Enhancing Online Meeting Experience through Shared Gaze-Attention

Chandan Kumar
chandan.kumar@iao.fraunhofer.de
Human-Technology Interaction
Fraunhofer Institute for Industrial Engineering IAO
Stuttgart, Germany

Bhupender Kumar Saini
bhupender.kumar.saini@iao.fraunhofer.de
Institute of Human Factors and Technology Management
University of Stuttgart
Stuttgart, Germany

Steffen Staab
s.r.staab@soton.ac.uk
University of Stuttgart
Stuttgart, Germany
University of Southampton
Southampton, UK

ABSTRACT

Eye contact represents a fundamental element of human social interactions, providing essential non-verbal signals. Traditionally, it has played a crucial role in fostering social bonds during in-person gatherings. However, in the realm of virtual and online meetings, the capacity for meaningful eye contact is often compromised by the limitations of the platforms we use. In response to this challenge, we present an application framework that leverages webcams to detect and share eye gaze attention among participants. Through the framework, we organized 13 group meetings involving a total of 43 participants. The results highlight that the inclusion of gaze attention can enrich interactive experiences and elevate engagement levels in online meetings. Additionally, our evaluation of two levels of gaze sharing schemes indicates that users predominantly favor viewing gaze attention directed toward themselves, as opposed to visualizing detailed attention, which tends to lead to distraction and information overload.

CCS CONCEPTS

• **Human-centered computing** → *User studies*; **User centered design**.

KEYWORDS

eye contact, gaze attention, virtual meeting, social interaction, collaboration

ACM Reference Format:
Chandan Kumar, Bhupender Kumar Saini, and Steffen Staab. 2024. Enhancing Online Meeting Experience through Shared Gaze-Attention. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3613905.3651068

1 INTRODUCTION

In recent years, online meeting systems like Zoom and Microsoft Teams have gained widespread popularity and usage. While remote meetings offer advantages such as time savings and a reduced carbon footprint associated with physical travel [21], they also come with several limitations in terms of user experience and engagement when compared to in-person meetings [11].

The effectiveness and overall productivity of any group meeting highly relies on how participants manage their attention and engage in the discourse. In face-to-face interactions, participants enjoy the advantage of easily assessing one another's participation status, aided by gaze cues about whom each participant is focusing on, which significantly contributes to this understanding. However, in online meetings, these crucial cues are not available to participants. Moreover, in traditional face-to-face meetings, participants have the ability to perceive each other's attention and interest, which often leads to spontaneous side conversations and plays a pivotal role in nurturing social connections. Online meetings fail to establishing such meaningful social connections. Another advantage that presenters or speakers possess in face-to-face meetings, is the capability to adapt their content or pitch based on the level of engagement displayed by their audience. They can gauge the reactions and non-verbal cues of listeners, allowing for real-time adjustments. Conversely, online discussions frequently leave presenters feeling distracted or disengaged when they lack visible signs of attention or feedback from fellow participants. This disconnect between presenters and their audience can impede the overall effectiveness of online meetings.

Towards this phenomena, various methods have been suggested to enhance communication by enabling better eye contact and sharing gaze information during remote collaborations [18]. However, previous studies in this area have primarily concentrated on representing relative positions and adjusting gaze directions with additional hardware to simulate eye contact among participants [12, 18]. Some research has explored the significance of joint attention in collaborative work settings, employing commercial eye tracking devices [6, 13, 19]. In comparison, our objective is to investigate the acceptability and practicality of incorporating gaze information in widespread multi-party online meetings conducted on platforms like Skype, Zoom, WebEx, and Microsoft Teams, all of which typically utilize laptops and webcams. Regarding acceptability, we aim to determine whether incorporating gaze information can enhance the online meeting experience for end-users by adding value, increasing engagement, and promoting social interaction, or if it tends to be distracting. Furthermore, our objective is to evaluate the optimal manner of conveying gaze information among participants through two distinct approaches. The first method involves complete visibility of gaze information among all participants (full visibility mode), enabling individuals to observe who is looking at whom. In contrast, the second method confines gaze information

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3651068

exclusively to the individuals themselves (direct visibility mode), allowing them to perceive who is looking at them.

