint attention could extend means for machine-mediated joint attention skill development (e.g., [59]) in people who experience challenges in reading social signals, as seen in individuals on the autism spectrum [60], [61]. Future work should therefore investigate the impact of the Mirror Eyes on people's attention and whether mirroring can be used for a selective modulation of eye model salience and gaze target salience.

E. References in physical space

Looking back at the original objective of the present investigation, i.e., the improvement of spatial reference communication with 2D eye models, one aspect that requires closer examination is depth of gaze. Although our eye model encodes focus distance in the interpupillary distance and reflections naturally become smaller with increasing distance, the present study did not investigate the effects of Mirror Eyes at different levels of focus depth.

Another direction for follow-up research can arise from an application of the approach within more physically instantiated machine interaction partners, such as robots that possess a movable head.

Previous work has shown that humans perceive the direct gaze of a robot more clearly when the robot's eye gaze is coordinated with its head movement, as opposed to when the head remains stationary [2], [26]. In the present study, all three eye conditions dynamically tracked the user's face, and the results demonstrate that participants could accurately judge the machine mediator's gaze direction, despite its static screen-based mock-up head. Previous research also suggests that combining eye and head movements while simulating visual perception processing enhances the perception of a robot's attentional gaze [2]. Also, potential cross-modal interactions with other output modalities such as speech (e.g., [62]) should be considered in future developments.

Moving forward, we aim to incorporate the Mirror Eye functionality into a physical robot to further examine the expressive potential that this approach may afford in coordination with physical action.

F. Conclusion

The introduction of a reflection-like overlay on the display of virtual eye models (Mirror Eyes) significantly improved reference identification accuracy compared to a pure, non-reflective eye model without affecting identification speed. In congruence with this accuracy gain, the subjective user experience in a game-like interaction with the eye models significantly improved with the Mirror Eyes compared to conditions that either lacked the reflection capability or the eye-like appearance. These results suggest that the benefits associated with Mirror Eyes cannot be attributed to the underlying eye model or the reflection alone, but only to their combination. Subsequent research should investigate the transfer of such effects to the reflections of other classes of gaze targets, explore options for a selective modulation of mirroring for attention guidance, and ascertain meaningful coordination with degrees of freedom available to more physically integrated machine agents.

ACKNOWLEDGMENTS

We thank Christian Arzate, Yotam Sechayk, and Heiko Wersing for their thoughtful discussions, which were instrumental in shaping the user study design.

REFERENCES

[1] H. Kobayashi and S. Kohshima, "Unique morphology of the human eye," *Nature*, vol. 387, no. 6635, pp. 767–768, 1997.

[2] Y. Fang, J. M. Pérez-Molerón, L. Merino, S.-L. Yeh, S. Nishina, and R. Gomez, "Enhancing social robot's direct gaze expression through vestibulo-ocular movements," *Advanced Robotics*, vol. 38, no. 19-20, pp. 1457–1469, 2024.

[3] S. R. Langton and V. Bruce, "Reflexive visual orienting in response to the social attention of others," *Visual cognition*, pp. 541–567, 1999.

[4] S. R. Langton, R. J. Watt, and V. Bruce, "Do the eyes have it? cues to the direction of social attention," *Trends in cognitive sciences*, vol. 4, no. 2, pp. 50–59, 2000.

[5] R. J. Itier, C. Villate, and J. D. Ryan, "Eyes always attract attention but gaze orienting is task-dependent: Evidence from eye movement monitoring," *Neuropsychologia*, vol. 45, no. 5, pp. 1019–1028, 2007.

[6] J. Driver IV, G. Davis, P. Ricciardelli, P. Kidd, E. Maxwell, and S. Baron-Cohen, "Gaze perception triggers reflexive visuospatial orienting," *Visual cognition*, vol. 6, no. 5, pp. 509–540, 1999.

[7] M. Tomasello, B. Hare, H. Lehmann, and J. Call, "Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis," *Journal of human evolution*, vol. 52, no. 3, pp. 314–320, 2007.

[8] S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, and I. Plumb, "The "reading the mind in the eyes" test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism," *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 42, no. 2, pp. 241–251, 2001.

[9] R. Adolphs, F. Gosselin, T. W. Buchanan, D. Tranel, P. Schyns, and A. R. Damasio, "A mechanism for impaired fear recognition after amygdala damage," *Nature*, vol. 433, no. 7021, pp. 68–72, 2005.

[10] J. K. Hietanen, J. M. Leppänen, M. J. Peltola, K. Linna-Aho, and H. J. Ruuhiala, "Seeing direct and averted gaze activates the approach–avoidance motivational brain systems," *Neuropsychologia*, vol. 46, no. 9, pp. 2423–2430, 2008.

[11] A. Senju and M. H. Johnson, "The eye contact effect: mechanisms and development," *Trends in cognitive sciences*, vol. 13, no. 3, 2009.

