# GazeLens: Guiding Attention to Improve Gaze Interpretation in Hub-Satellite Collaboration

Khanh-Duy Le, Ignacio Avellino, Cédric Fleury, Morten Fjeld, Andreas Kunz

# GazeLens: Guiding Attention to Improve Gaze Interpretation in Hub-Satellite Collaboration

Khanh-Duy Le[1], Ignacio Avellino[2], Cédric Fleury[3], Morten Fjeld[1], and Andreas Kunz[4]

[1] Chalmers University of Technology, Sweden
{khanh-duy.le,fjeld}@chalmers.se
[2] ISIR, CNRS, Sorbonne Université, France
ignacio.avellino@sorbonne-universite.fr
[3] LRI, Univ. Paris-Sud, CNRS, Inria, Université Paris-Saclay, France
cedric.fleury@lri.fr
[4] ETH Zurich, Switzerland
kunz@iwf.mavt.ethz.ch

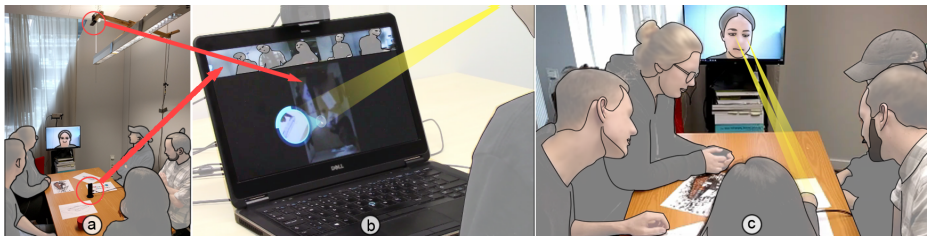


Fig. 1: *GazeLens* system. (a) On the hub side, a 360° camera on the table captures coworkers and a webcam mounted on the ceiling captures artifacts on the table. (b) Video feeds from the two cameras are displayed on the screen of the remote satellite worker; a virtual lens strategically guides her/his attention towards a specific screen area according to the observed artifact. (c) The satellite's gaze, guided by the virtual lens, is aligned towards the observed artifact on the hub space.

**Abstract.** In hub-satellite collaboration using video, interpreting gaze direction is critical for communication between hub coworkers sitting around a table and their remote satellite colleague. However, 2D video distorts images and makes this interpretation inaccurate. We present *GazeLens*, a video conferencing system that improves hub coworkers' ability to interpret the satellite worker's gaze. A 360° camera captures the hub coworkers and a ceiling camera captures artifacts on the hub table. The system combines these two video feeds in an interface. Lens widgets strategically guide the satellite worker's attention toward specific areas of her/his screen allow hub coworkers to clearly interpret her/his gaze direction. Our evaluation shows that *GazeLens* (1) increases hub coworkers' overall gaze interpretation accuracy by 25.8% in comparison to a conventional video conferencing system, (2) especially for physical artifacts on the hub table, and (3) improves hub coworkers' ability to distinguish between gazes toward people and artifacts. We discuss how screen space can be leveraged to improve gaze interpretation.

**Keywords:** remote collaboration · telepresence · gaze · lens widgets.

## 1 Introduction

In hub-satellite communication, a remote team member (satellite) collaborates at a distance with colleagues at the main office (hub). Typically, hub coworkers sit around a table with artifacts such as paper printouts, with a screen placed at one edge of the table showing a video feed of the satellite worker. The satellite worker sees the hub office in a deep perspective as it is captured by a camera placed at the edge of the table. Hub coworkers see a closer view of their colleagues, with a much more shallow perspective. Simply put, due to these differences in perspective, it is difficult to interpret the satellite's gaze (where s/he's looking at). While video conferencing systems can support non-verbal cues as people can see each others' faces and gestures, it is not always coherent: non-verbal cues such as gaze and deictic gestures are disparate between hub coworkers and the satellite, making communication asymmetric as co-located hub coworkers easily understand each others' non-verbal cues but not those of the satellite worker.

Gaze is important in collaboration - it is a reliable predictor of conversational attention [1, 4], offering effortless reference to spatial objects [29], supporting remote instruction [5, 30], and improving users' confidence in distributed problem solving on shared artifacts [2]. Kendon [27] argues that gaze is a signal through which a person relates their basic orientation and even intention toward another. Falling short on conveying gaze in remote collaboration can lead to confused communication [3], reduce social intimacy [3], decrease effectiveness [32] and increase effort for collaborative tasks [2, 5].

Previous work has tried to improve gaze perception in remote collaboration, but has mainly focused on conveying either gaze awareness between distant coworkers [7, 8, 25] or gaze on shared digital content [11, 14], leaving the problem of conveying gaze on physical artifacts rather under attended. Achieving this often requires specialized and complex hardware setups on the satellite side [9, 13, 16], which might be unrealistic for traveling workers. We focus on designing a mobile solution to improve hub coworkers' interpretation of the satellite worker's gaze both toward themselves and hub physical artifacts using minimal equipment.

We present *GazeLens*, a hub-satellite video conferencing system that improves hub coworker's accuracy when interpreting the direction of a satellite worker's gaze. At the hub side, *GazeLens* captures two videos: a view of the coworkers, using an off-the-shelf 360° camera, and a view of the artifacts on the table, using a ceiling-mounted camera. The system presents these videos simultaneously on the satellite worker's laptop screen, eliminating the need for stationary or specialized hardware on the remote end. *GazeLens* displays lenses on the satellite's screen, which the satellite worker can move to focus on different parts of the two videos, such as a hub coworker on the 360° view and an artifact on the table view. These lenses are strategically positioned to explicitly guide the direction of the satellite worker's gaze. As with conventional video conferencing, hub coworkers simply see a video stream of their remote colleague's face shown on the screen placed on the edge of their table. Our aim is to provide a more

coherent picture for hub coworkers of where exactly their satellite colleague is directing his/her attention, thus improving clarity of communication.

We evaluate the performance of *GazeLens* in two studies, where we compare it to conventional video conferencing (*ConvVC*) using a wide-angle camera on the hub side. The first study shows that *GazeLens* helps hub coworkers distinguish whether a satellite worker is looking at a person or at an artifact on the hub side. The second study shows that *GazeLens* helps hub coworkers interpret which artifact on the hub table the satellite worker is looking at. Early feedback on usability show the benefits for satellite workers, by improving visibility of hub artifacts and hub coworkers' activities while maintaining their spatial relations. We show that screen space can be better leveraged through strategic placement of interface elements to support non-verbal communication in video conferencing and thus convey a satellite worker's gaze direction.

