Effects of Shared Gaze on Audio- Versus Text-Based Remote Collaborations

GRETE HELENA KÜTT, Pomona College
TEERAPAUN TANPRASERT, Pomona College
JAY RODOLITZ, Pomona College
BERNARDO MOYZA, Pomona College
SAMUEL SO, Pomona College
GEORGIA KENDEROVA, Pomona College
ALEXANDRA PAPOUTSAKI, Pomona College

Remote collaborations are becoming ubiquitous, but, despite their many advantages, face unique challenges compared to collocated collaborations. Visualizing the collaborator's point of gaze on a shared screen has been explored as a promising way to alleviate some of these limitations by increasing shared awareness. However, prior studies on shared gaze have not considered the medium of communication and have only studied its effect on audio. This paper presents a study that compares the effects of shared gaze on collaboration performance during audio- and text-based communication. We find that for text, shared gaze improved task correctness and led collaborators to look at and talk more about shared content. Similar trends are found for gaze-augmented voice communication, but contrary to the slower performance in text, it also saw improvements in completion time as well as in cognitive workload. Our findings demonstrate the differences in how shared gaze impacts audio- versus text-based communication and highlight the need to further understand the nuances of the medium of communication when designing novel tools to support remote collaborators.

CCS Concepts: • **Human-centered computing → Collaborative and social computing systems and tools**; *Computer supported cooperative work*;

Additional Key Words and Phrases: Communication medium, eye tracking, shared gaze, remote collaboration

ACM Reference Format:
Grete Helena Kütt, Teerapaun Tanprasert, Jay Rodolitz, Bernardo Moyza, Samuel So, Georgia Kenderova, and Alexandra Papoutsaki. 2020. Effects of Shared Gaze on Audio- Versus Text-Based Remote Collaborations. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 136 (October 2020), 25 pages. https://doi.org/10.1145/3415207

1 INTRODUCTION

Today, remote collaboration is an essential part of the professional landscape. Working from distance has enabled companies and institutions to provide flexibility, accommodate different lifestyles, combine resources, and reduce expenses [26, 30]. Following the outbreak of the COVID-19 global

Authors' addresses: Grete Helena Kütt, Pomona College, gkva2015@mymail.pomona.edu; Teerapaun Tanprasert, Pomona College, ttaa2017@mymail.pomona.edu; Jay Rodolitz, Pomona College, jbra2017@mymail.pomona.edu; Bernardo Moyza, Pomona College, bdma2016@mymail.pomona.edu; Samuel So, Pomona College, ssah2018@mymail.pomona.edu; Georgia Kenderova, Pomona College, gkab2017@mymail.pomona.edu; Alexandra Papoutsaki, Pomona College, apaa2017@pomona.edu.

2573-0142/2020/10-ART136 $15.00
https://doi.org/10.1145/3415207

pandemic, remote work became overnight a crucial measure of public safety. By April 2020, 34.1% of US employees who used to commute started working from home [6] and virtually all academic institutions replaced in-person teaching by online sessions [3]. Even before COVID-19, a global survey of 15,000 employees showed that 50% of respondents work outside their main office at least 2.5 days per week [30]. This dramatic shift in the modern workplace has been supported by technological innovations that have enabled new ways of communicating during remote problem solving. Synchronous media, such as instant messaging and voice- and video-conferencing, are some of the many possibilities for remote collaborators to connect from virtually anywhere in the world and work simultaneously on the same problem. Nevertheless, remote collaborators face many unique challenges, such as lack of common ground [13], co-reference, implicit cues [56], and task awareness [39], that can severely impact the quality and outcome of their collaboration.

Recently, dual eye tracking technology has been explored as a novel way of enhancing remote synchronous communication [35]. Through shared gaze, the process where two collaborators can see each other's point of gaze visualized on their individual screens in real time, remote collaborators in diverse settings have seen improvements in the quality and outcome of their interactions [5, 16, 40, 64, 66]. For example, shared gaze has led to better collaboration performance as well as increased levels of mutual understanding and joint attention [5, 16, 40]. However, to our knowledge, real-time spatial-based gaze sharing has only been studied in settings with audio- or gaze-only communication, ignoring the potential benefits that collaborators engaging in instant messaging can reap during text-based communication. We believe that this is an important omission as text-based media are widely used in personal, educational, and professional circles [37, 38, 53]. For example, Slack has emerged as a popular instant messaging platform used by individuals and companies overseeing projects with multiple components and collaborators [37]. Therefore, chat is an important medium in remote group collaboration [50] and in environments where voice or video communication is not suitable, such as in shared households, open offices or in locations with lower Internet bandwidth. In addition, there is evidence that when remote collaborators communicate via text and their eye movements on shared content deviate, they are likely to misunderstand each other [55]. Hence, shared gaze has the potential to augment existing communication channels in collaborative applications like Google Docs, Online Microsoft Word, and Overleaf that have already integrated chats to allow for synchronous instant messaging between collaborators.

In this paper, we explore how the combination of traditional media (i.e. voice-conferencing and instant messaging) and the visualization of shared gaze influences collaboration performance. As a first step toward designing better tools to support gaze-augmented collaborations, we study how the presence of shared gaze impacts text-based versus audio-based remote collaborations. Following the long line of research on the differences between voice- and text-based communication [33, 45, 46], we approach this question through the lens of media synchronicity theory [22] and conduct a within-subjects experiment with 24 pairs that share parts of the contents in their individual screens. Similarly to Löber et al. [45], the pairs have to solve a mystery by transmitting critical information about suspects and reach a consensus, based on their mutual information and shared understanding. This experiment simulates tasks of high uncertainty [23, 45], where collaborators have both shared and unique pieces of information that are critical for the successful solution to a problem. For example, this task mirrors how a remote hiring committee might reach a conclusion after each member has completed an individual interview with the candidate as they simultaneously examine their application. In our setting, there is a unique correct solution to the problem, although one could argue that this can be less important in different settings, as long as both collaborators are satisfied with the conclusion they have reached.

Our results show that during text-based communication, remote pairs more often reach the correct conclusion and also look at and refer to the same parts of their screen in the presence of shared gaze. The same findings hold true for voice-based communication which additionally benefits from shared gaze in terms of shorter task completion time and reduced cognitive workload. Post-study interviews provide further explanations and interpretations of these findings.

Our research highlights how the effects of shared gaze on collaboration performance depend on whether it takes place in the context of voice or text-based communication. Our findings support the need for further investigation of the effect of the medium of communication when considering the design of novel interactive systems that can support remote collaborations.

