

Gaze in the Machine: How Mechanical and AI-Based Eye Contact Correction Shape Communication Experience and Gaze Behaviour in Video Calls

Leonie Wichers

Technische Universität Berlin

Leon Schreiber

Technische Universität Berlin

Sebastian Möller

Technische Universität Berlin

Robert P. Spang

`robert.spang@uni-weimar.de`

Bauhaus-Universität Weimar

Research Article

Keywords: eye contact, gaze correction, video-mediated communication, eye tracking, social presence, quality of experience

Posted Date: October 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11222460/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Gaze in the Machine: How Mechanical and AI-Based Eye Contact Correction Shape Communication Experience and Gaze Behaviour in Video Calls

Leonie Wichers¹, Leon Schreiber¹, Sebastian Möller^{1,2},
Robert P. Spang^{3,1*}

¹Quality and Usability Lab, Technische Universität Berlin, Berlin, Germany.

²German Research Center for Artificial Intelligence (DFKI), Berlin, Germany.

^{3*}Chair of Usability, Faculty of Media, Bauhaus-Universität Weimar, Weimar, Germany.

*Corresponding author(s). E-mail(s): robert.spang@uni-weimar.de;

Contributing authors: l.wichers@campus.tu-berlin.de;

leon.schreiber@tu-berlin.de; sebastian.moeller@tu-berlin.de;

Abstract

Video calls prevent mutual eye contact because the camera and the partner’s image on screen are spatially separated. Teleprompters and AI-based gaze correction promise to restore it, but their effect on users is largely unknown. In a within-subjects study ($N = 32$), participants played a conversational guessing game with a partner in four conditions: face-to-face, a standard video call, a video call in which the partner used a teleprompter, and a video call with AI-based gaze correction of the partner’s video. We measured social presence, gaze-specific experience and camera awareness by questionnaire, and gaze behaviour by mobile eye tracking. All video conditions showed a large and robust loss of co-presence compared with face-to-face conversation (Cohen’s $d_z \approx -1.5$ to -1.8), together with heightened camera awareness. Neither intervention narrowed this gap; equivalence tests confirmed comparable co-presence and camera awareness across video conditions. AI-based correction made eye contact feel less natural than the standard video call ($d_z = -0.54$), yet participants’ gaze behaviour remained equivalent. Gaze differences in the teleprompter condition

coincided with a smaller on-screen face and are attributable to framing. Restoring the appearance of eye contact thus does not restore what video calls lack. We discuss implications for evaluating and designing gaze technologies.

Keywords: eye contact, gaze correction, video-mediated communication, eye tracking, social presence, quality of experience

Word count: 5,430 words (main text, excluding title page, abstract, references, figures and tables). **Figures/tables:** 4.

1 Introduction

Eye contact is one of the most powerful signals in face-to-face conversation. It shows attention, regulates turn-taking and builds a sense of connection [1, 2]. Video calls, which have become a routine channel for work, education and private life, structurally remove this signal. Because the camera sits above or beside the screen, a person who looks at their partner’s face appears to look past them, and mutual eye contact becomes practically impossible [3].

Two families of technical remedies aim to restore it. Mechanical solutions, such as teleprompter-like setups, place the camera optically behind the partner’s image, so that looking at the partner and looking into the camera coincide [4]. Software-based solutions leave the hardware unchanged and instead redirect the eyes in the video stream with neural networks [5]. The latter have moved quickly from research prototypes into consumer products and can be switched on with a single click [6]. Their technical evaluation, however, typically relies on image-based metrics. Whether artificial eye contact actually improves the experience of a conversation, and whether it changes how people look at their partner, has received little systematic attention.

In a previous study, presented at the International Conference on Quality of Multimedia Experience (QoMEX 2025) [7], we compared a standard video call, a teleprompter setup and an AI-based gaze correction with face-to-face conversation in a within-subjects experiment. Contrary to the intuition behind these technologies, AI-based correction made eye contact feel *less* natural than in a standard video call, and participants reported more eye contact anxiety. Neither intervention brought the video call measurably closer to face-to-face interaction.

That study left three questions open, which we address here. First, the conference version reported uncorrected contrasts for a large set of measures, and its data had not yet been fully recovered from a logging error. The robustness of its findings was therefore unclear. Second, restoring eye contact targets the partner’s gaze, but video calls also make users aware of the camera and of being observed. Whether gaze interventions alleviate or aggravate this self-focus has not been examined. Third, subjective ratings do not show whether artificial eye contact changes where people actually look. Gaze correction might, for instance, draw attention to the partner’s eyes without being consciously noticed. We therefore ask:

RQ1 How do a teleprompter and AI-based gaze correction affect the communication experience, including social presence, gaze-specific experience and camera awareness, compared with a standard video call and with face-to-face conversation?

RQ2 Do these interventions change participants’ gaze behaviour towards their partner’s face?

To answer these questions, we re-analyse the full questionnaire data set with correction for multiple comparisons and several sensitivity analyses, and we add eye-tracking data recorded during the conversations. We find a large and robust presence gap between all video conditions and face-to-face conversation that neither intervention narrows. AI-based correction reduces the naturalness of eye contact while leaving gaze behaviour unchanged, and the teleprompter’s apparent effects on gaze behaviour reflect a difference in framing.

Extension of the conference version.

This article extends our QoMEX 2025 short paper [7] in four ways:

1. an eye-tracking analysis of participants’ gaze behaviour towards their partner’s face (RQ2);
2. a previously unreported set of measures on camera awareness and the feeling of being observed, as well as interpersonal closeness;
3. a revised statistical analysis of the complete data set, with correction for multiple comparisons, effect sizes with confidence intervals and sensitivity analyses. This analysis supersedes the results of the conference version;
4. an extended review of related work and a full description of the method.

2 Related Work

2.1 Functions of eye contact in face-to-face interaction

Gaze has been studied as a regulator of social interaction since Kendon’s seminal observations of conversational gaze [1]. Eye contact is commonly understood as the perception of another person’s direct gaze, which triggers specific social-cognitive processing in the observer [8]. It serves several functions: it conveys information about the partner’s focus of attention, regulates the flow of conversation, and expresses intimacy or social control [2]. In dyadic conversation, for example, speakers tend to avert their gaze at the start of a turn and to re-establish eye contact when yielding it [9], and sustained mutual gaze can signal both affiliation and dominance depending on the social context [10].

These functions are not tied to a fixed gaze pattern. Gaze to faces is dynamically allocated: viewers fixate the mouth more when the partner speaks and the nose region during rapid head movements [11], and the distribution of gaze across facial features depends on task and interaction demands [12]. Whether a gaze is experienced as directed at oneself is also not a matter of exact alignment. The *cone of gaze* describes a tolerance range of eye and head orientations within which observers still judge that they are being looked at [13]. When visual access to the partner is removed altogether, task performance can remain intact, but dialogue becomes less efficient, with more

interruptions and more verbal back-channelling to compensate for the missing non-verbal cues [14].