## 2 Related Work

### 2.1 Gaze Awareness Among Remote coworkers

**One-to-one Remote Collaboration:** Gemmel et al. [24] and Giger et al. [25] proposed using computer vision to manipulate eye gaze in the remote worker's video. They focused on achieving direct eye contact by correcting the disparity between the location of the video conferencing window and the camera.
**Multi-party Remote Collaboration:** In the Hydra system [7], each remote party was represented by a hardware device containing a display, a camera, and a microphone. These were spatially arranged in front of the local worker, helping convey the worker's gaze.
**Group-to-group Remote Collaboration:** For each participant, MultiView [8] used one camera and one projector to capture and display each person on one side from the perspective of each person on the other. Similarly, MMSpace [13] placed multiple displays around the table of a local group, each representing a worker on the remote side, replicating the sitting positions of the remote workers. Both systems maintained correct gaze awareness between remote coworkers.
**Hub-satellite Remote Collaboration:** Jones et al. [15] installed a large screen on the satellite's side to display the hub's video stream and employed multiple cameras to construct a 3D model of the satellite worker's face to help hub coworkers perceive their gaze. Pan and Steed [9] and Gotsch et al. [16] also used an array of cameras to capture the satellite worker's face from different angles, selectively displaying the images to hub coworkers on a cylindrical display.

While most previous work focused on direct eye contact and gaze awareness between remote coworkers, only a few have attempted to provide correct interpretation of gaze toward shared artifacts - either virtual artifacts shared through a synchronized system or physical artifacts at either location - mandatory in hub-satellite collaboration that involves shared objects on the hub table. In addition, the above systems often require specialized hardware, which is not suitable for a traveling satellite worker who needs a lightweight and mobile device.

### 2.2 Gaze Support for Shared Virtual Artifacts

ClearBoard [11] creates a write-on-glass metaphor by overlaying a shared digital canvas on the remote coworker's video feed, inherently conveying gaze between two remote people working on the canvas. Similarly, Holoport [14] captures images of the hub's workers using a camera behind a scree, which helps convey coworkers' gaze among each other as well as towards on-screen artifacts. GAZE and GAZE-2 [1, 18] introduce a 3D virtual environment where the video stream of each worker is displayed on a 3D cube that could change direction to convey the worker's gaze toward others. Hauber et al. [12] evaluated a setup where workers were equipped with a tabletop showing a shared display, coupled with a screen showing the video feeds of the remote workers. A camera was mounted on top of the screen to capture remote workers' faces. They also compared this technique with the a 3D virtual environment of GAZE [1]. Finally, Avellino et al. [31] showed that video can be used to convey gaze and deictic gestures toward shared digital content in large wall-sized displays.

All these systems convey gaze direction to shared virtual artifacts by keeping the spatial relation of the video feed to the digital content. While they demonstrate that people can interpret gaze direction from a video feed, these techniques are not applicable in the context of hub-satellite collaborations involving physical artifacts on the hub table. These systems are designed for symmetric and specific settings such as large interactive whiteboards or wall-sized displays, which are not appropriate for mobile workers or small organisations.

### 2.3 Gaze Support for Physical Artifacts

Visualizations that indicate remote gaze direction have been explored for supporting physical collaborative tasks [2, 5, 29]. Otsuki et al. [17] developed ThirdEye, an add-on display that conveys the remote worker's gaze into the 3D physical space. It projects a 2D graphic element, controlled by eye tracking data of the remote worker, onto a hemispherical surface that looks like an eye. However, such mediated representations might introduce spatial disparities when compared to unmediated gaze, potentially leading to confusion and reducing the value of the satellite's video feed. These solutions add complexity to the satellite worker's setup, by adding specialized hardware such as an eye tracker.

Xu et al. [6] introduce an approach for conveying the satellite worker's attention in hub-satellite collaboration. The satellite worker can view a panoramic video stream of the hub on their screen, captured by a 360° camera, and manually select the area of interest in the video. A tablet on the hub's side, horizontally placed under the 360° camera, showed an arrow pointing at the area selected by the satellite worker. This solution cannot convey the satellite's attention toward physical artifacts as it lacks the vertical dimension of their gaze, and using an arrow to represent gaze might also be distracting and unnatural for the hub coworkers as compared to an unmediated gaze.

Finally, CamBlend [19] used video effects to blur the 180° video of the remote side, encouraging the user to focus on an area of interest in order to view it in

high resolution. This mimics a human's visual system, where foveal (central) vision has much higher acuity than peripheral vision [20]. CamBlend did not however aim to convey the satellite's gaze. *GazeLens* leverages this technique to provide the satellite worker with an overview of the hub's space, while guiding the satellite worker's attention to strategic locations in order to explicitly convey their gaze to the hub workers.

## 3 GazeLens Design

*GazeLens* is designed to improve the hub coworkers' perception of the satellite worker's gaze. It is motivated by the limitations of current video conferencing systems in conveying gaze.

### 3.1 Gaze Perception in Video Conferencing

Stokes [22] and Chen [10] showed that when the angle between the gaze direction and the camera is less than 5° in video conferencing between two people, the remote person perceives direct eye contact. Moreover, when one person looks towards the right of the camera, the remote person feels they are looking at their right shoulder, and so on. While this effect can be leveraged to establish eye contact between pairs of video conferencing endpoints [10], it may also be used for gaze interpretation in groups, such as hub-satellite settings.


Fig. 2: Hub table captured by a camera placed (a) below and (b) above the hub screen (image courtesy requested).

### 3.2 Limitations of Hub-Satellite Communication Systems

The screen on the hub side showing the satellite's video often uses a wide-angle camera that captures an overview of the hub environment, so the satellite worker can view the hub (Figure 2). Two typical placements for this camera are just above the screen, such as in the Cisco MX Series [34] and Polycom RealPresence Group Series [35] or below the screen, as in the Cisco SX80 [36] or AVS solutions [37]. Neither of these setups effectively conveys the satellite worker's gaze back to hub coworkers nor at the artifacts on the table. When the camera is placed below the screen, artifacts on the table are largely occluded or difficult to see, but the satellite sees hub worker's faces straight on (Figure 2a).

With a placement above the screen, the hub's artifacts are less occluded, but the hub's environment as a whole seems distant, with a distortion of deep perspective where coworkers appear small (Figure 2b, note the distant table edge). This "mapping" of the hub's environment onto the satellite's computer screen leads to hub coworkers being unable to distinguish the satellite's gaze toward different people and artifacts. Additionally, hub coworkers near the camera appear lower on the satellite's computer screen, making it harder for them to discern whether the satellite worker is looking at a coworker or at an object on the table.

### 3.3 Design Requirements

With these limitations in mind, we derived the following design requirements for a video conferencing system that can convey the satellite worker's gaze toward their hub coworkers and physical artifacts:

*DR1*: the system needs to display a view of the hub to the satellite worker in which they can see both the hub coworkers' faces and the artifacts on the table without occlusions.

*DR2*: the system should allow hub coworkers to clearly distinguish if the satellite worker is gazing toward individual coworkers or hub table artifacts.

*DR3*: the system should allow hub coworkers to accurately interpret the satellite worker's gaze toward physical artifacts.

*DR4*: the system should only rely on video to convey the satellite worker's gaze, and avoid mediated gaze representations such as arrows, pointers or virtual arms, which introduce spatial and representational disparities.