2 RELATED WORK

Although critical to today's globally distributed workforce [30], remote collaborators still face significant challenges [56]. Identifying means of supporting situation awareness and conversational grounding in distributed settings has become one of the focal points in research on dual eye tracking. More specifically, shared gaze, enabled by dual eye tracking technology, has shown promise as a means to alleviate some of the limitations found in synchronous remote problem solving [16, 36, 40].

2.1 Gaze Awareness in Remote Collaboration

Compared to face-to-face settings, remote collaboration lacks some important characteristics, for example, task awareness [39] and conversational grounding [12, 29]. To overcome these limitations, researchers have proposed a new communication signal in the form of shared gaze. Shared gaze awareness, which refers to remote collaborators being able to see each other's point of gaze on their screen in real time, has been studied in contexts as diverse as collaborative writing, pair programming, visual search, education, and gaming [16, 40, 47, 61, 64, 66].

There are several ways in which real-time visualizations of shared gaze can benefit remote collaboration. Kütt et al. [40] showed that when incorporated into online collaborative writing, shared gaze increased levels of collaboration, produced tighter work coupling, and improved perceived collaboration quality. In pair programming tasks, D'Angelo and Begel [16] and Nüssli [54] demonstrated that shared gaze enabled pairs to refer to and find locations in shared code more effectively while strengthening shared awareness. Shared gaze has also been shown to be beneficial when applied to education, with potential uses for both students and teachers. For example, Schneider and Pea [61] showed that real-time gaze awareness improved students' quality of collaboration and learning gains, while Yao et al. [68] found that it helped teachers more effectively guide students through remote instruction. Additionally, both cooperative and competitive games have been the focus of many gaze sharing studies. In cooperative games, Maurer et al. [47] and Špakov et al. [66] discovered that the incorporation of shared gaze resulted in increased mutual understanding between teammates and overall increased team performance. On the other hand, Newn et al. [51, 52] showed that in competitive games, shared gaze visualizations increased the players' ability to infer the intent of their opponents. In both game types, shared gaze was shown to have a positive effect on perceived social presence in multiple studies [41–43, 47].

Previous studies on real-time shared gaze visualization have explored its effects on collaboration in several contexts and settings: gaze alone [5, 64], gaze with voice-based communication [16, 40], and gaze with voice-based communication in a virtual reality environment [66]. However, to the extent of our knowledge, no study has looked at the effect of shared gaze combined with a text-based medium, such as instant messaging, versus an audio-based medium, such as voice-conferencing. This is a significant oversight, as text-based media are widely used in personal and professional circles [37, 38, 53] and could benefit by improvements brought by shared gaze. This gap has also been acknowledged by recent research on shared gaze on collaborative writing, where participants mentioned the potential usefulness of gaze visualization to non-audio communication [40]. This paper builds upon these findings by investigating the effect of shared gaze on voice- versus text-based remote collaborations to better understand their suitability for remote problem solving.

2.2 Media Theories and Communication

Previous research has shown that communication media can help or hinder collaboration [7]. Synchronous media, such as instant messaging and audio- and video-conferencing, among others, have shown a great potential for improvement of effectiveness in communication and increase in production in remote teams [25, 34]. On the other hand, computer-mediated communication also carries potential risks of misunderstanding, distrust, and poor decision making, if used with disregard to different goals and tasks [8]. Thus, the capabilities of a medium are crucial for the communication performance between distant collaborators and the success of their task.

To investigate the effect of different media on the process of organizational communication, researchers have proposed a multitude of theories, some of which focus on media and information richness [15], media synchronicity [22], social influence [27], social presence [4, 59, 63], channel expansion [9], and conversational grounding [12]. Our work is inspired by Media Synchronicity Theory [22], which grew from one of the earliest media theories, Media Richness Theory [15]. While theories such as social presence [41–43, 47] and grounding theory [2, 18] have motivated scholarly work on certain aspects of shared gaze awareness, Media Synchronicity Theory has not appeared in gaze sharing research thus far. Given that it has been widely employed as a framework in studies comparing audio- and text-based media performance [33, 45, 46], it is highly relevant to this paper's focus on the effect of the real-time gaze visualization on the performance of remote collaborations supported by different communication media.

Media Synchronicity Theory (MST) reflects on modern, computer-mediated communication and aims to explain the capabilities of a communication medium to "support individuals working together at the same time with a shared pattern of coordinated behavior" [24]. This capability is known as *media synchronicity* and is distinguished from synchronous communication, where individuals simply communicate simultaneously. Dennis et al. claim that the following five properties of a medium influence its synchronicity [23]:

- Transmission velocity: the speed at which a medium can deliver a message.
- Parallelism: the ability of a medium to transmit multiple simultaneous messages.
- Symbol set: the different ways a medium allows information to be encoded.
- Rehearsability: the degree to which a medium allows for messages to be planned and edited.
- Reprocessability: the ability to reprocess a message once it has been received and decoded.

A higher transmission velocity and a more natural symbol set increase a medium's synchronicity, while parallelism, rehearsability, and reprocessability decrease its synchronicity.

In previous research, MST has served as a framework for examining the difference between commonplace voice- and text-based communication media for single tasks. According to the theory, audio communication is characterized by higher synchronicity, whereas text-based communication, such as email and chat, by lower synchronicity. Overall, the theory has found support in empirical studies. For example, Hassel et al. [33] compared the communication methods of email, voice, and a single email followed by voice in a hidden profiles task. They concluded that teams using media with lower synchronicity (i.e. email) outperformed teams using media with higher synchronicity (i.e. audio), while the best results were achieved with the use of a portfolio of media. However, a study by Löber and Schwabe [45, 46] found that groups perform faster with audio than text. This was true even for mystery-solving games which require an exchange of information without a need for explanation (i.e. task of high uncertainty), where MST posits that text would be better suited. They explain this discrepancy as a function of the transmission velocity of the medium—audio transmits information more quickly, which increases productivity. At the same time, their results show that text leads to communication with higher density of critical information.

Due to its novelty, shared gaze is an unexplored medium when combined with text communication within the context of media theories. In this paper, we extend the prior work to better understand the effect of shared gaze awareness on traditional communication media. While some studies treat shared gaze as an information channel on its own [5, 64], we consider real-time gaze visualization as a supplementary signal accompanying an existing medium, following the trend of most gaze sharing studies [1, 16, 40]. Many of these studies have regarded shared gaze as a way to approximate natural eye contact in remote settings. This comes in accordance with MST which considers face-to-face communication as having the highest synchronicity because of its abundant symbol set [22]. Since the spatial-based gaze visualization represents a visual cue that can be used to encode a message, its addition to a commonplace communication channel, according to MST, leads to an expansion of the symbol set. By definition, the added visual cue grants immediate feedback but cannot be seen as rehearsable or reprocessable in our real-time setting.