2.2 Eye contact in video-mediated communication

In conventional video calls, the camera is mounted above or beside the display. A user who looks at the partner’s face on screen therefore appears to look slightly downwards or sideways, and mutual eye contact becomes practically impossible [3]. In terms of the cone of gaze, the camera–screen offset of typical desktop setups often places the perceived gaze direction outside the range in which it is read as direct [13].

Beyond gaze geometry, mediated faces are processed differently from co-present ones. Live in-person eye contact elicits longer gaze durations, larger pupil diameters and stronger activation of social and attentional brain networks than the same partner seen over a video link [15]. Similarly, motor resonance is stronger in live than in video-based interaction [16]. Evidence is not uniform, however. Autonomic arousal and facial-muscle responses to eye contact did not differ between a live interaction and a video call in one study [17]. At the level of social judgement, people predict strangers’ cooperativeness less accurately after video calls than after face-to-face encounters [18]. Together, these findings suggest that video communication reproduces basic responses to a partner’s gaze but not the full depth of social processing found in co-present interaction. What remains open is whether, and how much, this gap is attributable to the missing eye contact itself.

2.3 Approaches to restoring eye contact

Attempts to restore eye contact in video communication fall into three groups: behavioural, mechanical, and software-based.

Behavioural strategies.

Simply looking into the webcam is not an effective remedy. In simulated video job interviews, directing one’s gaze at the camera yielded no advantage over looking at the screen [19]. Psychophysical work further shows that gaze directed exactly at the camera is not the gaze most often judged as eye contact. Observers most frequently perceived direct gaze when the looker fixated a point roughly 2° below the camera [20].

Mechanical approaches.

Mechanical solutions physically reduce the offset between camera and displayed eyes. Adjusting camera position, seating height and viewing distance so that the gaze angle stays within about 2° can create the impression of eye contact. Wider framing, however, increases perceived psychological distance compared with head-and-shoulder views [21]. In simulated teletherapy, facilitating eye contact in this way did not increase clients’ perceived empathy or telepresence [22]. Teleprompter-like setups place the camera behind a semi-transparent mirror that displays the partner’s image, so users can look at the partner and into the camera at the same time. Qualitative work with

such a setup reported increased connectedness and smoother turn-taking. Participants, however, also experienced occasional discomfort with an eye contact they were no longer used to in mediated conversations [4].

Software-based approaches.

Software-based methods leave the hardware unchanged and modify the transmitted image instead. Early work replaced or geometrically warped the eye region to simulate direct gaze [23, 24]. Deep neural networks now redirect gaze in real time by resynthesising the existing eye pixels [5]. Such models have reached consumer products, for example NVIDIA’s Maxine Eye Contact model in the NVIDIA Broadcast application [6, 25]. Technical evaluations of these systems typically rely on image-based error metrics rather than on how users experience the corrected gaze [5]. A related line of work conveys gaze information without photorealistic correction. Examples include gaze-aware 3D portraits that turn towards the attended participant [26] and explicit overlays that indicate who is looking at whom [27]. Such systems increased engagement and coordination but were partly experienced as unnatural [26]. A general concern for any synthetic modification of socially salient facial features is that near-realistic but imperfect renderings can evoke eeriness rather than affinity, a phenomenon known as the uncanny valley [28].

2.4 Research gap

Three gaps motivate the present work. First, mechanical and software-based approaches have mostly been evaluated in isolation or against a standard video call only. It is therefore unclear how close either comes to face-to-face interaction. Second, the two approaches differ fundamentally in how they intervene. A teleprompter changes the physical setup but transmits unaltered video, whereas AI-based correction keeps the setup but alters the video signal. Whether this difference matters for the communication experience has not been tested in a common design. Third, most evaluations rely on subjective ratings of connectedness or on task performance. They rarely capture how artificial eye contact changes where users actually look, or how aware users become of the camera and of being observed. Our QoMEX study addressed the first two gaps for a core set of experience measures [7]. The present article extends it with eye-tracking data and camera-related constructs to address the third.

3 Method

3.1 Participants

Thirty-three participants were recruited via the university’s participant pool. One data set was excluded due to substantial missing data, leaving $N = 32$ (age 20–62 years, $M = 31.8$, $SD = 9.5$; 17 male, 13 female, 1 intersex/diverse, 1 undisclosed). Sixteen participants were students, 14 were employed and 2 were not employed. Twenty-five reported German as their native language; the remaining participants’ native languages were Urdu, Italian, Russian, Cantonese and Korean. Sessions were conducted in German or English. Participants reported using video calls for $M = 3.3$ hours per

week ($SD = 3.1$). Because eyeglasses interfere with the mobile eye tracker, only participants who did not need to wear glasses were admitted. The study was approved by the ethics committee of Faculty IV of Technische Universität Berlin (reference number 331). Participants gave written informed consent and received 15 € for the approximately one-hour session.

3.2 Design and conditions

We used a within-subjects design with four communication conditions of five minutes each (see Fig. 1 for the laboratory layout). In all conditions, the participant interacted with a quiz master, a member of the research team who was not introduced as such (see Sect. 3.4).

- **Face-to-face (F2F):** Participant and quiz master sat opposite each other at a table in the same room.
- **Baseline video call (BaseVC):** Both parties used a standard desktop video call with the webcam mounted centrally on top of the monitor, so that mutual eye contact was not possible.
- **Teleprompter (Tele):** The quiz master used a teleprompter. The participant’s video was mirrored onto a semi-transparent glass in front of a second webcam, which allowed the quiz master to look at the participant’s image and into the camera simultaneously. To keep the viewing distance constant, the quiz master stood on a floor marker. The quiz master was therefore farther from the camera than in the other video conditions, and appeared smaller and with wider framing on the participant’s screen.
- **AI-based gaze correction (AI):** The setup was identical to BaseVC, but the quiz master’s video was processed in real time by the eye-contact feature of NVIDIA Broadcast (v2.0). This feature uses the Maxine Eye Contact model, an encoder-decoder convolutional network, to redirect the eyes towards the camera [6, 25].

The gaze interventions were applied to the quiz master only, so that the participant, as the evaluating party, perceived the effect of the interventions. Participants’ self-view was disabled in all video conditions to keep their attention on the partner. The order of the three video conditions was randomised. Because F2F required a room change, it was scheduled either first or last, alternating after approximately every seven to eight participants. As a result, F2F came first for 14 and last for 18 participants. The full order distribution is given in Online Resource 2.