*DR5*: the system for the satellite worker should consist of a lightweight and mobile device which does not require any calibration, suitable for traveling.

### 3.4 *GazeLens* Implementation

*Hub side:* to ensure *DR1*, *GazeLens* captures the panoramic video of the coworkers sitting around the hub's table using a 360° camera placed at the center of it, and it captures the scene using a camera mounted on the ceiling to avoid occluding artifacts on the hub table (see Figure 1a).

*Satellite side: GazeLens* presents the two video feeds to the satellite worker on a standard laptop with a camera, satisfying *DR5* (see Figure 1b). Their presentation is designed so that it improves the interpretation of the satellite's gaze. To fulfill *DR2*, the video feed displaying hub coworkers should be placed near the satellite's laptop camera, located above the screen. This panoramic video is then segmented on the satellite's display to maintain spatial fidelity: the hub coworkers sitting in front of their screen are shown in the center of the satellite's video, while those on the sides of the hub table are displayed on their corresponding sides. The overview video of the hub table is displayed below the panoramic video of the hub coworkers (Figure 3).

To address *DR3*, the video of the hub table view is scaled to fit the satellite's screen and to maximize the size of any objects on it, although this leads to different gaze patterns depending on table shape. Stretching this video to maintain

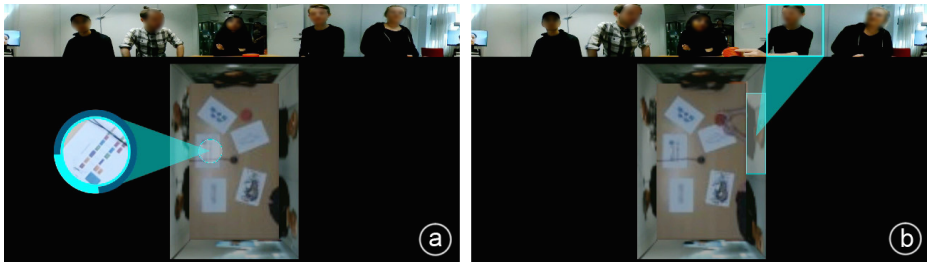


Fig. 3: *GazeLens* interface with (a) a lens showing a close-up of an artifact on the table and (b) a lens highlighting a hub worker's position around the table. Lenses are triggered when users click on the video feeds, the lens on artifacts is rotated either by dragging the handle or simply clicking on the border at the desired direction.

a specific size would solve this problem, but would also distort the objects. Instead, we chose a focus-based approach mimicking foveal and peripheral vision to maximize variation of the satellite worker's gaze, while preventing the hub table representation from becoming distorted.

The focus-based interaction is implemented as a widget in the form of a virtual lens that focuses on content. The hub's table video is displayed in the table's actual aspect ratio and "out of focus," using a video effect to mimic the indistinct quality of peripheral vision. The satellite worker thus sees an arrangement of artifacts but not their details. To see an object's details, the satellite selects it and a round virtual lens appears on the side showing a high-resolution detail of the selected area. This lens is strategically placed on the satellite's screen so that when the satellite worker gazes at it, the hub coworkers are able to correctly interpret which object is being looked at based solely from the direction of the satellite's gaze (see Figure 1c). This supports *DR4*. The lens position is interpolated by mapping the hub table's video onto as much screen space as possible below the hub coworker's panoramic video. As artifacts can be placed on the table from different directions, we implemented a rotation control on the lens, which the satellite worker can use to rotate its content if needed (Figure 3a).

To keep the satellite worker aware of the hub coworkers' spatial arrangement around the table, we segmented the hub's panoramic video based on the table's aspect ratio, and placed the segments around the table at their corresponding sides. These segments are then also displayed out of focus. When the satellite worker wants to look at one of their hub coworkers, they select it within the panoramic view at the top of the screen a square lens widget appears to guide their gaze toward a specific person (see Figure 3b).

*GazeLens* is implemented using C# and .NET 4.5 framework[5]. In its current implementation, the panoramic video height is equal to 20% of the entire screen height. When displayed on a 14-inch 16:9 conventional laptop screen, this creates a distance of around 4cm from the built-in camera to the top edge of the screen showing the panoramic video of the hub. Assuming that the satellite worker is 45

[5] https://docs.microsoft.com/en-us/dotnet/framework/

cm from their screen, and the hub screen is placed at the center of the table edge, this 4 cm distance creates the desired visual angle of 5° between the satellite's camera and the panoramic video showing their hub coworkers, establishing direct gaze [10]. This size is also sufficient to avoid distortions in the panoramic video. The lenses are activated by a left-button mouse click event on non-touch screen computers, and by a touch down event on touch devices.

## 4 Study 1: Accuracy in Interpreting Satellite's Gaze

We evaluate whether *GazeLens* can improve hub coworkers' ability to interpret a satellite's gaze by comparing it to a conventional video conference system (*ConvVC*). *ConvVC* displays the hub side in full screen on the satellite's screen, still guiding their gaze towards right direction. To our knowledge, no off-the-shelf video conferencing interfaces for laptop/tablet offer better unmediated gaze towards people and artifacts than a *ConvVC*. We test the following hypotheses:

- H1: *GazeLens* improves accuracy of gaze interpretation compared to *ConvVC*;
- H2: *GazeLens* outperforms *ConvVC* for gaze interpretation accuracy both when the hub coworker sits in front of and to the side of screen; and,
- H3: *GazeLens* incurs a lower perceived workload than *ConvVC*.

### 4.1 Method

The study has a within-subjects design with the following factors:

- Interface used by the satellite worker with levels: *GazeLens* and *ConvVC*;
- Position of the hub participants around the hub table with levels: *Front* and *Side* of the screen.

We controlled two secondary factors Actor and Target. We recorded 3 video sets of different Actors to mitigate possible effects tied to one of them in particular. Each Actor gazed at 14 Targets located on and around the table (Figure 4) as if he/she was the satellite worker.

Conditions were grouped by Position, then by Interface and then by Actor. The presentation order of these three conditions was counterbalanced using Latin squares. Each Latin square row was repeated when necessary. For each Position × Interface × Actor condition, the order of the 14 Targets was randomized so that successive videos never showed the same target as the previous one (and with a different Actors). Participants performed in total 168 trials (2 Positions × 2 Interfaces × 3 Actors × 14 Targets).

### 4.2 Participants

Twelve participants (7 male), aged 22 to 33 (median = 25), with backgrounds from computer science, interaction design, and social science participated in the study. This sample size is the average one reported in CHI studies [38] and also used in related work [17, 31]. Pilot studies determined that effects are strong

enough to be observed with this sample size. All participants had normal or corrected-to-normal vision. Three never used video conferencing applications, two used them on a monthly basis, five on a weekly basis, one on a daily basis and one multiple times a day. Each received a movie ticket for their participation.