3 USER STUDY

In our work, we investigate how remote collaboration performance is affected by traditional voice- and text-based media when coupled with shared gaze. We conducted a within-subjects study where pairs of collaborators played a mystery-solving game under four conditions: i) ***voice***, which corresponds to a remote setting where collaborators communicate via voice-conferencing while sharing part of their screens, ii) ***voice+gaze***, where the same voice-based communication is extended to include shared gaze, iii) ***text***, where communication happens through synchronous instant messaging, and iv) ***text+gaze***, where the text-based communication now also includes the shared gaze of the collaborator. Past studies have only compared voice versus text (e.g., [33, 45, 46]) or voice versus voice+gaze (e.g., [16, 40]). We chose to present participants with all four options for a more thorough comparison. These counterbalanced conditions were randomized, with each ordering occurring once, resulting in 24 unique permutations. For the remainder of the paper, we will often refer to these four conditions as the four media, although the combination of a voice- or text-based medium with shared gaze has not been formally recognized as a unique medium in literature before. Our reasoning is that adding the signal of shared gaze could be seen as equivalent to adding video on voice-conferencing which is considered a unique medium.

Participants were asked to solve four murder mysteries by identifying the correct suspect. This is an adaptation of Cruz, Henningsen, and Smith's work [14] which has been applied many times in studies of media synchronicity (e.g., [45, 49]). For each mystery, participants were provided with an image of a crime scene which contained some useful clues, a police report which contained more clues, profiles about each of the three potential suspects, and an optional chatroom, as seen in Figure 1, for the text and text+gaze conditions. Only the image was shared at its entirety by both participants, while the police report, the profiles, and the optional chats were at the same fixed positions for both participants but partially contained different information. The images were adopted from the game "Criminal Case" and were slightly modified to match the cases which were rewritten from scratch for simplicity and to make their difficulty comparable. Each mystery corresponded to a unique medium condition in order to limit the number of participants that would be required if the story was added in the design of the study and needed to be counterbalanced.

The clues were designed in the style of a hidden profiles task, similar to Stasser and Stewart [65] and its adaptation by Hassel and Limayem [33]. Participants were provided 9 clues in the suspect profiles, 6 of which were shared and 3 of which were unique to each participant. For each task, 6 out of the 24 total clues were deemed critical because they are unequivocally indicative of the correct suspect. Participants were made aware that not all of the information needed to solve the

Fig. 1. For each mystery, participants were provided with an image of a crime scene (top left), a police report (top right), and profiles about each of the three potential suspects (bottom left). An optional chat was included in the text and text+gaze conditions (bottom right). A timer reminded participants of the time limit of each task (top right). The shared gaze visualization used in the voice+gaze and text+gaze conditions can be seen in the right portion of the crime scene as a green circle, indicating that the gaze of the two-collaborators overlapped.

mystery was shared between them. The fact that each participant had some unique information in their suspect details and police report required them to share their clues and to discuss suspicions to reach a shared conclusion.

Despite its simple nature, this task models many real-world settings where remote collaborators have some unique and some shared information that is necessary to successfully solve a problem. By requiring both the conveyance of new and the discussion of already known pieces of information, this task allows us to compare the media with and without gaze in terms of how well they enable collaborators to work together at the same time with a common focus [22]. Another advantage of this type of task is that it allows us to focus on problems that have a known correct solution. This differs from prior applications of shared gaze on problems with multiple possible solutions (e.g., summarizing text [40] or debugging code [16] which can be achieved in many ways), therefore it is easier to objectively measure effects on the collaboration performance.

3.1 Gaze-Sharing System

We modeled our gaze-sharing platform after Eye-Write [40] and deployed it on Firepad, an open-source collaborative text editor. The point of gaze of each participant on their respective screen was continuously captured from their individual eye tracker and streamed to their collaborator using Firebase. During a session, a user could see their collaborator's point of gaze but not their own. If their gaze locations overlapped, the visualization turned green, as in [16, 40].

A common question in the literature on shared gaze is the shape of the visualization of the point of gaze [17, 19]. We chose to visualize the gaze of a collaborator as a circle when it hovered over the shared image with the crime scene. The circular visualization has been used before in different settings (e.g., [47, 52]) and it highlights in a discreet but clear manner the location of the collaborator's gaze while allowing them to see what is enclosed underneath. Beyond the crime scene, we were unsure about how to visualize the gaze and considered the following possibilities: as a circle only on the crime scene, as a circle everywhere on the screen, or as a circle on the crime scene and a rectangular border around the suspect profiles, police report, and chat. Normally, gaze sharing is done consistently on the whole screen, but in our hidden profiles setting, each screen shared only partial information as each user had some unique information in the suspect profiles and police reports and only the crime scene was shared in its entirety.

We conducted two rounds of iterative pilot testing to receive feedback on these aspects of the visualization of the shared gaze. First, we formed two pairs by pairing two authors with two recruited participants. Each pair was given the chance to test each option, then picked one to solve a task with. In both cases, the recruited participants chose to visualize the shared gaze as a circle continuously and regardless of the screen location. For the visualization where the circle appeared only on the crime scene, they cited being confused about it disappearing when their collaborator looked outside of it. For the one where the police report, suspect profiles, and chat would be highlighted with a rectangular border around them, they found the transition between different shapes and the sudden "snap" to be both jarring and confusing. A second question we asked was whether we would create a "shadow" effect around the circle to increase the likelihood of it being perceived, especially in the case of non-overlapping gaze when its color would be off-white. The same participants claimed that they preferred a mild shadow effect, as seen in Figure 1.

After these two pilot tests, we recruited and paired two new participants and tested the entire experiment. During it, we found that in the text and text+gaze conditions, the participants tried to copy and paste text from the police reports and suspect profiles. In the final form of our gaze-sharing platform, we deactivated the copy/paste option.

3.2 Apparatus

In order to simulate a remote collaborative environment, each participant was placed facing opposite directions of the study room and was given a Mpow noise-cancelling headset with a microphone. Each participant conducted the experiment on a Windows 10 PC with a Samsung SyncMaster 2233 monitor (1920 px × 1080 px) and a Tobii 4C eye tracker with a sampling rate of 90 Hz. Two Canon Vixia HF R800 camcorders were used to record the study, each being placed near a participant. The synchronous instant messaging for the text and text+gaze conditions was achieved through a chat that was implemented using Firechat and Firebase and the voice-conferencing for voice and voice+gaze was conducted via Slack. Screen recordings were captured with OBS Studio.