3.3 Apparatus

The study took place in two adjacent laboratory rooms. Each room had a desk with a 27-inch monitor (Lenovo ThinkVision T27p-30), a Logitech Brio webcam mounted centrally on top of the monitor, and a headset with a fixed volume level. The experimenter room additionally contained the teleprompter, whose screen mirrored the monitor and which housed a second Logitech Brio behind the glass. The participant room contained a second desk for the F2F condition, facing a wall with fiducial markers for the eye tracker.

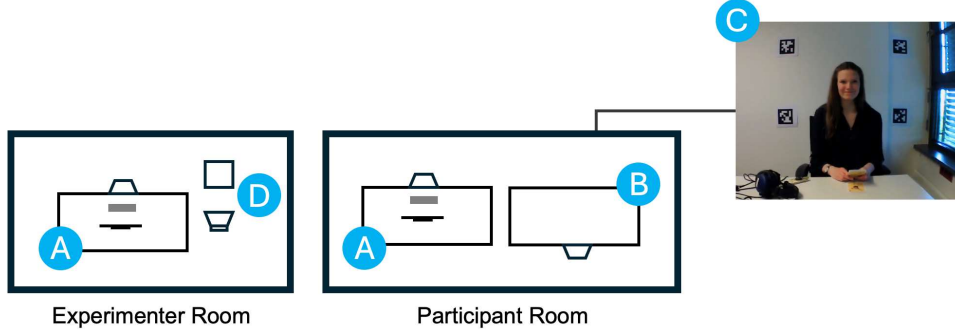


Fig. 1 Floor plan of the experimenter and participant rooms. A: video-call workstation (monitor with centrally mounted webcam, headset); B: table for the F2F condition; C: participant’s view in F2F, with the fiducial markers for the eye tracker on the wall; D: teleprompter with floor marker

All video conditions ran on a custom web-based experiment platform developed at our lab. A central server orchestrated the experiment: it randomised the conditions, controlled the sequence of instruction, call and questionnaire for each condition, and ended calls after five minutes. Two browser clients (one per party) displayed instructions, the video call and the questionnaires in synchrony. Audio and video were exchanged peer-to-peer via WebRTC, and all calls were recorded. The video window on the participant’s side had fiducial markers in its corners, which allowed gaze to be mapped onto the displayed video (Sect. 3.6).

3.4 Task and procedure

In each condition, participant and quiz master played a round of *Black Stories*, a cooperative guessing game. The quiz master read out a short, puzzling story, and the participant reconstructed the underlying scenario by asking yes/no questions. The quiz master knew the solution and guided the participant towards it. We chose the game because it sustains continuous, turn-based conversation without requiring personal disclosure. A new story was used in each condition.

Two experimenters ran each session and alternated between the roles of lead experimenter and quiz master across sessions. To reduce demand characteristics [29], the quiz master was not introduced as part of the research team. The lead experimenter welcomed participants, obtained consent and fitted the eye tracker, then left the room for the rest of the session. Participants first completed a pre-questionnaire. They then went through the four conditions, each consisting of task instructions, the five-minute interaction and the condition questionnaire. Finally, they completed a post-questionnaire.

3.5 Measures

Communication experience.

After each condition, participants rated their communication experience. Unless stated otherwise, items used continuous sliders from 0 (*does not apply*) to 100 (*fully applies*). For reporting, we grouped the measures into four construct groups. Full item wordings and scoring are given in Online Resource 1.

1. *Social presence and engagement:*

- the full Networked Minds Social Presence Inventory [30], with the subscales co-presence, engagement, emotional contagion, comprehension and interdependence, each also split into a self and an other perspective;
- the focused attention and perceived usability subscales of the User Engagement Scale (UES) [31];
- felt closeness, measured with the Inclusion of Other in the Self scale (IOS) [32] (seven pictorial options);
- a single item on perceived social closeness.

2. *Gaze-specific experience:* naturalness of eye contact (two items), perceived intensity of the eye contact, perceived intrusiveness of the eye contact, direct eye contact, eye contact anxiety, own attention, and partner’s attentiveness. Intensity and intrusiveness had been intended as one scale, but the two items were barely correlated ($r = .10$), so we analyse them separately.

3. *Camera awareness and being observed:* single items on webcam awareness, feeling observed, discomfort due to recording, discomfort when looking into the camera, intentional camera gaze (deliberately looking into the camera to return the partner’s gaze), and camera gaze avoidance.

4. *Affect and person perception:* valence and arousal on the Affect Grid [33], and the perceived friendliness and competence of the partner. The Affect Grid stores the click position relative to the top-left corner of the grid; we scored valence as the horizontal position and arousal as one minus the vertical position, so that higher values indicate higher arousal.

The construct groups 2 and 3 were developed for this study; group 3 was not reported in the conference version.

Person characteristics.

Before the first condition, participants reported their weekly video call usage. After the last condition, they provided demographic information. We also assessed sympathy towards the conversation partner, affinity for technology interaction, shyness and sociability. These scales are documented in Online Resource 1 but are not analysed in this article.

3.6 Eye tracking

Participants wore Pupil Invisible glasses (Pupil Labs) [34] throughout the session. Fixations were detected with the Pupil Cloud pipeline and mapped onto the video

window via the fiducial markers. To identify fixations on the quiz master’s face, we ran MediaPipe Face Mesh [35] on the screen recordings of the participant’s display. For each frame, we determined the facial region and tested whether the fixations falling in that frame lay inside it. Only the three video conditions could be analysed in this way. For F2F, only the scene-camera video was available, and it would have required a separate face-mapping pipeline.

A face was detected in 99.6 % of frames on average (minimum 87.3 % for a single recording), and 97.0 % of fixations fell on the video window. Each fixation was counted once, identified by its fixation ID. Valid eye-tracking data were available for 29 participants in BASEVC and TELE and for 28 in AI. For three participants no eye-tracking recording existed, and for one participant the AI recording contained no valid gaze data. Three recordings were shorter than five minutes (134–232 s).

The quiz master’s face occupied a smaller part of the screen in TELE than in the other two conditions (Sect. 3.2). Some recordings were also incomplete. We therefore computed metrics that depend neither on face size nor on recording length:

- the share of on-screen fixations that landed on the face;
- the share of on-screen dwell time spent on the face;
- the mean fixation duration (ms);
- the fixation rate (fixations per minute).

To describe the framing difference directly, we additionally report the face area as a share of the frame. Finally, we visualised the spatial distribution of fixations as heatmaps in face-normalised coordinates, that is, relative to the face centre and scaled by face width in each frame, displayed on a schematic face.

3.7 Data preparation and analysis

Data recovery.