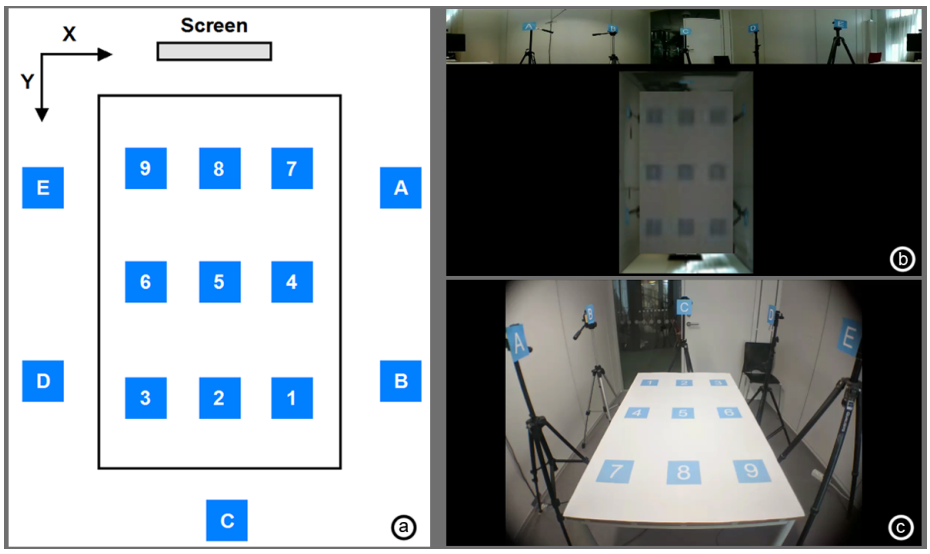


Fig. 4: (a) Hub space with target arrangement as used in Study 1. (b) *GazeLens* interface and (c) conventional interface *ConvVC* with the experimental setup.

### 4.3 Hardware and Software

For video recording the stimuli, we used a conventional laptop, a typical commercial foldable laptop with a screen size of $11-17$ inches (here, 14 inches) with a built-in front facing camera, for the satellite worker, as it is still one of the most common device used by travelers. Due to the low resolution of our laptop's built-in camera, we used a Plexgear 720p webcam[6] mounted at the same position as the laptop's built-in camera to record Actors. Using a high-res camera does not reduce the validity of the study as we are not investigating the effect of image quality. Also, many current conventional laptop models have high resolutions.

On the hub side, we used a 80cm × 140cm rectangular table with a height of 60cm to accommodate 6 people. The 14 Targets ($12cm \times 12cm$) were divided into two groups: 9 (labeled from 1 to 9) arranged in a $3 \times 3$ grid on the hub table represented artifacts, and 5 (labeled from A to E) around the table representing coworker targets. This left one edge occupied by the screen, two targets on each 140cm vertical edge and one on the 80cm horizontal edge (Figure 4). We used 5 hub coworkers in the study as it is a typical small team size and offers sufficient

[6] https://www.kjell.com/se/sortiment/dator-natverk/datortillbehor/webbkameror/plexgear-720p-webbkamera-p61271

challenge for interpreting gaze. Hub coworkers were 65cm higher than the table, approximated to the average eye-level height of a person sitting on a 45cm-high chair, the average for office chairs [23]. The distance from a coworker to its nearest neighbor (proxy or hub screen) was 80cm, and to the nearest edge of the table was 35cm. We used a 25-inch monitor to display the satellite's video stream, placed on a stand with the same height as the table.

To capture all the hub's targets in the *ConvVC* condition, due to our laboratory's hardware constraints, we simulated a wide-angle camera by coupling the 12-megapixel back camera of a LG Nexus 5X phone with a 0.67X wide-angled lens. The phone was mounted above the screen and adjusted so that the proxies were captured near the top edge of the video to convey the satellite worker's direct eye contact when looking at these targets (Figure 4-right/bottom). For the *GazeLens* condition, we used a Ricoh R 360° camera to capture panoramic video and a Logitech HD Pro Webcam C910 to capture the overview video of the table (Figure 4-right/top).

Participants sat at two positions around the hub's table: in *Front*, opposite the screen and at the of the screen (positions A and C respectively on 4a). The distance from the participant's body to the nearest edge was around 35cm. As the table was symmetric, the evaluation result from one side could be applied to the other. We chose position A as it was closer to the screen than B or D, causing the so-called Mona Lisa effect (where the image of a subject looking into the camera is seen by remote participants as looking at them, irrespective of their position) that could affect the gaze interpretation. The recorded videos were displayed at full screen. The videos' aspect ratio (4:3) mismatched the hub's screen (16:9). However, we did not modify the videos' size to avoid a partial or distorted view of the hub.

### 4.4 Procedure

After greeting participants, they signed a consent form and read printed instructions. Participants answered a pre-study questionnaire providing their background and self-assessing their technological expertise. They completed a training session before starting the experiment. We encouraged them to take a 5-minute break between the two POSITION conditions and a 2-minute break between each 21 videos (middle and at the end of each INTERFACE condition). It took 1 hour 15 minutes for a participant to complete the study. Finally, they answered the post-study questionnaires and received a movie ticket.

**Video Recordings.** We recorded 6-second videos of 3 different ACTORs gazing at 14 targets in the satellite worker interface displayed on a 14-inch laptop screen, for both *GazeLens* and *ConvVC*. We observed in pilots that 6 seconds are long enough to make ACTOR's gaze movements perceivable while avoiding fatigue. *ActorA* was a 29-year-old man with brown medium-length hair and hazel eyes, *ActorB* a 34-year-old man with short blonde hair and brown eyes, and *ActorC* a 44-year-old woman with brown pulled back hair and hazel eyes.

ACTORs sat on a 45cm-high office chair 45cm away from the laptop screen, which was placed on a 70cm-high office desk. In order to recreate a more realistic

gaze, actors first looked at a starting point and then at the target. This causes relative movements in the satellite worker's gaze, which provides a context with easier interpretation for the viewer. A target's starting point was decided by choosing its nearest neighboring label at an arbitrary point of 50cm, with the exception of labels on the same grid row and column as the target. As humans are less sensitive to vertical changes of gaze [10] and the distance between two targets on the same row in conventional videos is smaller, satellite workers eye movements become be noticeable. For each target, we recorded actors gazing at them from three different starting points. We did not use a chin-rest for the actors to make the recording realistic, however they were instructed to keep their head straight. They were also instructed to look at the targets in natural ways (i.e. they could turn their head if needed).

**Task.** Participants were advised to sit upright at POSITIONs, and could lean back if they got tired. However, if seated at the *Side*, they were not allowed to lean toward the screen. Participants watched each video playing in an infinite loop to avoid missing gaze movements due to distractions. There was thus no time pressure for the participants as we focused on accuracy. When they were ready to answer which target the ACTOR was looking at, they tapped a large "Stop" button on an Asus Nexus 9 tablet. The tablet then showed a replica of the table with the targets and hub coworkers laid out in the same fashion as on the participant's screen to make selection easier.