We have developed a web-based extension for online meetings that can introduce gaze information through webcam streams. Using the extension framework, we have conducted a study encompassing 13 online group meetings involving a total of 43 participants. In this paper, we report the feasibility and reception of visualizing eye contact in virtual meetings. Specifically, we investigate whether eye contact enhances the virtual meeting experience and assess the suitable methods for visualizing eye contact in such settings.

2 RELATED WORK

The significance of eye gaze information has garnered considerable attention from researchers in psychology and affective computing [1, 10]. More specifically, in physical settings, the importance of eye contact and attention has been investigated from various perspectives, including turn-taking, social roles, and engagement within group conversations [14, 15]. These studies typically utilized head-mounted eye trackers or readily available cameras to track eye and head movements. The outcomes of these investigations underscore the crucial role of eye contact in promoting natural organization and nurturing social connections during group discussions in face-to-face settings. Consequently, this emphasis on the significance of eye contact in physical meetings has stimulated research into its role within virtual meetings. Below, we outline the primary focus of prior research in this area.

2.1 Representation of Eye Contact in virtual meetings

Considerable research has focused on gaze correction methods in video conferencing, facilitating individuals in maintaining eye contact with their fellow participants [18]. True-view [22] aims to provide accurate views for two meeting participants and create the illusion of close proximity by using two cameras, one on the left and one on the right, to generate a virtual camera view in the center. MMSpace [16] offers realistic social telepresence in small group conversations via "kinetic display avatars" that mirror remote users' head movements. TeleHuman [12] introduces a cylindrical display with 6 Kinects and a 3D projector. These methods primarily rely on hardware solutions with complex setups to facilitate eye contact. In contrast, there are also 2D video-based software approaches that utilize computer vision techniques to calculate pupil positions and apply image operations to adjust the pupil's appearance, making it seem like individuals are looking directly into the camera [17, 23]. More recently, commercial solutions like NVIDIA Maxine Eye Contact[1] have emerged. Our work, on the other hand, focuses on conveying users' genuine intentions, translating their attention and eye contact when they are looking at another participant's video frame.

[1]https://developer.nvidia.com/blog/improve-human-connection-in-video-conferences-with-nvidia-maxine-eye-contact

2.2 Implicit Gaze analysis in virtual meetings

Numerous studies have leveraged implicit gaze information to gain insights into people's behavior during virtual meetings. For instance, Isaacs and Tang [11] conducted a study to discern the disparities between in-person and virtual meetings. In virtual meetings, participants typically used verbal cues like addressing each other by name and explicitly requesting individuals to speak, whereas in physical meetings, eye gaze was frequently employed to indicate the recipient of their communication. Vertegaal et. al [20] discovered that the absence of eye contact led to 88% of participants experiencing difficulties in identifying the target of others' conversations. George et. al [7] investigated what users actually focus on during online video meetings.While these studies analyzed gaze information and yielded valuable social insights, they did not asses the acceptance and feasibility of enabling explicit gaze information in online meetings.

2.3 Explicit Gaze sharing in virtual meetings

There is existing research focused on sharing explicit gaze information among participants to evaluate its advantages in collaboration and learning contexts. For instance, D'Angelo and Gergle [6] used remote eye tracking to share gaze information, investigating how remote pairs utilize graphical representations of each other's eye gaze during tightly-coupled collaborative tasks. Burch [2] employed mobile eye tracking devices to continuously monitor students' eye movements as they paid visual attention to lecture slides. Langner et al. [13] conducted studies with student groups working remotely on course assignments, utilizing Tobii eye trackers to explore how eye-based joint attention enhances efficient collaboration. These studies primarily employed commercial eye trackers to detect gaze information within specific task environments. In contrast, our research aims to assess the impact of general-purpose webcam-based gaze information in everyday online group conversations.