Due to a software error, each condition questionnaire was prefilled with the participant’s answers from the previous condition. The application logged only values that participants changed, so 1270 of 7434 item responses (17.1 %) were not logged. These responses are prefilled values that participants submitted without changing them. We therefore coded each of them as the value displayed, that is, the participant’s answer to the same item in the preceding condition. Experimenter notes for three participants state that they adjusted only those answers that differed from the previous condition. The affected responses were concentrated in the video conditions (BASEVC 23.2 %, TELE 26.1 %, AI 18.0 %, F2F 0.1 %). An unchanged value may reflect genuine agreement or reduced diligence. In the latter case, these values attenuate differences between successive video conditions, which makes the respective contrasts conservative. We therefore repeated all analyses with logged values only.

For one participant, the questionnaire files were missing, and we recovered the responses from the application log using the same procedure. For two participants, the F2F conversation, which was first in their sequence, took place without the corresponding step being started in the software, so no F2F questionnaire was recorded. The software logged the subsequent calls as its first three sessions, and call logs and recordings confirm exactly three complete five-minute video calls for both participants.

Because the experimenters configured the gaze interventions manually following the session protocol, we assigned the three questionnaires in protocol order. Excluding both participants did not change any result. One participant’s AI questionnaire was incomplete (29 of 59 items). Missing responses were not imputed.

The conference version [7] excluded values beyond $\pm 2 SD$ and reported uncorrected contrasts. The analyses reported here supersede it.

Statistical analysis.

Omnibus tests. For each dependent variable, we fitted a linear mixed-effects model with condition as a fixed effect and a random intercept per participant, estimated by maximum likelihood. We compared it against an intercept-only model with a likelihood-ratio test ($\chi^2(3)$ for the questionnaires, $\chi^2(2)$ for the eye-tracking metrics). The false discovery rate (FDR) across the omnibus tests was controlled with the Benjamini–Hochberg procedure [36].

Pairwise contrasts. We computed all pairwise contrasts as paired t -tests on pairwise complete data, with Holm correction within each variable [37]. Each contrast was tested with its own error term rather than a pooled model residual, because response variability differed markedly between F2F and the video conditions. Effect sizes are Cohen’s d_z with exact 95 % confidence intervals based on the noncentral t distribution. In the text, we interpret contrasts for variables with a significant omnibus effect. For the comparisons of AI with BASEVC and TELE, we report effect sizes regardless. To distinguish the absence of an effect from a lack of evidence, we tested the comparisons of AI and TELE with BASEVC on key measures for equivalence. We used two one-sided tests (TOST) with bounds of $d_z = \pm 0.5$, standardised by the standard deviation of the paired differences [38]. With $N = 32$, the design had 80 % power to detect effects of $|d_z| \geq 0.51$ (two-sided, $\alpha = .05$; 0.65 at the strictest Holm-adjusted level). For the eye-tracking contrasts ($n = 28$), the corresponding threshold was 0.55.

Sensitivity analyses. All analyses were repeated in the following variants:

- with outlier removal ($\pm 2 SD$, as in the conference version);
- with logged values only;
- with rank-transformed data (rank-based mixed models and Wilcoxon signed-rank tests);
- with serial position as an additional covariate, since F2F was not fully counterbalanced;
- excluding the two participants with missing F2F questionnaires.

Results are reported in Online Resource 2. *Use of AI tools.* The re-analysis code was written with the assistance of the large language model Claude Opus 5.5 (Anthropic), used as a coding agent via the Claude Code extension for Visual Studio Code under the authors’ direction. The authors reviewed its output and verified the reported results against the original analysis. The same model assisted in drafting the manuscript text from the authors’ materials and decisions. The authors reviewed and revised all content and take full responsibility for it. All analyses were conducted in Python 3.12 using statsmodels 0.14.4 [39], pingouin 0.5.5 [40], SciPy 1.14.1 and pandas 2.2.3.

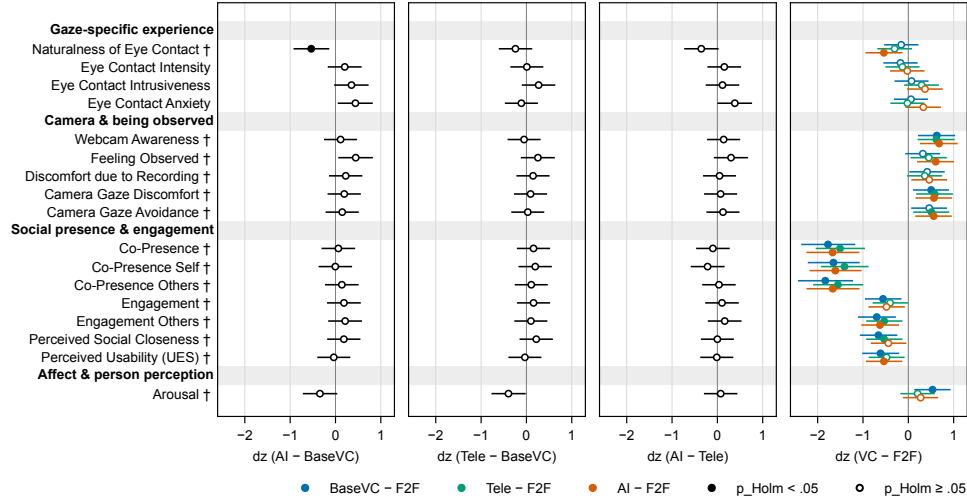


Fig. 2 Pairwise effect sizes (d_z with 95% CI) for all measures with a significant omnibus effect, plus eye contact anxiety, intensity and intrusiveness, grouped by construct. Panels show the contrasts AI – BASEVC, TELE – BASEVC, AI – TELE, and each video condition – F2F. Positive values indicate higher ratings in the first-named condition. Filled markers: $p_{\text{Holm}} < .05$; †: omnibus $p_{\text{FDR}} < .05$

4 Results

We report the full omnibus tests, all pairwise contrasts and descriptive statistics for every dependent variable in Online Resource 2. After FDR correction, 14 of the 36 dependent variables showed an effect of condition. Figure 2 summarises the pairwise effect sizes for the key measures. Table S4 in Online Resource 2 shows for every decision whether it held across the sensitivity analyses; we note below where it did not.

4.1 Social presence and engagement

The largest differences concerned co-presence. In all three video conditions, participants felt far less co-present with their partner than in F2F (F2F: $M = 90.3$; video conditions: $M = 51.6$ – 53.4). The effects were very large for BASEVC ($d_z = -1.77$ [$-2.34, -1.19$]), AI ($d_z = -1.67$ [$-2.23, -1.10$]) and TELE ($d_z = -1.51$ [$-2.03, -0.97$]; all $p_{\text{Holm}} < .001$). The same pattern held for both the self and the other perspective of co-presence.