3.3 Participants

Following the guidelines and approval of our Institutional Review Board, 48 participants were recruited via local mailing lists that mostly consist of undergraduate students. The participants were split into 24 pairs that were selected based on scheduling availability. It was made clear to participants when scheduling, and confirmed before the experiment, that they could only be paired with a partner whom they were not familiar with in advance in order to avoid any effects on the outcome of the collaboration from prior familiarity among the pairs. Twenty-eight participants identified as women, 19 as men, and 1 as non-binary. All participants were between the ages of 18 and 24 years old apart from one who was in the range of 25 and 29 years old. Each pair was exposed to a unique ordering of the four media (voice, voice+gaze, text, and text+gaze).

3.4 Procedure

Since participants not knowing each other was a requirement of the study, they were introduced prior to commencing the experiment. Once introduced, they were seated at their respective work-stations with their backs facing each other for the duration of the study. Participants were then given consent forms. Upon consent, they were tasked with completing the Tobii Eye Tracking Core Software calibration process and wearing their headphones. An experimenter activated all recording and logging processes.

Participants were then shown an outline of the experiment's structure and expectations. Following this, they completed a practice task before beginning each mystery task. The practice task consisted of a collaborative environment structurally identical to the one used in the actual tasks, but filled with placeholder information. The communication medium corresponding to the task was also present. Participants were given one minute to complete each practice task.

After the practice task was completed, participants moved on to the task's preparatory section. During it, they were presented with the information needed to solve the crime (i.e. the crime scene, police report, and suspects profiles) and were given time to read these before commencing the task. This process was repeated for all four conditions. In all four, participants were free to establish their own strategy of how they would communicate their given information and decide on the suspect that committed the crime within 10 minutes.

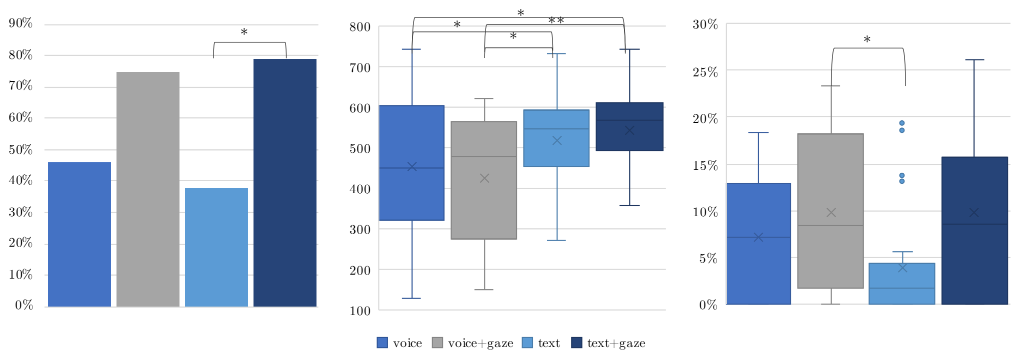
After each task, participants completed a series of individual questionnaires to gauge their experience with the medium. The first of these was adopted from NASA's Task Load Index (NASA-TLX) [32], which assessed the participant's perceived cognitive workload, on a scale of 0−100. We chose to focus on four of the six subscales of NASA-TLX, namely the mental demand, performance, effort, and frustration, ignoring the physical and temporal demand questions due to the nature of our experiment. The NASA-TLX was followed by a set of questions that asked participants to self-evaluate their collaboration with each other based on five questions adopted by the rating scheme by Meier, Spada, and Rummel [48]. At the end of voice+gaze and text+gaze, the "Perceived Usefulness and Ease of Use Survey" [21] was presented, asking participants about the perceived usefulness of the shared gaze visualization. After all four tasks were completed, the experiment was ended with a brief semi-structured interview in which the participants discussed the study with the experimenter. Each participant received a compensation of 20 USD.

4 RESULTS

We collected data on the following objective and subjective measures of collaboration performance: correctness, completion time, references to the shared image, critical clues transmitted, gaze overlap, cognitive workload, perceived quality of collaboration, and utility of shared gaze visualization, along with post-experiment responses. Figures 2, 3, and 4 capture the objective measures of collaboration performance, while Table 1 and Figure 5 illustrate the subjective measures. Eye tracking data were missing from one of the pairs for the text condition due to the failure of the logging server. Statements and interpretations from the post-study interviews of 46 participants are included as the interview from one pair was not properly recorded.

For each valid pair, a transcript for each of the four tasks was created and transcribed. The transcripts for voice and voice+gaze tasks were created using the online audio transcription software Sonix. The transcripts for text and text+gaze tasks were created using the chat logs recorded through Firechat on Firebase. For the voice and voice+gaze conditions, the time duration was calculated through Sonix. For the text and text+gaze conditions, we calculated the duration of the task as the difference between the timestamp of the last and first message. In addition, for each utterance/message, two authors simultaneously confirmed the transcripts and labeled whether the pair referred to the shared image that corresponds to the crime scene.

For the majority of the analysis reported below, we perform omnibus tests with repeated measures across the four different media: voice, voice+gaze, text, and text+gaze. For each statistical test, the assumptions have been tested (e.g., normality with Shapiro-Wilk's tests and sphericity with Mauchly's test for one-way repeated measures ANOVA) and, if violated, a correction (e.g.,



(a) Correctness

(b) Completion Time (sec)

(c) Image Referral Proportion

Fig. 2. Correctness (a), completion time (sec) (b), and image referral proportion (c) across the four conditions. Statistically significant differences are indicated for pairwise comparisons (* for $p < .05$, ** for $p < .01$, and *** for $p < .001$). Shared gaze leads to more correct answers and higher referrals on the shared image for both voice- and text-based communication. Completion time is shorter for voice and longer for text-based communication, with shared gaze having the opposite effect on the two.

Greenhouse-Geisser) or a non-parametric equivalent test (e.g., Friedman's test) has been conducted instead, as appropriate. The appropriate post hoc analysis with Bonferroni correction follows to identify the pairs of conditions that led to statistically significant differences in the measures.