In addition to gaze awareness, which is the central focus of this paper, prior studies have delved into examining the effects and utilization of various other modalities within the realm of online meetings. These include investigations into the role of spatial cues [3], which encompass the positioning of individuals within virtual spaces to convey information or establish social dynamics. Proximity cues [9] have also been scrutinized, exploring how the perception of personal space and distance influences communication and interaction in virtual environments. Furthermore, head-turning cues [4] have been studied to understand how subtle movements of the head can signal attention or engagement, contributing to the overall dynamics of group meetings.

3 GAZE-ENABLED ONLINE MEETING FRAMEWORK

The proposed gaze-enabled meeting framework is based on extension of the Open Source Jitsi Meet Web platform by integrating webcam-based gaze detection and visualization. As shown in Figure 1, the proposed meeting platform retains the familiar layout found in popular online meeting platforms, with video frames distributed proportionally among participants. Gaze attention is denoted by bubbles located in the top right corner of each participant's video frame, while eye gaze calibration and the option to select between

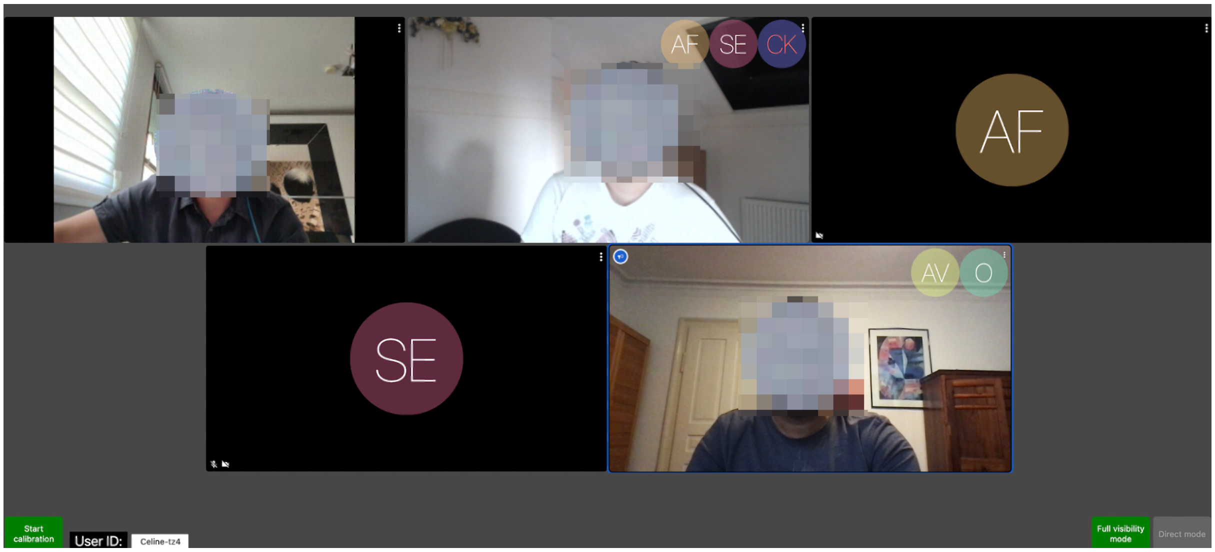

Figure 1: This screenshot illustrates a session of the gaze-enabled online meeting framework featuring five participants (in Full Visibility mode). Each participant is depicted within a rectangular frame, with attention bubbles displayed in the top right corner of each frame, indicating the attention they are receiving from other participants. Notably, the participant positioned in the top middle frame is currently being observed by three other participants, while the participant in the bottom right frame is the focus of attention for two other participants.

two proposed visualization modes are available in the bottom layout. In the following sections, we provide insights into the visual design (gaze attention representation) and functional design (gaze tracking and sharing) choices.

3.1 Visual Design

To determine the most suitable approach for integrating gaze attention information, we initiated an ideation process comprising two focus group discussions. In these discussions, participants were presented with a hypothetical scenario wherein gaze information, indicating who is looking at whom in a virtual meeting, was available. The aim was to gather insights on how people preferred to visualize this information.