### 4.5 Data Collection and Analysis

We collected participants' responses for each trial, i.e. which target they thought was being gazed at, and their confidence in their answer (on 5-point Likert scale: 1 = not confident, 5 = very confident). We also recorded response time. When two INTERFACE conditions for each POSITION were completed, participants answered a post-questionnaire indicating their perceived workload (based on NASA TLX [33]), perceived ease to differentiate gazes at targets on and around the table, perceived ease to interpret the satellite's gaze and their interpretation strategies in both conditions.

We define *Gaze Interpretation Accuracy* as the proportion of correct trials. We define *Differentiation Accuracy*, i.e. the participant's ability to differentiate gaze at targets *around* or *on* the table, as the proportion of trials with gaze at the correct set of targets on or around table.

### 4.6 Results

To analyze *Gaze Interpretation Accuracy* we perform a two-way factorial ANOVA (INTERFACE × POSITION). The result (Figure 5) shows an effect of INTERFACE ($F_{1,44} = 7.33$, $p < 0.001$), POSITION ($F_{1,44} = 6.88$, $p < 0.01$) and no interaction effect INTERFACE × POSITION ($p > 0.1$). *GazeLens* significantly improves interpretation of the satellite gaze in comparison to *ConvVC* ($31.45\% \pm 4.67\%$ vs $25\% \pm 3.51\%$, an increase of 25.8%, 6.45% effect size), supporting H1. As expected, participants interpreted the satellite's gaze significantly more precisely

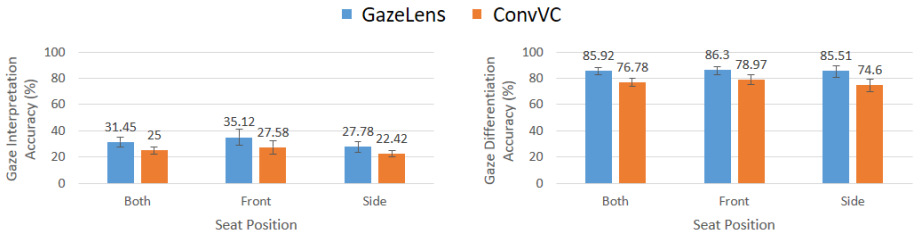


Fig. 5: *Gaze Interpretation Accuracy* (in %) (left) and *Gaze Differentiation Accuracy* (in %) (right) for INTERFACE ×POSITION. Error bars show 95% confidence interval (CI).

at the *Front* than *Side* POSITION (31.35% ± 4.96% vs 25% ± 3.13%). As there is no interaction effect, we cannot reject H2. Data in Figure 5 suggests that participants performed better with *GazeLens* at both sitting positions.

To analyze *Differentiation Accuracy* we perform a two-way factorial ANOVA (INTERFACE × POSITION). The result (Figure 5) shows an effect of INTERFACE ($F_{1,44} = 20.77$, $p < 0.001$), no effect of POSITION nor an interaction effect of INTERFACE ×POSITION ($p > 0.1$). Participants using *GazeLens* could better differentiate gaze at targets around or on the table compared to *ConvVC* (85.92%±2.38% vs. 76.78%±3.14%, $p < 0.001$).

A two-way factorial ANOVA analysis (INTERFACE × POSITION) did not yield any effect of INTERFACE × POSITION on perceived workload (not supporting H3), neither for answer time nor confidence (all $p$'s $> 0.1$). Finally, we did not find any effect of ACTOR on *Gaze Interpretation Accuracy* nor *Differentiation Accuracy* at targets on or around table, neither learning effects.

## 5 Study 2: Accuracy in Interpreting Gaze at Hub Artifacts

Study 1 showed that *GazeLens* improves gaze interpretation accuracy in general. We wanted to further investigate how accurately hub coworkers can interpret the satellite worker's gaze towards physical artifacts on the hub table. In reality, arrangements of physical artifacts on the table can vary from sparse (e.g. meetings with some paper documents) to dense (e.g. brainstorming with sticky notes, phones, physical prototypes). This prompts the need to explore how the granularity of artifact arrangement impacts a hub coworker's gaze interpretation. We also investigate if *GazeLens* can increase hub coworkers' accuracy at interpreting the satellite's gaze compared to *ConvVC*, especially regarding error distance along the two table dimensions: horizontal (X) and vertical (Y). We used a similar experiment design to Study 1, where participants have to determine the satellite's gaze in prerecorded videos displayed on the screen at the hub table.

We operationalize artifacts arrangements through the granularity of layouts:

- 3×3 (9 objects in a 3×3 grid): low-granularity arrangements to investigate gaze interpretation accuracy in meeting scenarios involving paper documents,
- 5×5 (25 objects in a 5×5 grid): high-granularity arrangements to investigate gaze interpretation accuracy in scenarios such as brainstorming.

We formulate the following hypotheses:

- H1: *GazeLens* improves the hub coworkers' interpretation accuracy for gaze toward objects on the hub table compared to *ConvVC*;
- H2: *GazeLens* outperforms *ConvVC* for gaze interpretation accuracy at both levels of granularity;
- H3: *GazeLens* reduces X and Y error distance compared to *ConvVC*.

### 5.1 Method

The within-subject study design has the following factors:

- Interface used by the satellite to view the targets, with two conditions: *GazeLens* and *ConvVC*; and,
- Layout of the artifacts on the table with two conditions: 3×3 and 5×5 grid.

For each participant, the conditions were grouped by Layout, then by Interface and then by Actor. Actors were the same as in Study 1.

The order of presentation was counterbalanced across conditions using Latin squares for the first three conditions and randomized order for Target. Each Latin square was repeated when necessary. For each Layout × Interface × Actor condition, the order of the targets (9 for 3×3 and 25 for 5×5) was randomized so a different succession of videos was shown for each target. Participants took a 5-minute break between the two layout conditions, and performed a training session before starting the experiment, where we ensured they covered all Targets, Interfaces and Layouts.

### 5.2 Participants

Twelve participants—different from those in Study 1—8 males, aged 22 to 38 (median = 29), with backgrounds from computer science, interaction design, and social science participated in the study. All had normal or corrected-to-normal vision. Three used their computer on daily basis, eight multiple times a day and one on a weekly basis. One had never used video conferencing applications, eight used them on a monthly basis, one on a weekly basis, one on a daily basis, and one multiple times a day. Each received a movie ticket for their participation.

### 5.3 Hardware and Software

We used the same cameras, hub table, hub screen, and screen placement as in Study 1. To investigate gaze interpretation accuracy for different artifact sizes, we used two different layouts on the table (Figure 4). We removed targets representing hub coworkers in Study 1 to avoid distracting actors and participants.

### 5.4 Procedure

We employed a similar procedure as in Study 1. However, participants took a 2-minute break after every 18 videos in the 3×3 layout, and after every 15 videos in the 5×5 layout (the dense layout was more tiring).

**Video Recordings.** We recorded 306 6-second videos for the hub's targets of the same three ACTORs as in study 1: 81 videos for the 3×3 layout and 225 for the 5×5 layout. We used the same laptop, camera, placements of the devices and ACTORs as in Study 1. Each video was also recorded in a similar procedure as in Study 1: each ACTOR first looked at a starting point and then at the target. We used the same criteria for choosing starting points for targets.