4.1 Objective Measures of Collaboration Performance

4.1.1 Correctness. Correctness was measured based on a pair's correct identification of the suspect for each mystery. Four pairs disagreed in their responses; we labeled these as incorrect. As shown in Figure 2a, the percentages of pairs that correctly identified the suspect were 45.8%, 75.0%, 37.5%, and 79.2% for voice, voice+gaze, text, and text+gaze, respectively. Cochran's Q test determined that these percentages were statistically significantly different, $\chi^2(3) = 12.634$, $p = .005$. Identifying the suspect correctly was found to statistically significantly increase from text to text+gaze. The addition of shared gaze increased the correctness of responses both for voice- and text-based communication, with text decisively improving in the presence of shared gaze. This is justified by MST, confirming our expectation that the expansion of the symbol set by the addition of visual cues increases the synchronicity of the medium and therefore leads to improved information processing [22]. Interestingly, collaborators achieved the highest levels of correctness during text+gaze (79.2%), which outperformed even voice+gaze communication. While the improvement in correctness from text-only to voice-only communication is in accordance with previous research which showed the significant advantage of the transmission velocity of audio for a similar mystery task [45], the overall higher correctness of the text+gaze medium is a surprising finding. This suggests that the combination of text and gaze worked well in the novel communication context. We hypothesize that while the gaze signal could have helped with the collaborators' unfamiliarity with each other as well as with maintaining shared focus, the transmission and analysis of large amounts of information benefited from the more deliberate and reprocessable text medium, leading to a more correct outcome. In this way, text+gaze seems to strike a balance by combining the capabilities of the traditional text and the slightly higher synchronicity enabled by shared gaze.

4.1.2 Completion Time. Figure 2b shows the completion times across the four conditions. We observe that pairs spent more time on their tasks during text-based compared to voice-based conditions. A one-way repeated measures ANOVA determined there were statistically significant differences, $F(3, 69) = 10.435$, $p < .001$, partial $\eta^2 = 0.312$, with statistically significant increases from voice to text, voice to text+gaze, voice+gaze to text, and voice+gaze to text+gaze. Most communication theories, including MST, would support these findings as voice-based communication supports higher synchronicity due to its higher transmission velocity. The effect of shared gaze was more subtle here, reducing the total time for voice but increasing it for text. Prior research has been inconclusive on the effect of shared gaze on completion time for audio [16, 40]. Although not statistically significant, the increase we observe in the text+gaze condition could be justified by the additional cognitive workload that incurs by the shared gaze. Supported by previous findings, the decrease from voice to voice+gaze could be explained by more efficient referencing [16] and increased mutual understanding [40] enabled by the shared gaze.

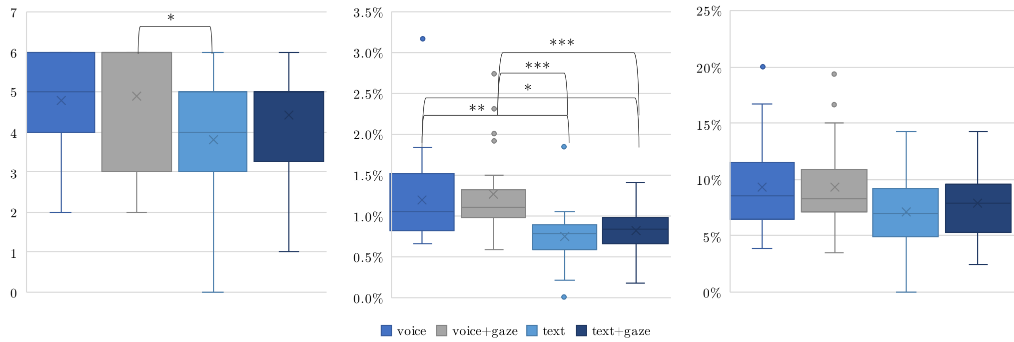
4.1.3 Image Referral Proportion. While reviewing the transcripts of all pairs, we noticed that some of the exchanges made explicit use of the shared image that corresponded to the crime scene. We analyzed the image referral proportions as the duration of utterances that referenced the image normalized over completion time for voice and voice+gaze and the number of messages that referenced the image normalized over the total messages sent for text and text+gaze. A one-way repeated measures ANOVA revealed statistically significant differences across the four conditions, $F(3, 69) = 4.689$, $p = .005$, partial $\eta^2 = .195$. The proportions can be seen in Figure 2c, with the increase from text to voice+gaze being statistically significant. Although not statistically significant, it is interesting to note the increase during the text+gaze condition, where participants referred to the shared image of the crime scene in 9.74% of the messages exchanged in contrast to 3.85% of the messages in text condition. These data suggest that the presence of shared gaze during voice, and particularly during text communication leads pairs to look more at shared parts of their screens, as further elaborated in Section 4.1.5, and therefore, are more likely to refer to them in their exchanges.

4.1.4 Critical Clues. Inspired by Löber et al. [45], we next focus on the critical clues that were transmitted across the four different conditions. We analyze the raw number of critical clues, the normalized proportion over the total duration (critical clue rate), and the normalized proportion over all information pieces sent (ratio of critical information). For text and text+gaze, the number of information pieces is the number of messages while for voice and voice+gaze, we consider the number of information pieces as the total number of turns taken.

Figure 3a shows that the pairs exchanged more critical clues during voice-based conditions. A Friedman test determined there were differences in the raw number of critical clues transmitted across the four conditions, $\chi^2(3) = 11.656$, $p = .009$, with statistically significant differences between voice+gaze and text.

When analyzing the critical clue rate (critical clues normalized by task duration), a Friedman test showed that there were statistically significant differences across the four different conditions, $\chi^2(3) = 30.600$, $p < .001$, with voice and text, voice and text+gaze, voice+gaze and text, and voice+gaze and text+gaze being statistically significantly different, as seen in Figure 3b. Contrary to [45], we did not find a statistically significant difference for the ratio of critical information (critical clues normalized over total number of information pieces), as seen in Figure 3c ($\chi^2(3) = 7.569$, $p = .056$).

Knowing that audio is faster than text-based communication [45] and that shared gaze enables faster referencing [16], we would expect pairs to transmit more critical information when communicating over a medium of higher synchronicity. Although shared gaze did not significantly increase the critical clue rate, we do see, as expected, an overall higher rate of critical information transmitted during voice and voice+gaze. This finding agrees with [45], emphasizing the importance



(a) Critical Clues (0–6) (b) Critical Clue Rate (clues/sec) (c) Ratio of Critical Information

Fig. 3. Critical clues (a), critical clue rate (b), and ratio of critical information (c) across the four conditions. Statistically significant differences are indicated for pairwise comparisons (* for $p < .05$, ** for $p < .01$, and *** for $p < .001$). Shared gaze leads to more exchanges of critical clues for both voice- and text-based communication.

of the transmission velocity of the medium which allows collaborators to correctly identify the critical information under time pressure. Interestingly, while shared gaze led to a decrease in the critical clue rate for voice, it benefited text-based communication. This could suggest that although the combination of text and gaze caused participants to take longer to complete the task, they also transferred more critical clues over that time, which would explain the higher scores on correctness.