Participants also perceived their partner as less engaged in all video conditions than in F2F (Engagement Others, $d_z = -0.53$ to -0.70 , all $p_{\text{Holm}} \leq .027$; for TELE not with outlier removal). Perceived social closeness was lower in BASEVC ($d_z = -0.66$, $p_{\text{Holm}} = .007$) and TELE ($d_z = -0.53$, $p_{\text{Holm}} = .033$) than in F2F, and perceived usability was lower in BASEVC ($d_z = -0.61$, $p_{\text{Holm}} = .013$) and AI ($d_z = -0.54$, $p_{\text{Holm}} = .031$) than in F2F. Except for social closeness in BASEVC, these contrasts were not significant with logged values only. The overall engagement score was lower

only in BASEVC than in F2F ($d_z = -0.56$, $p_{\text{Holm}} = .029$), and this effect did not hold in all sensitivity analyses.

Crucially, none of these measures differed between the three video conditions: neither the teleprompter nor the AI-based correction reduced the gap to F2F. Equivalence tests supported this conclusion. Co-presence in both AI and TELE was equivalent to BASEVC within $\pm 0.5 d_z$ ($p_{\text{TOST}} = .011$ and $.031$). Equivalence was also established for perceived social closeness in AI and for the partner’s engagement in TELE. It remained inconclusive for the partner’s engagement in AI ($p_{\text{TOST}} = .060$) and for social closeness in TELE ($p_{\text{TOST}} = .060$). Emotional contagion, comprehension, interdependence, felt closeness (IOS) and focused attention did not differ between conditions.

Of the affect and person-perception measures, only arousal differed between conditions ($\chi^2(3) = 11.35$, $p_{\text{FDR}} = .026$). Participants were more aroused in BASEVC than in F2F ($M = 0.56$ vs. 0.43 ; $t(29) = 2.93$, $p_{\text{Holm}} = .039$, $d_z = 0.53$ [$0.15, 0.91$]), with TELE (0.49) and AI (0.50) in between. The omnibus effect was not significant in the rank-based analysis. The conference version reported this difference in the opposite direction because the vertical axis of the Affect Grid had been coded inversely there. Valence, friendliness and competence did not differ.

4.2 Gaze-specific experience

Condition affected the perceived naturalness of eye contact ($\chi^2(3) = 14.84$, $p_{\text{FDR}} = .007$). Eye contact felt least natural with AI-based correction ($M = 67.8$), compared with BASEVC ($M = 76.9$), TELE ($M = 74.2$) and F2F ($M = 79.6$). It was rated as less natural in AI than in BASEVC, $t(30) = -2.98$, $p_{\text{Holm}} = .034$, $d_z = -0.54$ [$-0.91, -0.15$], and than in F2F, $t(28) = -2.91$, $p_{\text{Holm}} = .035$, $d_z = -0.54$ [$-0.93, -0.15$]; the latter contrast was not robust across sensitivity analyses. The difference between AI and TELE pointed in the same direction but was not significant ($d_z = -0.35$ [$-0.72, 0.01$], $p_{\text{Holm}} = .230$). The teleprompter did not differ from BASEVC or F2F. The decrease for AI relative to BASEVC was significant in all sensitivity analyses except the one based on logged values only.

Eye contact anxiety showed no significant omnibus effect ($\chi^2(3) = 6.67$, $p_{\text{FDR}} = .182$). Descriptively, however, anxiety was highest with AI ($M = 24.0$ vs. 16.7 – 18.8 in the other conditions). The contrasts with BASEVC ($d_z = 0.44$ [$0.07, 0.80$]) and TELE ($d_z = 0.39$ [$0.02, 0.75$]) were of small-to-medium size but did not survive Holm correction ($p_{\text{Holm}} = .123$ and $.200$). The anxiety effect reported in the conference version [7] is therefore not supported once multiple comparisons are taken into account.

Perceived intensity of the eye contact did not differ between conditions ($\chi^2(3) = 1.60$, $p_{\text{FDR}} = .678$). Perceived intrusiveness was descriptively highest in AI ($M = 25.0$; TELE 23.9 , BASEVC 18.7 , F2F 15.9). The omnibus effect was not significant, however ($\chi^2(3) = 7.41$, $p_{\text{FDR}} = .135$), and neither was the AI–BASEVC contrast ($d_z = 0.35$ [$-0.01, 0.71$], $p_{\text{Holm}} = .343$). Direct eye contact, own attention and the partner’s attentiveness did not differ between conditions.

4.3 Camera awareness and being observed

Video calls, regardless of intervention, made participants aware of the camera. Webcam awareness was higher in all three video conditions than in F2F ($d_z = 0.62$ – 0.68 , all $p_{\text{Holm}} \leq .009$; for BASEVC not with logged values only), whereas the video conditions were virtually identical (all $|d_z| \leq 0.14$; equivalence within $\pm 0.5 d_z$ for AI and TELE vs. BASEVC, both $p_{\text{TOST}} \leq .018$). Discomfort when looking into the camera was likewise higher in all video conditions than in F2F ($d_z = 0.50$ – 0.58 , all $p_{\text{Holm}} \leq .040$; equivalence between the video conditions, $p_{\text{TOST}} \leq .046$), although only the contrast for AI held in every sensitivity analysis. Participants also reported avoiding the camera more in AI ($d_z = 0.56$, $p_{\text{Holm}} = .028$) and TELE ($d_z = 0.51$, $p_{\text{Holm}} = .045$; not in the rank-based analysis) than in F2F.

Participants felt more observed in AI than in F2F ($d_z = 0.60$ [$0.21, 0.99$], $p_{\text{Holm}} = .015$). The contrasts between AI and BASEVC ($d_z = 0.45$ [$0.08, 0.81$], $p_{\text{Holm}} = .085$) and between TELE and F2F ($d_z = 0.45$, $p_{\text{Holm}} = .086$) did not reach significance after correction. Discomfort due to being recorded showed an omnibus effect ($p_{\text{FDR}} = .023$), but no pairwise contrast survived correction, and the omnibus effect was not robust across all sensitivity analyses. Intentional camera gaze did not differ between conditions.

Descriptively, AI had the highest mean among the video conditions on all six camera-related items. None of these differences to BASEVC or TELE was significant, however.