[12] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori, "The icub humanoid robot: An open platform for research in embodied cognition," in *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, ser. PerMIS '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 50–56.

[13] C. Breazeal, M. Siegel, M. Berlin, J. Gray, R. Grupen, P. Deegan, J. Weber, K. Narendran, and J. McBean, "Mobile, dexterous, social robots for mobile manipulation and human-robot interaction," in *ACM SIGGRAPH 2008 new tech demos*. ACM, 2008, pp. 1–1.

[14] A. Zaraki, D. Mazzei, M. Giuliani, and D. De Rossi, "Designing and evaluating a social gaze-control system for a humanoid robot," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 2, 2014.

© 2026 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Human-Machine Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/THMS.2026.3651818

[15] H. Admoni and B. Scassellati, "Social eye gaze in human-robot interaction: a review," *Journal of Human-Robot Interaction*, 2017.

[16] R. Gomez, D. Szapiro, K. Galindo, and K. Nakamura, "Haru: Hardware design of an experimental tabletop robot assistant," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 233–240.

[17] N. Yoshida, S. Yonemura, M. Emoto, K. Kawai, N. Numaguchi, H. Nakazato, S. Otsubo, M. Takada, and K. Hayashi, "Production of character animation in a home robot: A case study of lovot," *International Journal of Social Robotics*, vol. 14, no. 1, pp. 39–54, 2022.

[18] H. Kiilavuori, V. Sariola, M. J. Peltola, and J. K. Hietanen, "Making eye contact with a robot: Psychophysiological responses to eye contact with a human and with a humanoid robot," *Biological Psychology*, vol. 158, p. 107989, 2021.

[19] S. Linnunsalo, D. Küster, S. Yrttiaho, M. J. Peltola, and J. K. Hietanen, "Psychophysiological responses to eye contact with a humanoid robot: Impact of perceived intentionality," *Neuropsychologia*, vol. 189, p. 108668, 2023.

[20] M. Krüger, C. B. Wiebel, and H. Wersing, "From tools towards cooperative assistants," in *Proceedings of the 5th International Conference on Human Agent Interaction*, 2017, pp. 287–294.

[21] B. Sendhoff and H. Wersing, "Cooperative intelligence-a humane perspective," in *2020 IEEE international conference on human-machine systems (ICHMS)*. IEEE, 2020, pp. 1–6.

[22] A. Bhaskara, M. Skinner, and S. Loft, "Agent transparency: A review of current theory and evidence," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 3, pp. 215–224, 2020.

[23] M. Teyssier, M. Koelle, P. Strohmeier, B. Fruchard, and J. Steimle, "Eyecam: Revealing relations between humans and sensing devices through an anthropomorphic webcam," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–13.

[24] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley [from the field]," *IEEE Robotics and Automation Magazine*, vol. 19, no. 2, 2012.

[25] Y. Fang, L. Merino, S. Thill, and R. Gomez, "Designing visual and auditory attention-driven movements of a tabletop robot," in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2023, pp. 2232–2237.

[26] Y. Fang, J. M. Pérez-Molerón, L. Merino, and R. Gomez, "Enhancing human perception of direct gaze from a social robot through eye-head coordination," in *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, 2024.

[27] I. P. Howard and B. J. Rogers, *Binocular vision and stereopsis*. Oxford University Press, USA, 1995.

[28] J. C. Lee, "Hacking the nintendo wii remote," *IEEE Pervasive Computing*, vol. 7, no. 3, pp. 39–45, 2008.

[29] T. Bates, J. Kober, and M. Gienger, "Head-tracked off-axis perspective projection improves gaze readability of 3d virtual avatars," in *SIGGRAPH Asia 2018*. Association for Computing Machinery (ACM), 2018, p. 29.

[30] F. Delaunay, J. De Greeff, and T. Belpaeme, "Towards retro-projected robot faces: an alternative to mechatronic and android faces," in *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*. Ieee, 2009, pp. 306–311.

[31] G. Van Rossum and F. L. Drake Jr, *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[32] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[33] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "Mediapipe: A framework for perceiving and processing reality," in *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR)*, 2019.

[34] M. Schrepp, J. Thomaschewski, and A. Hinderks, "Design and evaluation of a short version of the user experience questionnaire (ueq-s)," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 6, pp. 103–108, 12/2017 2017.

[35] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[36] K. Finstad, "The usability metric for user experience," *Interacting with Computers*, vol. 22, no. 5, pp. 323–327, 05 2010.

[37] M. Hassenzahl, S. Diefenbach, and A. Göritz, "Needs, affect, and interactive products–facets of user experience," *Interacting with computers*, vol. 22, no. 5, pp. 353–362, 2010.

[38] T. pandas development team, "pandas-dev/pandas: Pandas," Feb. 2020.