4.1.5 Gaze Overlap Proportion. The final objective measure of collaboration performance focuses on eye tracking data. Throughout the experiment, we recorded the gaze locations of the pairs whenever they overlapped within one of the following areas of interest (AOIs), as seen in Figure 1: *image* that corresponds to the picture of the crime scene (top-left), *report* which is the area where the police report appeared (top-right), *suspects* that matches the profiles about each of the three potential suspects (bottom-left), and finally *chat* for the text-based conditions (bottom right).

We considered the duration of the overlapping gaze locations for each of the above AOIs. We grouped them together if they overlapped for at least 150 ms and unique records were separated by at least 100 ms in order to avoid accidental gaze overlaps and address the noise by the eye trackers. We then calculated the total gaze overlaps per AOI and per pair and normalized it as a proportion over the total duration of the task.

For the image, report, and suspects AOIs, we conducted a one-way repeated measures ANOVA test to determine whether there were statistically significant differences in the gaze overlap durations on the image across the different media. Figure 4 illustrates the proportions of gaze overlap across these three AOIs for each medium. There were no statistically significant differences in the gaze overlap proportions on the **chat**, $t(23) = 0.312$, $p = .758$, $d = 0.02$. This is a surprising finding, indicating that the gaze representation does not affect the proportion of time spent on looking at the chat box together. It shows that despite the fact that typically, participant's gaze visualization mostly stays within the chat box while typing, their collaborator's gaze is not drawn to the chat box more often than without the presence of the partner's gaze. This could suggest that participants are not distracted easily by the shared gaze during the text condition. A possible interpretation is that our setup is similar to social media platforms such as Facebook where users are comfortable with switching their attention between the chat box and the browsing content.

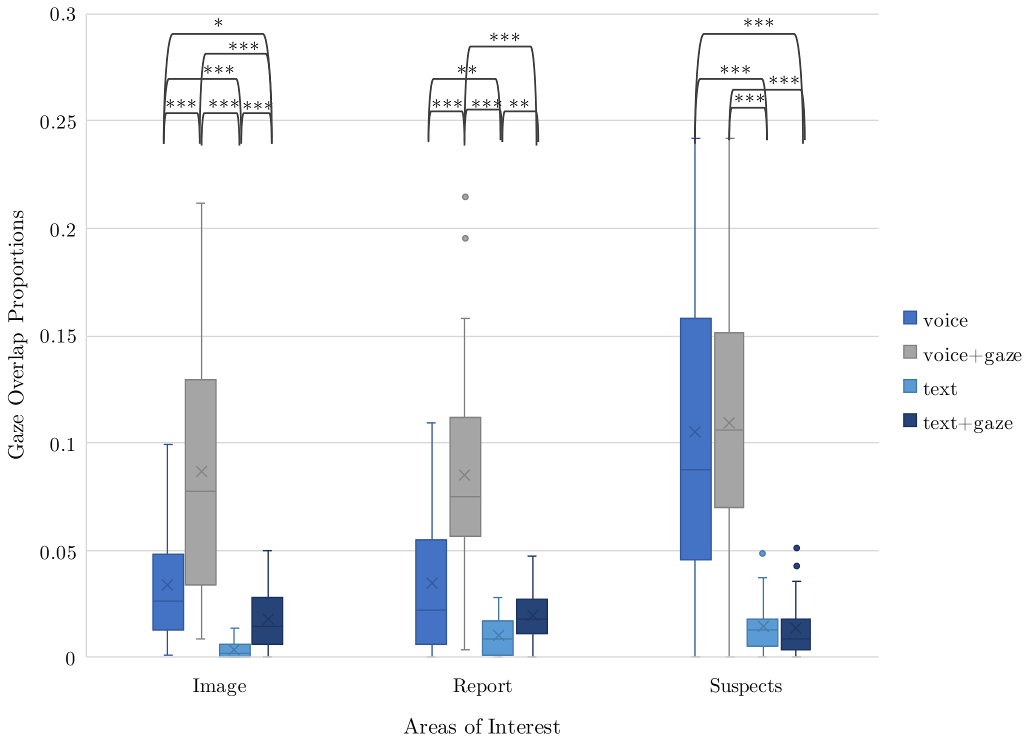


Fig. 4. The normalized gaze overlap proportions for the three AOIs: image, report, and suspects, across each of the four media. Statistically significant differences in scores are indicated for pairwise comparisons (* for $p < .05$, ** for $p < .01$, and *** for $p < .001$). Participants looked at the same AOIs for longer during voice-based communication. Shared gaze led to increases in gaze overlap both for voice- and text-based communication.

For the gaze overlap proportion on the **image**, the medium elicited statistically significant differences on the gaze overlap of pairs, $F(1.402, 32.246) = 33.910$, $p < .001$, partial $\eta^2 = 0.596$, with statistically significant increases from voice to voice+gaze, and text to text+gaze, and statistically significant decreases from voice to text, voice to text+gaze, voice+gaze to text, and voice+gaze to text+gaze. Similarly, for the **report** $F(1.456, 33.497) = 33.145$, $p < .001$, partial $\eta^2 = 0.590$, the gaze overlap statistically significantly increased from voice to voice+gaze, and from text to text+gaze, and statistically significantly decreased from voice to text, voice+gaze to text, and voice+gaze to text+gaze. Finally, for the gaze overlap proportion on the **suspects**, $F(2.007, 46.171) = 38.988$, $p < .001$, partial $\eta^2 = 0.629$, the gaze overlap statistically significantly decreased from voice to text, voice to text+gaze, voice+gaze to text, and voice+gaze to text+gaze.

For each of the AOIs, we observe that shared gaze is more likely to lead pairs to look at the same area of their screens. This happens regardless of whether the entire content is identical (as in image) or not (as in the report). The increase in gaze overlap in voice-based communication was also observed by D' Angelo and Begel in a completely-shared setting [16]. Our data suggest that this phenomenon also applies to text-based communication. This is an important finding since prior studies have shown that simultaneously attending to the same object or information can lead to better performance [11, 28, 55]. It also agrees with our findings on improvements in correctness. Furthermore, the significant increase of overlap on the report in both voice+gaze and text+gaze