The initial focus group, composed of 5 university students, and the subsequent group, comprising 8 university employees, both brainstormed ideas aimed at enhancing video conference platforms. Suggestions ranged from implementing arrows between video frames to adjusting frame sizes, incorporating color coding, and introducing virtual avatars. However, ideas involving explicit visualizations such as avatars, 3D images, rotating videos [8, 19] were dismissed due to their potential to disrupt the current layout and user experience of online video platforms, to which users have become accustomed.

The preferred design concept centered around embedding circular bubbles within video frames. This choice was made to maintain the familiar layout of traditional meeting platforms, as users were already familiar with such bubbles, commonly used to indicate attention in platforms like Overleaf and Google Docs. Our user-centered design approach meticulously refined visualization elements, addressing concerns such as size, placement, and animation of gaze bubbles. Initially, real-time gaze tracking resulted in distracting icon movements, prompting the introduction of a fading effect to ensure a smoother user experience. When it came to visualizing gaze with bubbles, two options were considered: using colored dots or letter abbreviations derived from usernames.. Despite the simplicity of colored dots, feedback and our preference favored letter abbreviations as color dots are hard to distinguish especially in larger meetings.

3.2 Functional Design

The gaze-enabled framework was implemented as a web extension, ensuring easy integration into users' browsers. We selected Jitsi Meet as our video conferencing platform due to its resemblance to commercial remote meeting software in design and functionality, coupled with the advantage of open-source flexibility for design adjustments and data analysis. Gaze tracking was achieved through the integration of the WebGazer library [2]. In Jitsi meetings, visualizations were created by overlaying HTML elements using JavaScript. The browser extension is designed for adaptability in virtual conference meetings on the Jitsi Meet platform and requires installation in the web browser. The recorded gaze data is transmitted to a NodeJS server for processing and conversion into gaze bubble visualization data. This server is hosted on a cloud provider's instance. Subsequently, the converted visualization data is sent back to the browser extension and directly displayed within the online meeting.

On the server side, we performed additional processing of WebGazer data to enhance gaze prediction accuracy. We implemented

[2]https://webgazer.cs.brown.edu

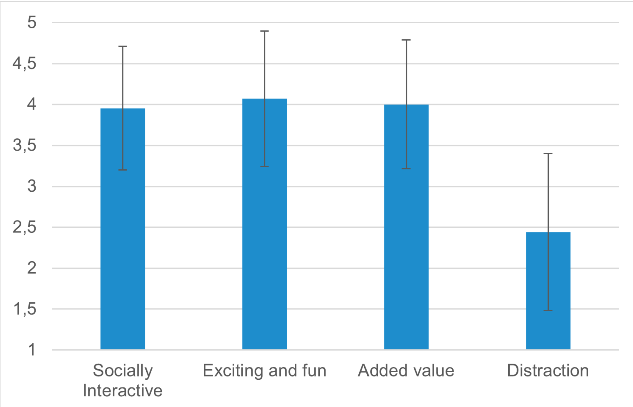

Figure 2: Participant feedback for the proposed gaze-enabled framework with inherent comparisons with conventional virtual meeting. Errors bars indicate Standard Deviation

WebGazer's Kalman filter and further improved precision by averaging the last 20 WebGazer predictions. The purpose of this smoothing filter is to minimize unwanted jittering effects. The filter computes an average of the most recent data points, running alongside the current eye gaze as a secondary reference. Consequently, significant deviations are mitigated as much as possible, resulting in a noticeably smoother user experience with the framework. We also conducted checks to ensure that predicted gaze coordinates fell within specific HTML elements designated for each user (video frame associated with a participant). Attention was realized by gaze fixation for at least one second, adhering to the standard dwell time criteria established in eye-tracking literature*[5]. The calibration of the eye tracker was designed to be user-friendly, involving intuitive clicks at various points on the screen. Initially, these calibration points are positioned at the screen corners and center on the calibration screen, and subsequently on individual windows corresponding to the participants during the meeting.