**Task.** We used a similar task as in Study 1, although participants only sat at position A (*Front*) to watch the videos. The positions of the screen and those of participants in relation to it remained the same.

### 5.5 Data Collection and Analysis

We collected data as in Study 1. We measure *Gaze Interpretation Accuracy* as in Study 1 and two error measures: *X-Axis Error* and *Y-Axis Error*, denoting the error between the correct and selected target along the table's horizontal and vertical orientation (X and Y axis in Figure 4a) respectively.

### 5.6 Results

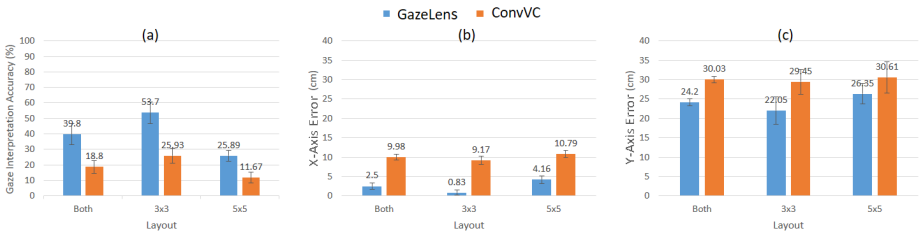


Fig. 6: (a) Gaze Interpretation Accuracy (in %) for each INTERFACE × LAYOUT condition. (b) X-Axis Error (in cm) and (c) Y-Axis Error (in cm) for each INTERFACE × LAYOUT condition. Bars indicate 95% CI.

We analyze *Gaze Interpretation Accuracy* as in Study 1 by performing a two-way factorial ANOVA (INTERFACE and LAYOUT). The result shows an effect of INTERFACE ($F_{1,44} = 69.26$, $p < 0.001$), supporting H1, LAYOUT ($F_{1,44} = 69.50$, $p < 0.001$) and INTERFACE × LAYOUT ($F_{1,44} = 7.214$, $p < 0.05$). A post-hoc Tukey HSD test showed that *GazeLens* significantly improves *Gaze Interpretation Accuracy* in both 3×3 layout (53.7%±7.06% vs 25.93%±4.81%,$p <$ 0.001) and 5×5 layout (25.89%± 3.55% vs 11.67%± 3.5%, $p < 0.001$) supporting H2. Post-hoc Tukey HSD tests showed significant differences between *GazeLens* with 3×3 and *GazeLens* with 5×5 ($p < 0.001$), between *GazeLens* with 3×3 and *ConvVC* with 5×5 ($p < 0.001$), between *ConvVC* with 3×3 and *ConvVC* with

5×5 ($p < 0.01$). Figure 6a shows gaze interpretation accuracy in each INTERFACE × LAYOUT condition.

To examine *X-Axis Error*, we perform a two-way factorial ANOVA analysis with INTERFACE and LAYOUT as factors. The analysis shows an effect of INTERFACE ($F_{1,44} = 251.59$, $p < 0.001$), partially supporting H3, LAYOUT ($F_{1,44} = 27.54$, $p < 0.001$) and no effect of INTERFACE × LAYOUT ($F_{1,44} = 3.255$, $p > 0.05$). For *Y-Axis Error* we perform a two-way factorial ANOVA analysis with INTERFACE and LAYOUT as factors. The analysis shows an effect of INTERFACE ($F_{1,44} = 11.29$, $p < 0.005$), partially supporting H3, but no effect of LAYOUT and INTERFACE × LAYOUT (all $p > 0.1$).

We did not find any effect of INTERFACE on answer time, self-confidence, perceived workload and perceived ease of gaze interpretation (all $p > 0.1$). No learning effect was found in term of gaze interpretation accuracy, X and Y-axis error. Figure 7 visualizes the X and Y error distances at each target by INTERFACE and LAYOUT. Figure 6 (b,c) shows X and Y error distance in each INTERFACE × LAYOUT condition.

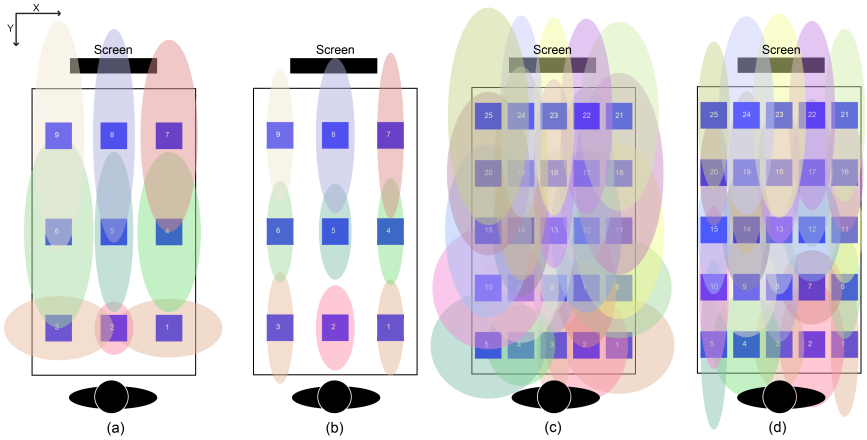


Fig. 7: X and Y-Axis Error visualization at each target in Study 2, in (a) 3×3 layout using *ConvVC*, (b) 3×3 layout using *GazeLens*, (c) 5×5 layout using *ConvVC*, (d) 5×5 layout using *GazeLens*. Zero error is shown by an ellipse-axis equal to the target size.

## 6 Early User Feedback of GazeLens

The two previous experiments evaluated *GazeLens* on the hub side. We gather in a last study early user feedback on its usability from the satellite worker's perspective. We recruited five pairs of participants (8 male, 2 female, aged from 23 to 50, median 31) to solve a remote collaborative task. Participants had various backgrounds from computer science, software engineering, and social sciences. Participants in each pair knew each other well. Designing an experimental collaborative task for hub-satellite collaboration involving physical artifacts is complicated by the complex communication required between coworkers and artifacts.

To our knowledge, there is still no standardized experimental task for this. As we focus on gathering feedback on the satellite's side, for simplicity, we chose a standard task commonly used when investigating remote collaboration on physical tasks: solving a puzzle by arranging a set of pieces into a predefined picture.

Each pair of participants consisted of a *worker* on the hub side and an *instructor* on the satellite side. The *worker* had all the puzzle pieces on the hub table, but did not know the solution. The *instructor* knew the solution and communicated with the worker via audio and video to guide them selecting and arranging the pieces. This task can trigger movements of the hub worker, their hands, and the artifacts on and around the table, which could be perceived differently by the satellite worker on different video conferencing interfaces. Each puzzle consists of 16 rectangular pieces, chosen so that they were hard to be verbally described by color and visual patterns. We used the same laboratory setup as in Study 1 and 2.