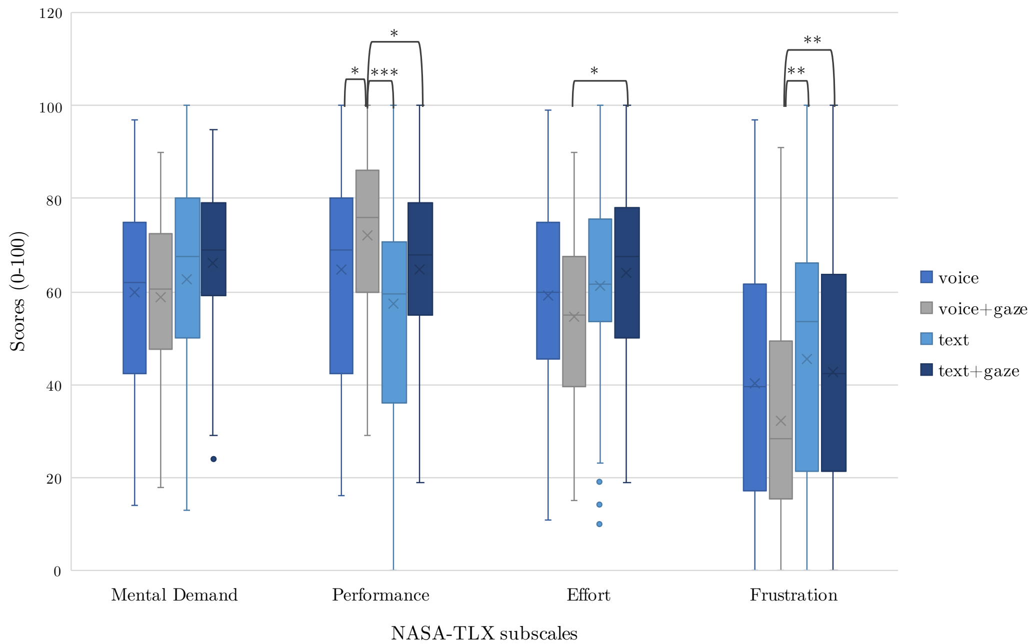


Fig. 5. The four subjective subscales of the NASA-TLX questionnaire that were asked at the end of each task to capture the cognitive workload. Statistically significant differences in scores are indicated for pairwise comparisons (* for $p < .05$, ** for $p < .01$, and *** for $p < .001$). Voice leads to less cognitive workload than text communication. Shared gaze further improves voice while having the opposite effect on text.

condition indicates that collaborators tended to attend to the same part of the layout despite them being aware that their content was not necessarily identical. While conducting the experiments, we observed that multiple pairs' strategies included reading out loud (for voice+gaze) or typing out (for text+gaze) the report, to find out where their information diverged. Thus, we believe that looking simultaneously at the same area enabled them to discover differences in the pieces of information and consequently, fill in the gaps in their understanding of the problem.

4.2 Subjective Measures of Collaboration Performance

4.2.1 Cognitive Workload. Next, we analyze the differences in the distributions of scores assigned by each individual participant during the NASA-TLX questionnaire and across the four conditions. The four questions focused on mental demand, performance, effort, and frustration, resulting in a "raw TLX" questionnaire and an analysis on the individual components rather than a weighted and cumulative score on cognitive workload [31]. For each of the four subscales, we conducted a Friedman's test followed by a post hoc analysis with a Bonferroni correction for pairwise comparisons. Figure 5 summarizes the scores on the four subjective subscales across each of the four conditions and illustrates pairwise statistically significant differences.

The reduction that shared gaze brought in **mental demand** for voice and the increase for text were not statistically significantly different, $\chi^2(3) = 4.620$, $p = .202$.

The medium elicited statistically significant responses in **performance**, $\chi^2(3) = 18.390$, $p < .001$, with statistically significant increases from voice to voice+gaze, text to voice+gaze, and text+gaze to voice+gaze. Shared gaze decisively improved perceived performance for both voice and text.

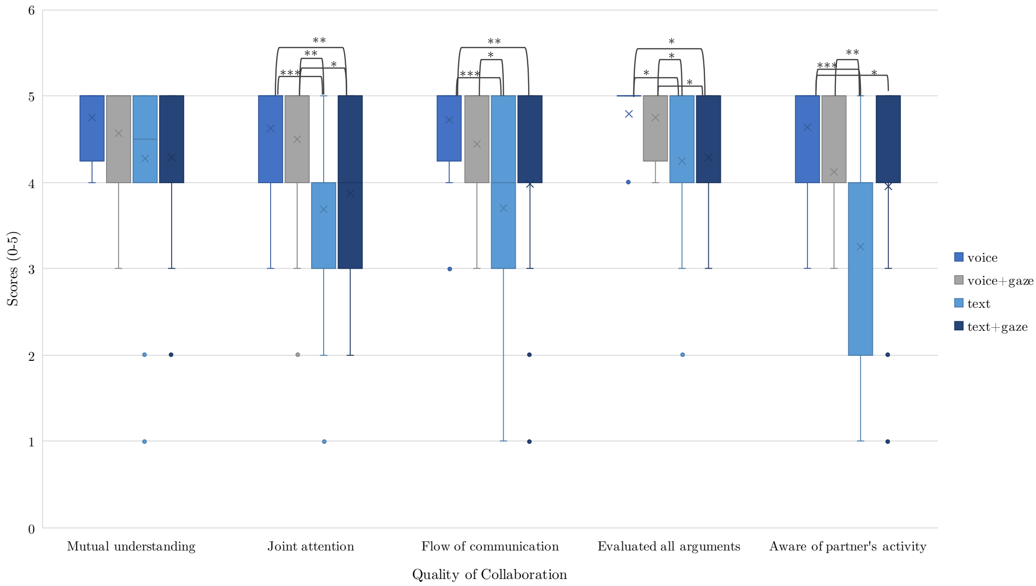


Fig. 6. The five dimensions on the quality of collaboration that were asked at the end of each task along with the mean rank scores (on a 5-point Likert scale). Statistically significant differences in scores are indicated for pairwise comparisons (* for $p < .05$, ** for $p < .01$, and *** for $p < .001$). Voice generally led to higher scores on perceived quality. Shared gaze had a positive effect on joint attention, flow of communication, and awareness of partner's activity for text communication.

Similarly, the scores for **effort** were statistically significantly different, $\chi^2(3) = 10.165$, $p = .01$, with a statistically significant increase from voice+gaze to text+gaze. Shared gaze reduced the perceived effort during voice-based communication and increased it during text-based communication, perhaps because participants had to track too much information [62]. This might justify why the scores for **frustration** were also statistically significantly different, $\chi^2(3) = 16.069$, $p < .001$, with statistically significant increases from voice+gaze to text and from voice+gaze to text+gaze.

In accordance with the results on correctness, our data indicate that shared gaze improves subjective scores of performance both for voice and text communication. At the same time, the addition of shared gaze reduced the perceived effort for voice while increasing it for text. Although we would expect that the expansion of the symbol set through shared gaze adds to the mental demand, the data indicate that participants did not perceive such a significant difference. It is worth noting the juxtaposition of the increase in effort and mental demand from text to text+gaze alongside the higher scores of perceived performance. Finally, the significant differences found between voice+gaze and text+gaze agree with MST and prior studies that explain the contrast between voice and text through the former's relatively fewer cognitive demands on a communicator [10].