4 EVALUATION STUDY

The study's objective is to evaluate the acceptance and feasibility of remote meetings enhanced with gaze-based attention information, and compare the two modes of attention visualization. To achieve this, we invited multiple users to use the proposed framework in remote group conversation scenario and gather their subjective feedback through online questionnaire.

4.1 Participants

A total number of 43 participants (32 males and 11 females; aged 22 to 55, mean = 30.2, *SD* = 7.5) took part in the study. 15 participants were university students, 12 researchers, 8 industry employees, and 8 from other occupations (nurse, teachers etc.). All the participants were competent in using computers and other online services such as virtual meeting applications. 30 participants mentioned that physical meetings are their preferred type of meeting, while the remaining 13 said they prefer virtual meetings. We also asked what's their main reason or use case for using virtual meetings. In this regard, 23 of the participants mentioned their main use case for work (business, conferences, etc.), 14 mainly use it for school (lecture, exercises, etc.), and 6 of the participants mainly use virtual meeting for personal needs (meeting friends, family etc.). The participants were not paid and did not receive any other compensation.

4.2 Procedure

The entire study was conducted online. We scheduled the group meeting using a standard online calendar. Once the schedule was confirmed, we sent the participants an online study details link, which included brief information about the study and requested their informed consent for data recording and analysis. Additionally, it provided the Jitsi meeting link for them to join at the scheduled time slot. Within the study details link, once participants consented to the study's terms, they were directed to the next page to download the extension. This page included a set of instructions, including an installation guide, a brief pictorial tutorial on the calibration process, and information on the two visualization modes. Before joining the meeting, participants were required to calibrate the webcam-based eye tracking system. The meeting host welcomed everyone and spent the initial 5 minutes addressing any technical queries related to the framework from participants. Afterward, the meeting followed its agenda, which varied from a regular research group meeting to student project discussions or casual social gatherings. During the meeting, participants had the option to re-calibrate if they believed that improved accuracy was necessary. Each participant received an anonymous user ID, which they used to complete a post-meeting questionnaire. These user IDs were unique to each participant, ensuring consistency even if they dropped out and rejoined the meeting.

After the meeting they received an online questionnaire link consisted several parts: Demographic details and preferences including age, gender, and general preferences regarding physical versus virtual meetings, and the frequency of webcam usage; Framework assessment where participants provided feedback on the comfort and accuracy of using eye tracking during the meeting. Additionally, they assessed the optional feature of providing attention information without turning on the camera; Visualization mode comparison where participants compared their experience using full visibility versus direct mode visualizations; Option for additional feedback in textual form was also provided.

4.3 Results

In total 13 group meetings took place using the framework with a minimum of 3 and maximum of 7 participants in a single meeting (including the host who did not participate in the questionnaire), *i. e.*, average of 4.3 person per meeting. The total meeting duration was 8 hours, *i. e.*, average of 36 minutes per meeting. After attending these meetings, all the participants expressed their opinion through questionnaire, which we report in the following:

4.3.1 Gaze calibration and accuracy. First, we asked the participants their feedback on the integrated calibration process in the framework, and the perceived accuracy of eye tracking. The calibration process was rated by participants with an average score of 2.59, *SD* = 1.2 (on the Likert scale from 1 = very easy to 5 = very complicated). Participants rated the accuracy of eye gaze with an average score of 2.79, *SD* = 1.11 (1 = very accurate to 5 = not accurate