Participants performed the tasks in both interfaces on the satellite side, *GazeLens* and *ConvVC*, in order to have comparative views on their usability. They familiarized themselves for about 15 minutes with each interface, and had 10 minutes to solve the task in each condition. Two different $50cm \times 80cm$ puzzles with comparable levels of visual difficulty were used for two conditions. The conditions and puzzle tasks were counter-balanced. We gathered qualitative feedback in an interview after participants went through both conditions.

Only one *instructors* our of four reported perceiving inconvenient using *GazeLens* lenses. He reported that activating the lens on the table by mouse click was quite tiring and suggested using mouse wheel scroll events to make the activation similar to zooming. All *instructors* reported that it was easier for them to see the puzzle pieces' content in *GazeLens*, as those at the far edge of the table were hard to see in *ConvVC*, where they sometimes had to ask the performer to hold and show it to the camera. One *instructor*, who often uses Skype for hub-satellite meetings, really liked the concept of people around the table in a panoramic video connected via a virtual lens. He could imagine that it could help him clearly see everyone while knowing where they are sitting around the table and what they are doing on the table during a meeting. Besides that, two *instructors* reported that when the virtual lens was on the top of *GazeLens*' table area it might obscure *workers'* hand gestures.

## 7 Discussion

### 7.1 GazeLens Improves Differentiating Gaze Towards People vs. Artifacts

Study 1 showed that *GazeLens* improves the satellite worker's gaze interpretation accuracy toward hub coworkers and, in particular, they are able to distinguish with more than 85% accuracy if the satellite worker is gazing towards them or towards the physical artifacts on the table. This is due to the position of the panoramic video of hub coworkers' at the top of the satellite interface, close to the

camera, and the position of the artifact overview at the bottom of the interface. To further improve this, we could explore increasing the gap between these two views to obtain a larger, more distinguishable, difference in gaze direction.

### 7.2 GazeLens Improves Gaze Interpretation Accuracy

Study 2 showed that *GazeLens* improved gaze interpretation accuracy for table artifacts not only in sparse ($3 \times 3$) but also dense ($5 \times 5$) arrangements. This can be explained by how the entire laptop screen is used to make the satellite's worker's gaze better aligned with the hub artifacts as compared to the *ConvVC* condition, as stated by P10: *"To determine where the satellite worker was gazing, I could imagine a line of sight from his eyes to the table/person"*. We argue that in *ConvVC*, due to the perspective projection of the hub table, distances between objects at the far edge of the table appear too small, making gaze toward them indistinguishable. This was confirmed by participant comments about the satellite's gaze in *ConvVC*: *"It was hard compared to the other condition [*GazeLens *] because they just stared at the table and I had no clue which number was the exact one" (P6)*; *"the differences between different gazes felt very small" (P5)* and *"They were looking more at the center" (P9)* . In contrast to *ConvVC*, participants perceived the satellite worker's gaze in the *GazeLens* condition as *"more obvious"*, *"clearer"* and *"easier to determine where they are looking"*.

Small between-object distances in *ConvVC* also caused negligible eye movements in the videos where the satellite worker gazes from the starting point to the target: *"They moved less and thus gave me fewer references to be able to get a picture of what they were looking at"*; *"Eye movements were very small and the angles were hard to calculate in my head"* and *"The eyes did not move and I got confused"*. In contrast, participants perceived eye movement in *GazeLens* condition as *"clearer"*, *"easy to distinguish from side to side"*, *"enough to follow"*, *"sometimes added with head movements, easier to determine"*.

When investigating X-Axis and Y-Axis Error, it was not surprising that horizontal gaze changes were perceived more accurately than vertical ones, as people are more sensitive to horizontal gaze changes, especially when the gaze is below the satellite's camera [10]. Furthermore, laptops have landscape screens, leaving less vertical space to position the lens than in the horizontal direction-making gaze differences more distinguishable in the horizontal orientation. In future work, we want to explore how to improve gaze perception in the vertical dimension.

### 7.3 Limitations and Future Work

Although most of the participants reported that the satellite worker's gaze was clear and easy to interpret with *GazeLens*, two participants in Study 1 reported that they did not feel the satellite worker was looking at any markers in particular, and their answers were just an approximation based on gaze. This can be explained by the fact that at that moment *GazeLens* did not precisely calculate the screen mapping based on the actual size of the table and the distance from

the coworkers to the hub table. Achieving geometrically corrected gaze in video communication is almost impossible, as it depends on several parameters that cannot all be easily acquired in real-life hub-satellite scenarios, such as camera focal length, camera position, video size, camera-scene, and screen-viewer distance. *GazeLens*' mapping strategy is effective at improving gaze interpretation and yet simple enough to be deployed in realistic scenarios. In future, we will consider replacing the ceiling-mounted camera with a depth-sensing camera, which can acquire the table size and coworkers's distance from the table in order to improve mapping. We are also interested to further study *GazeLens* with different hub table shapes, sizes and layouts, using the current screen mapping strategy and others. Likewise, due to the emerging use of tablets for work purposes, it would be valuable to study *GazeLens* on tablet devices both in portrait and landscape display mode.

In our last study, participants perceived *GazeLens* positively, without usability issues. Still, we plan to improve the system by making the virtual lens over the table less occlusive in the future setup using a depth-sensing camera, by detecting the presence of hub workers' hand gestures and dynamically adjusting the opacity of the lens. Besides that, we plan to study how expertise might influence time needed to learn *GazeLens*, as we think that probably this is not enough to make an impact on the hub side, which is unaware of what is shown on the satellite's interface. Lastly, we plan to extended *GazeLens* to support multiple satellites, for instance by representing each one by a screen placed around the hub table and the corresponding video feeds adjusted accordingly (e.g. re-segment panoramic video, change orientation of the table's video).

## 8 Conclusion

While conventional hub-satellite collaboration typically employs video conferencing, it is difficult for hub coworkers to interpret the satellite worker's gaze. Previous work supporting gaze between remote workers has not addressed shared physical artifacts used in collaboration, and support for conveying gaze in remote collaboration with asymmetric setups is still limited. We designed *GazeLens*, a novel interaction technique supporting gaze interpretation that guides the attention of the satellite worker by means of virtual lenses focusing on either hub coworkers or artifacts. In our first study, we showed that *GazeLens* significantly improves gaze interpretation over a conventional video conferencing system; and also that it improves hub coworkers' ability to differentiate the satellite's gaze toward themselves or artifacts on the table. In our second study, we found that *GazeLens* improves hub coworkers' interpretation accuracy for gaze toward objects on the table, for both sparse and dense arrangements of artifacts. Early user feedback informed us about the advantages and potential drawbacks of *GazeLens*' usability. *GazeLens* shows that the satellite worker's laptop screen can be fully leveraged to guide their attention and help hub coworkers more accurately interpret their gaze.