4.4 Gaze behaviour

AI-based gaze correction left participants' gaze behaviour unchanged (Fig. 3). Participants directed the same share of their fixations to the partner's face in AI as in BASEVC (.526 vs. .518; $d_z = 0.06$ [$-0.31, 0.43$]). Their fixations were equally long (414 vs. 418 ms; $d_z = -0.09$), and they fixated at the same rate (89.1 vs. 89.6 per minute). None of the AI–BASEVC contrasts exceeded $|d_z| = 0.12$. Equivalence tests established that AI and BASEVC were equivalent within $\pm 0.5 d_z$ on all gaze metrics and on face size (all $p_{\text{TOST}} \leq .026$). For the shares of fixations and dwell time on the face, equivalence held even within $\pm 0.4 d_z$.

The teleprompter condition differed from both other video conditions. Participants directed a smaller share of their fixations to the face in TELE (.448) than in AI ($d_z = 0.59$ [$0.18, 0.99$], $p_{\text{Holm}} = .013$) and BASEVC ($d_z = -0.49$ [$-0.87, -0.10$], $p_{\text{Holm}} = .027$). Their fixations were also longer in TELE than in AI (458 vs. 414 ms; $d_z = -0.56$, $p_{\text{Holm}} = .018$). These differences coincided with a much smaller face on screen. In TELE, the quiz master's face covered 6.4% of the frame, compared with 9.5% in BASEVC and 9.7% in AI ($\chi^2(2) = 75.39$, $p < .001$). For dwell-time share and fixation rate, the omnibus tests were significant, but no pairwise contrast survived correction.

In all three conditions, face fixations clustered in the upper, central part of the face (Fig. 4). They were shifted slightly to the left of the face centre (0.11–0.14 face widths). Because this shift occurred to the same extent in BASEVC, where the video was not processed, it cannot be attributed to either intervention. The difference map

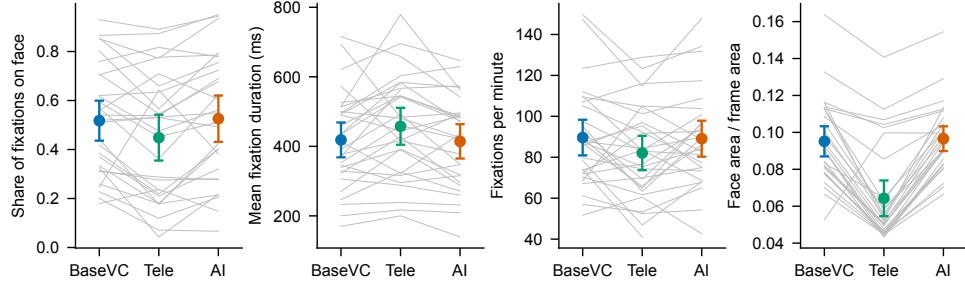


Fig. 3 Gaze behaviour in the three video conditions. From left to right: share of fixations on the partner’s face, mean fixation duration, fixation rate, and face area as a share of the video frame. Thin lines show individual participants; markers show condition means with 95 % CIs

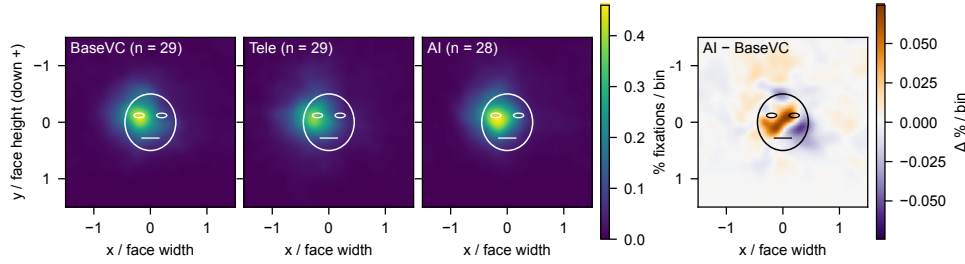


Fig. 4 Spatial distribution of face fixations in face-normalised coordinates for BASEVC, TELE and AI (shared colour scale; participants weighted equally), and the difference AI – BASEVC (diverging scale), shown on a schematic face

showed slightly more fixations around the eyes in AI and slightly fewer on the right cheek. These differences were small relative to the overall density (at most about 0.06 compared with a peak of about 0.45 percentage points per bin), and we did not test them inferentially.

5 Discussion

We compared a teleprompter and AI-based gaze correction with a standard video call and with face-to-face conversation, measuring both the communication experience and gaze behaviour. Four findings stand out.

5.1 The presence gap is not a gaze gap

The difference between video calls and face-to-face conversation was large and robust. Co-presence dropped by about 1.5 to 1.8 standard deviations, and perceived engagement, social closeness and usability declined as well. Neither intervention narrowed this gap: on none of these measures did TELE or AI differ from BASEVC. Restoring

the *appearance* of eye contact therefore does not restore the sense of being together, in line with findings from simulated teletherapy [22]. This fits evidence that live and mediated faces are processed differently because of the social context, not because of visual differences in the face itself [15]. Co-presence rests on a shared physical space and on a broad set of cues, of which eye contact is only one.

The camera-related measures point in the same direction. In every video condition, participants were more aware of the camera and more uncomfortable looking into it than face to face. The size of these effects did not depend on the intervention. Participants were also more aroused in the standard video call than face to face, which fits this picture of heightened self-monitoring, although the other two video conditions did not differ significantly from F2F. Gaze correction changes how the partner is displayed, but it leaves untouched the participant’s own experience of being on camera. This self-focused side of video communication is not addressed by current eye contact technologies.

5.2 Artificial eye contact feels less natural

The one robust effect of AI-based correction was the opposite of its purpose: eye contact felt less natural than in a standard video call. Two mechanisms are plausible.

- *Rendering artefacts.* In the screen recordings, we observed delayed pupil alignment and occasionally enlarged-looking eyes [7]. Near-realistic but imperfect renderings of socially salient features are a classic trigger of eeriness [28].
- *Disrupted gaze dynamics.* The correction continuously redirects gaze towards the camera. This removes the natural alternation of looking at and away from the partner that structures turn-taking [9]. Sustained direct gaze is experienced as intense [2], and gaze aimed exactly at the camera is not even the gaze that observers most readily perceive as eye contact [20].

Eye contact anxiety, perceived intrusiveness and the feeling of being observed followed the same pattern, with small-to-medium effects for AI relative to BASEVC ($d_z = 0.35$ – 0.45). These effects did not survive correction for multiple comparisons, and the study was powered only for effects of $|d_z| \geq 0.51$. A confirmatory test of effects of this size would require roughly 40 to 65 participants, and more under correction for multiple contrasts. Until then, the anxiety effect reported in the conference version should be regarded as a hypothesis rather than a finding.

5.3 Experience and gaze behaviour are decoupled

AI-based correction changed how eye contact felt, but not where participants looked. The share of fixations on the face, fixation durations and fixation rates were virtually identical to those in BASEVC. The unnaturalness of corrected gaze was therefore registered without any measurable redistribution of visual attention. This has a methodological consequence: gaze metrics alone would have classified the intervention as ineffective but harmless, while ratings alone could not show that participants kept attending to the face as usual. Evaluating gaze technologies requires both levels.