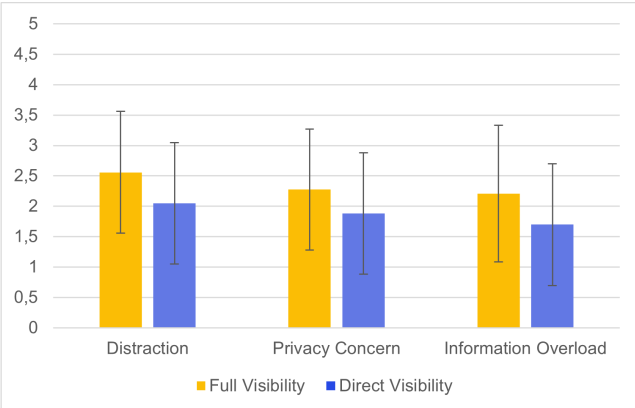

Figure 3: Participant feedback while comparing the two modes of attention visualization. Errors bars indicate Standard Deviation

at all). Some participants commented – *"Calibration not easy in bad light conditions" "tracking sometimes "drifted" off to person below" "Had some accuray / calibration issues but after the third calibration the accuracy was fine"*. In summary, we found that the calibration and accuracy of the system depended on individual participant characteristics such as seating position, camera quality, lighting conditions, and more. Consequently, the perception of accuracy varied among participants, as evidenced by the significant standard deviation in the results. However, despite this variability, the scores we recorded are promising, especially considering the limitations of webcam-based tracking in current eye-tracking research. Looking ahead, we anticipate that advancements in camera technology and innovative tracking methods will greatly broaden the scope of our research and its practical applications.

4.3.2 Assessment and acceptance of gaze-enabled meeting framework. The proposed framework enables eye contact information compared to conventional virtual meeting platform, however, it also introduces additional overhead of visual information processing for its participants. We wanted to assess how users perceive the usefulness of eye contact information, therefore we asked them *"In comparison to virtual meetings with no eye-contact, if the framework made the meeting more socially interactive/exciting and fun/has added value"* (1 = strongly disagree to 5 = strongly agree). We also asked if the framework distracted them from the actual meeting. The results of this assessment are shown in Figure 2. It is evident from the results that participants found the framework interesting, and agreed that enabling gaze aids value to the meetings by making it more exciting, and socially interactive. At the same time, it is important to know that participants did not find the additional information distracting them from actual meeting severely. The positive assessment also reflected in the overall acceptance of proposed framework, *i. e.*, to the question *"If possible: would you use the framework for future meetings?"*. 31 out of the 43 participants said they would like to use it in future meetings, only 1 participant said she would not, while the rest 11 were neutral. In the pilot study we observed that adjusting calibration and accuracy can be sometimes demanding or erroneous for participants, and hence they judge the concept of enabling eye contact in future meetings chiefly based on the tracking accuracy. Therefore to further understand the reasoning behind their non-acceptance, all the participants who selected the options of "No" or "I would not care" had an additional pop-up question *"if the accuracy of eye tracking is improved in future, would they like to use the framework in future meetings"*. 46% off the participants who were initially neutral about using the framework in future with the answer "I would not care", changed their decision to "Yes" showcasing that they like the concept but they were mainly hesitant due to the accuracy of eye tracking.

4.3.3 Sharing attention without switching-on video feed. In many situations, individuals attending a meeting may either be unable or unwilling to activate their cameras. Reflecting this, in the demographic questionnaire we have asked the participants *how often do you use your webcam in virtual meetings*, to which we received average response of 3.09, $SD = 1.19$ (1 = never to 5 = very often). The neutral score indicates that it is highly subjective to people preferences and context. Nevertheless, an absence of video feed highly depreciates the social context for meeting participants, both as a presenter/speaker who does not know if people are paying attention, and as a participant who does not have any other source to show their attention and active participation. In this regard, we asked the participants (i) *"Being a presenter/speaker - Would you like to have attention information from the participants whose camera are switched off"*, and (ii) *"Being a participant - In situations where you are uncomfortable in switching on the camera, you would have no problem in sharing the gaze attention information"*. In response to (i), most participants agree that they would like to get the attention information and provided an average score of 4.2, $SD = 1$ (on the Likert scale from 1 = strongly disagree to 5 = strongly agree). Answering (ii) average score was 3.8, $SD = 1.2$.