## References


1. Vertegaal, R.: The GAZE groupware system: mediating joint attention in multiparty communication and collaboration. In: CHI '99 Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 294–301, ACM New York,1999.
2. Akkil, D., James, J.M.,Isokoski, P., Kangas, J.: GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks. In: CHI EA '16 Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 1151–1158, ACM New York,2016.
3. Vertegaal, R.,van der Veer, G., Vons, H.: Effects of Gaze on Multiparty Mediated Communication. In: Proceedings of Graphics Interface 2000, pp. 95–102, ACM New York,2010.
4. Vertegaal, R., Slagter, R., Van der Veer, G., Nijholt, A.: Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In: CHI '01 Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 301-308, ACM New York, 2001.
5. Higuch, K., Yonetani, R., Sato, Y.: Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. In: CHI '16 Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 5180–5190, ACM New York, 2016.
6. Xu, B., Ellis, J., Erickson, T.: Attention from Afar: Simulating the Gazes of Remote Participants in Hybrid Meetings. In: DIS '17 Proceedings of the 2017 Conference on Designing Interactive Systems, pp. 101–113, ACM New York, 2017.
7. Sellen, A., Buxton, B., Arnott, J.: Using spatial cues to improve videoconferencing. In: CHI '92 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 651–652, ACM New York, 1992.
8. Nguyen, D., Canny, J.: MultiView: spatially faithful group video conferencing. In: CHI '05 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 799–808, ACM New York, 2005.
9. Pan, Y., Steed, A.: A Gaze-preserving Situated Multiview Telepresence System. In: CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2173–2176, ACM New York, 2014.
10. Chen, M.: Leveraging the Asymmetric Sensitivity of Eye Contact for Videoconferencing. In: CHI '02 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 49–56, ACM New York, 2002.
11. Ishii, H., Kobayashi, M.: ClearBoard: a seamless medium for shared drawing and conversation with eye contact. In: CHI '92 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 525–532, ACM New York, 1992.
12. Hauber, J., Regenbrecht, H., Billinghurst, M., Cockburn, A.: Spatiality in videoconferencing: trade-offs between efficiency and social presence. In: CSCW '06 Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, pp. 413–422, ACM New York, 2006.
13. Otsuka, K.: MMSpace: Kinetically-augmented telepresence for small group-to-group conversations. In: Proceedings of 2016 IEEE Virtual Reality (VR), IEEE, 2016.
14. Küchler, M., Kunz, A.: Holoport-a device for simultaneous video and data conferencing featuring gaze awareness. In: Proceedings of Virtual Reality Conference 2006, pp. 81–88, IEEE, 2006.
15. Jones ,A. , Lang, M., Fyffe, G., Yu, X., Busch, J., McDowall, I., Bolas, M., Debevec, P.: Achieving eye contact in a one-to-many 3D video teleconferencing system. In: ACM Transactions on Graphics (TOG), **28**(3),2009.

16. Gotsch, D., Zhang, X., Meeritt, T., Vertegaal, R.: TeleHuman2: A Cylindrical Light Field Teleconferencing System for Life-size 3D Human Telepresence. In: CHI '18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 552, ACM New York, 2018.
17. Otsuki, M., Kawano, T., Maruyama, K., Kuzuoka, H., Suzuki, Y.: ThirdEye: Simple Add-on Display to Represent Remote Participant's Gaze Direction in Video Communication. In: CHI '17 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 5307–5312, ACM New York, 2017.
18. Vertegaal, R., Weevers, I., Sohn, C., Cheung, C.: GAZE-2: conveying eye contact in group video conferencing using eye-controlled camera direction. In: CHI '03 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 521–528, ACM New York, 2003.
19. Norris, J., Schnädelbach, H., Qiu, G.: CamBlend: An Object Focused Collaboration Tool. In: CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 627-636, ACM New York, 2012.
20. Bailey, R., McNamara, A., Sudarsanam, N., Grimm, C.: Subtle gaze direction. In: ACM Transactions on Graphics (TOG) **28**(4), 2009.
21. Hata, H., Koike, H., Sato, Y.: Visual Guidance with Unnoticed Blur Effect. In: AVI '16 Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 28–35, ACM New York, 2016.
22. Stokes, R.: Human Factors and Appearance Design Considerations of the Mod II PICTUREPHONE Station Set. In: ACM Transactions on Graphics (TOG) **17**(2), 1969.
23. Average Human Sitting Posture Dimensions Required in Interior Design, https://gharpedia.com/average-human-sitting-posture-dimensions-required-in-interior-design/.
24. Gemmell, J., Toyama, K., Zitnick, C.L., Kang, T., Seitz, S.: Gaze awareness for video-conferencing: a software approach. In: IEEE MultiMedia **7**(4),2000.
25. Giger, D., Bazin, J-C, Kuster, C., Popa, T., Gross, M.: Gaze correction with a single webcam. In: 2014 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2014.
26. Venolia, G., Tang, J., Cervantes, R., Bly, S., Robertson, G., Lee, B., Inkpen, K.: Embodied social proxy: mediating interpersonal connection in hub-and-satellite teams. In: CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1049–1058, ACM New York, 2010.
27. Kendon, A.: Some functions of gaze-direction in social interaction. In: Acta Psychologica **26**, 1967.
28. Brennan, S. E., Chen, X., Dickinson, C.A., Neider, M.B., Zelinsky, G.J.: Coordinating cognition: the costs and benefits of shared gaze during collaborative search. In: Cognition **106**(3),2008.
29. Akkil, D., Isokoski, P.: I See What You See: Gaze Awareness in Mobile Video Collaboration. In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, ETRA '18, p. 32, ACM New York, 2018.
30. Yao, N., Brewer, J., D'Angelo, S., Horn, M., Gergle, D.: Visualizing Gaze Information from Multiple Students to Support Remote Instruction. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, p. LBW051, ACM New York, 2018.
31. Avellino, I., Fleury, C. and Beaudouin-Lafon, M.: Accuracy of Deictic Gestures to Support Telepresence on Wall-sized Displays. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 2393–2396, ACM New York, 2015.

32. Monk, A.F., Gale, C.: A look is worth a thousand words: Full gaze awareness in video-mediated conversation. In:Discourse Processes **33**(4), pp. 257-278, 2002.
33. Hart, S.G. and Staveland, L.E: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Advances in psychology **52**, pp. 139-183, 1998.
34. Cisco TelePresence MX Series, https://www.cisco.com/c/en/us/products/collaboration-endpoints/telepresence-mx-series/index.html. Last accessed 26 Jan 2019
35. RealPresence Group Series, http://www.polycom.com/products-services/hd-telepresence-video-conferencing/realpresence-room/realpresence-group-series.html. Last accessed 26 Jan 2019
36. Cisco CTS-SX80-IPST60-K9 TelePresence (CTS-SX80-IPST60-K9), https://www.bechtle.com/ch-en/shop/cisco-cts-sx80-ipst60-k9-telepresence–896450-40–p. Last accessed 26 Jan 2019
37. Enterprise Video Conference, http://www.avsolutions.com/enterprise-video-conference. Last accessed 26 Jan 2019
38. Caine, K.: Local standards for sample size at CHI. In: Proceedings of the 2016 CHI conference on human factors in computing systems, pp. 981–992, ACM New York, 2016.