The teleprompter, in contrast, did change gaze behaviour: participants looked less at the face, and with longer fixations. These differences, however, coincided with a face that was a third smaller on screen, because the teleprompter required a greater distance between quiz master and camera. We therefore attribute them to framing rather than to gaze alignment. Framing is known to affect perceived distance in video calls [21]. Mechanical eye contact solutions thus change more than the gaze angle, and comparisons between approaches must control or at least report the resulting differences in framing.

5.4 Implications for design

Our results argue against pursuing ever more convincing *imitations* of face-to-face gaze. More promising directions are corrections that preserve natural gaze dynamics, for example partial correction or redirection towards the screen centre rather than the camera, and interventions that address the self-focus of being on camera. More broadly, our results support designing for the specific affordances of mediated communication rather than replicating co-presence [7].

5.5 Limitations

- *One AI system.* We tested a single AI implementation (NVIDIA Broadcast 2.0) that produced visible artefacts. Newer models may perform better, so our results do not generalise to AI-based gaze correction as such.
- *Asymmetric setup and task.* Only the quiz master’s gaze was corrected. In the guessing task, the quiz master held the solution and occasionally read from cards, which interrupted eye contact and created an asymmetric role structure.
- *Order.* F2F was scheduled only first or last. Ratings of several measures rose over the session, most likely due to growing rapport. The condition effects persisted when serial position was controlled, but position and condition are not fully separable.
- *Logging error.* About 17% of the responses were unchanged prefilled values. These may attenuate differences between video conditions. The naturalness effect of AI holds only if these values are treated as valid responses; it was not significant with logged values only.
- *Sample and eye tracking.* The sample was small and excluded people who wear glasses. Eye tracking was not available for F2F. A constant horizontal gaze offset of the eye tracker was not corrected; this affects absolute positions but not the comparisons between conditions.

6 Conclusion

Neither a teleprompter nor AI-based gaze correction brought video calls closer to face-to-face conversation. The large gap in co-presence remained unchanged, and awareness of the camera was a property of the medium rather than of the intervention. AI-based correction made eye contact feel less natural, while leaving participants’ gaze behaviour unchanged. Apparent effects of the teleprompter on gaze behaviour reflected differences in framing. Restoring the appearance of eye contact is thus not sufficient

to restore what video calls lack. Future work should preserve natural gaze dynamics, address the experience of being on camera, and evaluate gaze technologies at the level of both experience and behaviour.

Supplementary information. Online Resource 1 documents all questionnaire items and their scoring. Online Resource 2 contains the omnibus tests, all pair-wise contrasts, descriptive statistics, sensitivity analyses, position models, reliabilities, eye-tracking results, equivalence tests and the order distribution.

Declarations

- **Funding:** No funding was received for conducting this study.
- **Competing interests:** The authors have no competing interests to declare that are relevant to the content of this article.
- **Ethics approval and consent to participate:** The study was approved by the ethics committee of Faculty IV of Technische Universität Berlin (reference number 331). All participants gave written informed consent.
- **Consent for publication:** The person shown in Fig. 1 (first author) consented to the publication of the image.
- **Data availability:** The data are available from the corresponding author on reasonable request.
- **Materials availability:** The questionnaires are documented in Online Resource 1. Further materials are available from the corresponding author on reasonable request.
- **Code availability:** The analysis code is available from the corresponding author on reasonable request.
- **Author contribution:** Leonie Wichers: Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Writing – review & editing. Leon Schreiber: Project administration, Resources, Software, Writing – review & editing. Sebastian Möller: Supervision, Writing – review & editing. Robert P. Spang: Conceptualization, Methodology, Supervision, Validation, Writing – original draft. All authors read and approved the final manuscript.

References

- [1] Kendon, A.: Some functions of gaze-direction in social interaction. *Acta Psychologica* **26**, 22–63 (1967) [https://doi.org/10.1016/0001-6918\(67\)90005-4](https://doi.org/10.1016/0001-6918(67)90005-4)
- [2] Kleinke, C.L.: Gaze and eye contact: A research review. *Psychological Bulletin* **100**(1), 78–100 (1986) <https://doi.org/10.1037/0033-2909.100.1.78>
- [3] Bohannon, L.S., Herbert, A.M., Pelz, J.B., Rantanen, E.M.: Eye contact and video-mediated communication: A review. *Displays* **34**(2), 177–185 (2013) <https://doi.org/10.1016/j.displa.2012.10.009>
- [4] Kaiser, N., Henry, K., Eyjólfssdóttir, H.: Eye contact in video communication: Experiences of co-creating relationships. *Frontiers in Psychology* **13**, 852692 (2022) <https://doi.org/10.3389/fpsyg.2022.852692>

- [5] Isikdogan, L.F., Gerasimow, T., Michael, G.: Eye contact correction using deep neural networks. In: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 3307–3315 (2020). <https://doi.org/10.1109/WACV45572.2020.9093554>
- [6] NVIDIA Corporation: NVIDIA Broadcast App. <https://www.nvidia.com/en-gb/geforce/broadcasting/broadcast-app/>. Accessed 13 June 2025 (2025)
- [7] Wichers, L., Schreiber, L., Möller, S., Spang, R.P.: Gaze in the machine: The paradox of artificial eye contact in video calls. In: 2025 17th International Conference on Quality of Multimedia Experience (QoMEX) (2025). <https://doi.org/10.1109/QoMEX65720.2025.11219955>
- [8] Senju, A., Johnson, M.H.: The eye contact effect: mechanisms and development. *Trends in Cognitive Sciences* **13**(3), 127–134 (2009) <https://doi.org/10.1016/j.tics.2008.11.009>
- [9] Ho, S., Foulsham, T., Kingstone, A.: Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PLOS ONE* **10**(8), 0136905 (2015) <https://doi.org/10.1371/journal.pone.0136905>
- [10] Jarick, M., Kingstone, A.: The duality of gaze: eyes extract and signal social information during sustained cooperative and competitive dyadic gaze. *Frontiers in Psychology* **6**, 1423 (2015) <https://doi.org/10.3389/fpsyg.2015.01423>
- [11] Võ, M.L.-H., Smith, T.J., Mital, P.K., Henderson, J.M.: Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision* **12**(13), 3 (2012) <https://doi.org/10.1167/12.13.3>
- [12] Hessels, R.S.: How does gaze to faces support face-to-face interaction? A review and perspective. *Psychonomic Bulletin & Review* **27**, 856–881 (2020) <https://doi.org/10.3758/s13423-020-01715-w>
- [13] Gamer, M., Hecht, H.: Are you looking at me? Measuring the cone of gaze. *Journal of Experimental Psychology: Human Perception and Performance* **33**(3), 705–715 (2007) <https://doi.org/10.1037/0096-1523.33.3.705>
- [14] Boyle, E.A., Anderson, A.H., Newlands, A.: The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech* **37**(1), 1–20 (1994) <https://doi.org/10.1177/002383099403700101>
- [15] Zhao, N., Zhang, X., Noah, J.A., Tiede, M., Hirsch, J.: Separable processes for live “in-person” and live “zoom-like” faces. *Imaging Neuroscience* **1**, 1–17 (2023) https://doi.org/10.1162/imag_a_00027
- [16] Prinsen, J., Alaerts, K.: Eye contact enhances interpersonal motor resonance: comparing video stimuli to a live two-person action context. *Social Cognitive*