4.3.4 Comparison and acceptance of visualization schemes. We have integrated two different visualization schemes in the framework, *i. e.*, Full Visibility, and Direct Mode. During the meeting, participants were free to switch between these modes as per their convenience and liking. After the meeting, we asked them to weigh the visualizations (by the inherent means of icons frequency and attention information) distracted them from actual meeting, if they felt overwhelmed to process the information, or with respect to privacy aspect of being watched by others. The findings, depicted in Figure 3, align with our expectations, showing that participants favored the Direct mode concerning distraction, information overload, and privacy concerns. To assess the statistical differences in ordinal Likert Scale responses, we first conducted a Shapiro-Wilk test for normality and subsequently applied the Wilcoxon rank-sum test. The results indicated a significant distinction between Full Visibility and Direct mode for the dependent variables of distraction ($p < .01$), information overload ($p < .01$), and privacy ($p < .001$). Furthermore, upon surveying the participants to ascertain their preferred mode, it was found that approximately 60% favored the direct mode, indicating a clear preference for its straightforward approach. In contrast, 38% of participants opted for the full visibility mode, appreciating its comprehensive display of information, and 2% did not express a preference for either mode. Additional subjective feedback from the participants also reflects the usage scenario of two schemes: *"when I am in business meeting or giving a presentation, I would use Direct Mode as it lets me focus, but at the same time provides me motivation that people are paying attention. However, in social meetings with colleagues, friends, and family I would choose Full visibility scheme to connect with everyone"*. We argue that both schemes have certain pros and cons and end user should have the choice to enable and disable their preferred mode. This could also be driven by the privacy regulations of the particular organization or region, *e. g.*, companies can opt for only Direct mode integration in their virtual meeting platform, or it could also depend on the use case scenario, a moderator of the meeting can have the full visibility but the participants can only see who are looking at them with a direct mode.

5 CONCLUSIONS AND FUTURE WORK

We introduced a framework that allows users to perceive gaze attention in online group meetings using webcam feeds. Our results demonstrated that incorporating gaze attention can enhance interactive experiences and increase engagement levels in online meetings, with users expressing a preference for its integration in online meeting platforms. This framework provide the option of without the need to activate the camera, offering the option to enhance attention and engagement in meeting and lectures, without compromising privacy. We also explored the right level of granularity for conveying gaze attention, and users preferred the scheme of receiving attention information directed at them in terms of distraction, privacy, and information processing. However, this preference is also subject to situations, such as people prefer complete share (everyone can see who is looking at whom) in social meeting scenarios.

We anticipate that advancements in camera technology and tracking algorithms on modern computers will significantly improve gaze attention accuracy. This progress will enable the integration of gaze communication into commercial online meeting platforms and open new research avenues for exploring non-verbal behavior in virtual collaborative environments. Our future plans involve incorporating additional non-verbal signals, including emotion and gesture recognition, into the framework. Emotion recognition has already been integrated, and we are planning a study involving teachers being informed about shared gaze and emotion status in a classroom setting.

ACKNOWLEDGMENTS

We extend our sincere gratitude to Oliver Ferch, Alexander Fischer, Jan Fuhrmann, and Alexander Völker from the University of Stuttgart for their invaluable contributions towards the development of the prototype for the proposed framework.

REFERENCES

[1] Janet Beavin Bavelas, Linda Coates, and Trudy Johnson. 2002. Listener responses as a collaborative process: The role of gaze. *Journal of communication* 52, 3 (2002), 566–580.

[2] Michael Burch. 2023. Gaze-Based Monitoring in the Classroom. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*. 1–3.

[3] Bill Buxton. 2009. Mediaspace–meaningspace–meetingspace. In *Media space 20+ years of mediated life*. Springer, 217–231.

[4] William AS Buxton, Abigail J Sellen, and Michael C Sheasby. 1997. Interfaces for multiparty videoconferences. *Video-mediated communication* (1997), 385–400.

[5] Benjamin T Carter and Steven G Luke. 2020. Best practices in eye tracking research. *International Journal of Psychophysiology* 155 (2020), 49–62.

[6] Sarah D'Angelo and Darren Gergle. 2016. Gazed and confused: Understanding and designing shared gaze for remote collaboration. In *Proceedings of the 2016 chi conference on human factors in computing systems*. 2492–2496.