[39] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[40] M. L. Waskom, "seaborn: statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021.

[41] S. Seabold and J. Perktold, "statsmodels: Econometric and statistical modeling with python," in *9th Python in Science Conference*, 2010.

[42] M. A. Terpilowski, "scikit-posthocs: Pairwise multiple comparison tests in python," *Journal of Open Source Software*, vol. 4, no. 36, 2019.

[43] R. Vallat, "Pingouin: statistics in python," *Journal of Open Source Software*, vol. 3, no. 31, p. 1026, Nov. 2018.

[44] G. McCarthy, A. Puce, J. C. Gore, and T. Allison, "Face-specific processing in the human fusiform gyrus," *Journal of cognitive neuroscience*, vol. 9, no. 5, pp. 605–610, 1997.

[45] G. Besson, G. Barragan-Jason, S. J. Thorpe, M. Fabre-Thorpe, S. Puma, M. Ceccaldi, and E. J. Barbeau, "From face processing to face recognition: Comparing three different processing levels," *Cognition*, vol. 158, pp. 33–43, 2017.

[46] N. George, J. Driver, and R. J. Dolan, "Seen gaze-direction modulates fusiform activity and its coupling with other brain areas during face processing," *Neuroimage*, vol. 13, no. 6, pp. 1102–1112, 2001.

[47] Y. H. Li and N. Tottenham, "Seeing yourself helps you see others." *Emotion*, vol. 11, no. 5, p. 1235, 2011.

[48] S. Mulkens and A. Jansen, "Mirror gazing increases attractiveness in satisfied, but not in dissatisfied women: A model for body dysmorphic disorder?" *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 40, no. 2, pp. 211–218, 2009.

[49] E. M. Barnier and J. Collison, "Experimental induction of self-focused attention via mirror gazing: Effects on body image, appraisals, body-focused shame, and self-esteem," *Body Image*, vol. 30, 2019.

[50] A. Ypsilanti, A. Robson, L. Lazuras, P. A. Powell, and P. G. Overton, "Self-disgust, loneliness and mental health outcomes in older adults: an eye-tracking study," *Journal of affective disorders*, vol. 266, 2020.

[51] L. Luo, N. Koyama, K. Ogawa, and H. Ishiguro, "Robotic eyes that express personality," *Advanced Robotics*, vol. 33, no. 7-8, 2019.

[52] Y. Tao, C. Y. Wang, A. D. Wilson, E. Ofek, M. González-Franco, and A. D. Wilson, "Embodying physics-aware avatars in virtual reality," in *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, April 2023.

[53] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell, "A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception," *Computer Graphics Forum*, vol. 34, no. 6, pp. 299–326, 2015.

[54] J. Yamashita, Y. Takimoto, H. Oishi, and T. Kumada, "How do personality traits modulate real-world gaze behavior? generated gaze data shows situation-dependent modulations," *Frontiers in Psychology*, vol. 14, p. 1144048, 2024.

[55] R. Gomez, Y. Fang, S. Thill, R. Ragel, H. Brock, K. Nakamura, Y. Vasylkiv, E. Nichols, and L. Merino, "Developing a robot's empathetic reactive response inspired by a bottom-up attention model," in *International Conference on Social Robotics*. Springer, 2021.

[56] M. Oliveira, C. Fernandes, F. Barbosa, and F. Ferreira-Santos, "Differential correlates of fear and anxiety in salience perception: A behavioral and erp study with adolescents," *Cognitive, Affective, & Behavioral Neuroscience*, pp. 1–13, 2024.

[57] S. R. Langton, A. S. Law, A. M. Burton, and S. R. Schweinberger, "Attention capture by faces," *Cognition*, vol. 107, no. 1, 2008.

[58] T. Gliga, M. Elsabbagh, A. Andravizou, and M. Johnson, "Faces attract infants' attention in complex displays," *Infancy*, vol. 14, no. 5, 2009.

[59] Z. Zheng, L. Zhang, E. Bekele, A. Swanson, J. A. Crittendon, Z. Warren, and N. Sarkar, "Impact of robot-mediated interaction system on joint attention skills for children with autism," in *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, 2013, pp. 1–8.

[60] G. Dawson, K. Toth, R. Abbott, J. Osterling, J. Munson, A. Estes, and J. Liaw, "Early social attention impairments in autism: social orienting, joint attention, and attention to distress." *Developmental psychology*, vol. 40, no. 2, p. 271, 2004.

[61] T. Charman, "Why is joint attention a pivotal skill in autism?" *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 358, no. 1430, pp. 315–324, 2003.

[62] V. Srinivasan, C. L. Bethel, and R. R. Murphy, "Evaluation of head gaze loosely synchronized with real-time synthetic speech for social robots," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 6, pp. 767–778, 2014.

© 2026 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.