- and Affective Neuroscience **14**(9), 967–976 (2019) <https://doi.org/10.1093/scan/nsz064>
- [17] Hietanen, J.O., Peltola, M.J., Hietanen, J.K.: Psychophysiological responses to eye contact in a live interaction and in video call. *Psychophysiology* **57**(6), 13587 (2020) <https://doi.org/10.1111/psyp.13587>
 - [18] Roghanizad, M.M., White, R.E.: Predicting strangers’ cooperativeness in face-to-face vs. video-to-video interactions: A case of inaccurate social prediction in mediated communication. *International Journal of Information Management* **84**, 102913 (2025) <https://doi.org/10.1016/j.ijinfomgt.2025.102913>
 - [19] Jelson, A., Tausif, M.T., Lim, S.I., Khanna, S., Lee, S.W.: Investigating the effects of simulated eye contact in video call interviews. In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. CHI ’25. ACM, New York, NY, USA (2025). <https://doi.org/10.1145/3706598.3713282>
 - [20] Gao, A., Jayakumar, S., Maniglia, M., Curless, B., Kemelmacher-Shlizerman, I., Seitz, S.M., Seitz, A.R.: Don’t look at the camera: Achieving perceived eye contact in remote video communication. *Journal of Vision* **25**(11), 8 (2025) <https://doi.org/10.1167/jov.25.11.8>
 - [21] Grondin, F., Lomanowska, A.M., Békés, V., Jackson, P.L.: A methodology to improve eye contact in telepsychotherapy via videoconferencing with considerations for psychological distance. *Counselling Psychology Quarterly* **34**, 1–14 (2020) <https://doi.org/10.1080/09515070.2020.1781596>
 - [22] Grondin, F., Lomanowska, A.M., Poiré, V., Jackson, P.L.: Clients in simulated teletherapy via videoconference compensate for altered eye contact when evaluating therapist empathy. *Journal of Clinical Medicine* **11**(12), 3461 (2022) <https://doi.org/10.3390/jcm11123461>
 - [23] Wolf, L., Freund, Z., Avidan, S.: An eye for an eye: A single camera gaze-replacement method. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 817–824 (2010). <https://doi.org/10.1109/CVPR.2010.5540133>
 - [24] Yip, B.: Face and eye rectification in video conference using affine transform. In: *IEEE International Conference on Image Processing 2005*, vol. 3, pp. 513–516 (2005). <https://doi.org/10.1109/ICIP.2005.1530441>
 - [25] NVIDIA Corporation: Eye Contact – Model Overview. <https://build.nvidia.com/nvidia/eyecontact/modelcard>. Accessed 14 November 2025 (2025)
 - [26] He, Z., Wang, K., Feng, B.Y., Du, R., Perlin, K.: GazeChat: Enhancing virtual conferences with gaze-aware 3D photos. In: *The 34th Annual ACM Symposium on User Interface Software and Technology*. UIST ’21, pp. 769–782. ACM, New

York, NY, USA (2021). <https://doi.org/10.1145/3472749.3474785>

- [27] Greer, R., Dubnov, S.: Restoring eye contact to the virtual classroom with machine learning. In: Proceedings of the 13th International Conference on Computer Supported Education (CSEDU), pp. 698–708 (2021). <https://doi.org/10.5220/0010539806980708>
- [28] Mori, M., MacDorman, K.F., Kageki, N.: The uncanny valley [from the field]. IEEE Robotics & Automation Magazine **19**(2), 98–100 (2012) <https://doi.org/10.1109/MRA.2012.2192811>
- [29] Orne, M.T.: On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American Psychologist **17**(11), 776–783 (1962) <https://doi.org/10.1037/h0043424>
- [30] Biocca, F., Harms, C.: Guide to the Networked Minds Social Presence Inventory v. 1.2. <http://cogprints.org/6743/> (2003)
- [31] O’Brien, H.L., Cairns, P., Hall, M.: A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. International Journal of Human-Computer Studies **112**, 28–39 (2018) <https://doi.org/10.1016/j.ijhcs.2018.01.004>
- [32] Aron, A., Aron, E.N., Smollan, D.: Inclusion of other in the self scale and the structure of interpersonal closeness. Journal of Personality and Social Psychology **63**(4), 596–612 (1992) <https://doi.org/10.1037/0022-3514.63.4.596>
- [33] Russell, J.A., Weiss, A., Mendelsohn, G.A.: Affect grid: A single-item scale of pleasure and arousal. Journal of Personality and Social Psychology **57**(3), 493–502 (1989) <https://doi.org/10.1037/0022-3514.57.3.493>
- [34] Tonsen, M., Baumann, C.K., Dierkes, K.: A High-Level Description and Performance Evaluation of Pupil Invisible. arXiv preprint arXiv:2009.00508 (2020)
- [35] Kartynnik, Y., Ablavatski, A., Grishchenko, I., Grundmann, M.: Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs. arXiv preprint arXiv:1907.06724 (2019)
- [36] Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological) **57**(1), 289–300 (1995) <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- [37] Holm, S.: A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics **6**(2), 65–70 (1979)
- [38] Lakens, D.: Equivalence tests: A practical primer for t tests, correlations, and

- meta-analyses. *Social Psychological and Personality Science* **8**(4), 355–362 (2017) <https://doi.org/10.1177/1948550617697177>
- [39] Seabold, S., Perktold, J.: statsmodels: Econometric and statistical modeling with Python. In: *Proceedings of the 9th Python in Science Conference*, pp. 92–96 (2010). <https://doi.org/10.25080/Majora-92bf1922-011>
- [40] Vallat, R.: Pingouin: statistics in Python. *Journal of Open Source Software* **3**(31), 1026 (2018) <https://doi.org/10.21105/joss.01026>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESM1.pdf](#)
- [ESM2.pdf](#)