[7] J George, A Mirsadikov, M Nabors, and Kent Marett. 2022. What do users actually look at during videoconference calls? Exploratory research on attention, distraction effects and gender. In *Proceedings of the 55th Hawaii international conference on system sciences*. 4779–4787.

[8] Zhenyi He, Keru Wang, Brandon Yushan Feng, Ruofei Du, and Ken Perlin. 2021. Gazechat: Enhancing virtual conferences with gaze-aware 3d photos. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 769–782.

[9] Erzhen Hu, Jens Emil Sloth Grønbæk, Austin Houck, and Seongkook Heo. 2023. OpenMic: Utilizing Proxemic Metaphors for Conversational Floor Transitions in Multiparty Video Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[10] Michael Xuelin Huang, Jiajia Li, Grace Ngai, and Hong Va Leong. 2016. Stressclick: Sensing stress from gaze-click patterns. In *Proceedings of the 24th ACM international conference on Multimedia*. 1395–1404.

[11] Ellen A Isaacs and John C Tang. 1994. What video can and cannot do for collaboration: a case study. *Multimedia systems* 2, 2 (1994), 63–73.

[12] Kibum Kim, John Bolton, Audrey Girouard, Jeremy Cooperstock, and Roel Vertegaal. 2012. Telehuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2531–2540.

[13] Moritz Langner, Peyman Toreini, and Alexander Maedche. 2022. EyeMeet: A Joint Attention Support System for Remote Meetings. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.

[14] Philipp Müller, Michael Xuelin Huang, Xucong Zhang, and Andreas Bulling. 2018. Robust eye contact detection in natural multi-person interactions using gaze and speaking behaviour. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. 1–10.

[15] Catharine Oertel and Giampiero Salvi. 2013. A gaze-based method for relating group involvement to individual engagement in multimodal multiparty dialogue. In *Proceedings of the 15th ACM on International conference on multimodal interaction*. 99–106.

[16] Kazuhiro Otsuka, Shiro Kumano, Ryo Ishii, Maja Zbogar, and Junji Yamato. 2013. Mm+ space: nx 4 degree-of-freedom kinetic display for recreating multiparty conversation spaces. In *Proceedings of the 15th ACM on International conference on multimodal interaction*. 389–396.

[17] Yalun Qin, Kuo-Chin Lien, Matthew Turk, and Tobias Höllerer. 2015. Eye gaze correction with a single webcam based on eye-replacement. In *Advances in Visual Computing: 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, December 14-16, 2015, Proceedings, Part I 11*. Springer, 599–609.

[18] Holger Regenbrecht and Tobias Langlotz. 2015. Mutual gaze support in videoconferencing reviewed. *Communications of the Association for Information Systems* 37, 1 (2015), 45.

[19] Roel Vertegaal. 1999. The GAZE groupware system: mediating joint attention in multiparty communication and collaboration. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 294–301.

[20] Roel Vertegaal, Gerrit Van der Veer, and Harro Vons. 2000. Effects of gaze on multiparty mediated communication. In *Graphics interface*. 95–102.

[21] Juncheng Wu, Anushka Rajesh, Yu-Ning Huang, Karishma Chhugani, Rajesh Acharya, Kerui Peng, Ruth D Johnson, Andrada Fiscutean, Carla Daniela Robles-Espinoza, Francisco M De La Vega, et al. 2022. Virtual meetings promise to eliminate geographical and administrative barriers and increase accessibility, diversity and inclusivity. *Nature Biotechnology* 40, 1 (2022), 133–137.

[22] L-Q Xu, A Loffler, PJ Sheppard, and D Machin. 1999. True-view videoconferencing system through 3-D impression of telepresence. *BT technology journal* 17, 1 (1999), 59–68.

[23] Ben Yip. 2005. Face and eye rectification in video conference using artificial neural network. In *2005 IEEE International Conference on Multimedia and Expo*. IEEE, 690–693.