4.2.2 Quality of Collaboration. At the end of each task, each participant was given five statements of the form "We sustained mutual understanding during the collaboration" that aimed to capture the quality of collaboration they experienced under each medium. Figure 6 summarizes these five statements that were ranked on a 5-point Likert scale. These five statements, a subset of the original nine [48], were found to statistically significantly increase in Eye-Write [40] from the baseline voice-only communication to the voice+gaze conditions. Although the experimenters took standard precautions to avoid the observer-expectancy effect, we note the ceiling effect. For each statement, we conducted a Friedman test to determine if there were any differences in the assigned scores across the four media. Post hoc analysis with a Bonferroni correction followed.

Interestingly, shared gaze brought borderline statistically significant increases to awareness of partner's activity for text-based communication. This shows that some of the findings of Kütt et al. [40] in voice-based environments could extend to text-based communication. Although, in the text+gaze condition, participants are not able to read the chat and utilize the gaze visualization for spatio-visual references at the same time, the increase in awareness of partner's activity highlights the importance of gaze as an awareness tool. In addition, the slight increase in perceived joint attention comes in agreement with the gaze overlap results for image and report, further confirming that the time spent looking at the same area increases for text-based communication. Together with the increase in image referral, we see that indeed the presence of gaze led to more engagement with the area of joint attention. However, we did not find any statistically significant increases from voice to voice+gaze, contrary to the findings of Kütt et al. On the contrary, most scores dropped, although not substantially. This is surprising, as post-study interviews indicated that attitudes toward shared gaze were either positive or neutral during voice-based communication. The difference between text+gaze and voice+gaze data could suggest that for this particular task that requires the conveyance of a great amount of information, voice outperformed text as a baseline medium in terms of perceived quality of collaboration.

4.2.3 Utility of Shared Gaze. At the end of the two tasks that were conducted under the presence of shared gaze (i.e. voice+gaze and text+gaze), we asked 14 statements from the "Perceived Usefulness and Ease of Use Survey" [21] on a 7-point Likert scale. Table 1 summarizes the average scores assigned to each statement for the two media ranked by increasing p-values. The scores were consistently higher for voice+gaze, with seven statements having a statistically significant decrease from voice+gaze to text+gaze according to Wilcoxon signed rank-tests. This comes to agreement with the statistically significantly higher effort and frustration perceived during text+gaze compared to voice+gaze. These results highlight the nuances that shared gaze brings to the baseline medium (i.e. voice and text). Above all, the differences in user attitudes between text+gaze and voice+gaze indicate that even if shared gaze brings objective improvements during text communication, developers of novel tools and interaction techniques that aim to support remote gaze-augmented text-based collaboration need to further understand how to improve user satisfaction.

4.3 Post-study interviews

The post-study interviews, conducted after pairs completed all tasks, reveal additional information on the results of the objective measures and the participants' perceived experience with the four communication media and offer qualitative explanations to some of our quantitative findings.

Between voice and text—with or without the shared gaze—10 pairs stated that communicating with text is more thought-out and takes more time, while voice allows for quicker but also less filtered communication. When asked to select their favorite communication medium out of the four, 22 participants voted for voice+gaze, 3 for voice alone, 17 for voice either with or without shared gaze, and 4 for text—either with or without shared gaze. Regardless of the presence of shared gaze, voice was clearly preferred over text in general. This agrees with our findings on completion time and with MST which would justify favoring voice-based communication due to its higher transmission velocity that leads to higher synchronicity. One participant with a strong preference for voice over text stated that "It's just easier to communicate through audio... The discussions were more clear." This seemed to override any advantages of text-based communication on rehearsability and reprocessability and could be justified by the type of the tasks, the time limits imposed on them, and the small size of the teams. With respect to shared gaze, the consensus across the three

Perceived Utility Questions on Gaze Visualization	voice+gaze	text+gaze	*p*	
Shared gaze enabled me to accomplish tasks more quickly.	5.08	4.31	< .001	***
Shared gaze increased my productivity.	4.97	4.31	.001	**
Shared gaze improved my performance.	4.89	4.37	.011	*
Shared gaze made it easier to complete the task.	4.93	4.39	.014	*
Shared gaze made the collaboration easier.	5.33	4.77	.014	*
I found shared gaze useful for the task.	5.06	4.62	.020	*
Shared gaze enhanced my effectiveness on the task.	4.81	4.35	.031	*
Learning to operate shared gaze was easy for me.	6.14	6.00	.165	
I found shared gaze easy to use.	5.97	5.81	.199	
I found shared gaze easy to get it to do what I want it to do.	5.89	5.72	.227	
Shared gaze was not distracting.	4.56	4.37	.288	
My interaction with shared gaze was clear and understandable.	5.91	5.75	.365	
I found shared gaze to be flexible to interact with.	5.70	5.60	.433	
It was easy for me to become skillful at using shared gaze.	5.54	5.45	.694	

Table 1. The average scores on the fourteen statements on perceived utility of the shared gaze visualization that were asked during tasks under the voice+gaze and text+gaze conditions ranked by increasing p-value. The asymptotic p-values for the Wilcoxon test statistics are reported (* for $p < .05$, ** for $p < .01$, and *** for $p < .001$). Overall, voice+gaze gathered higher scores than text+gaze.

participants who preferred voice alone was that shared gaze was simply distracting. Only four participants preferred text over voice; their justification was that the clarity and impersonality of the text medium were appropriate for communicating with a stranger. A different attitude toward the medium is possible if the pairs knew each other before the study.

Among those participants who preferred having the additional signal of shared gaze, one commented that "The gaze just speeds up some things more. I didn't find it too distracting... but it was helpful when we were looking at the same thing." They thought that the gaze facilitated a real-time synchronization, so they could focus on the same piece of information, which was helpful especially for this particular type of collaborative task. In the context of MST, that would be justified by the expanded symbol set where shared gaze between remote collaborators could be seen as a possible, albeit limited, alternative for natural eye contact in face-to-face settings [16]. On the other hand, twelve participants stated that while gaze was an interesting feature, it was neither distracting nor helpful. In this, we recognize that pairs only had a limited time to interact with the shared gaze and learn to use it for their benefit. In terms of the general impression toward shared gaze, five pairs had strongly positive feelings, while most felt neutral, saying that its effect was rather subtle: "We didn't use it as efficiently, because we had other forms of communications, but I don't know. I still found it a little bit useful." Only two pairs strongly opined that they found the gaze distracting and unhelpful, both with voice- and text-based communication. This is not surprising and agrees with previous studies, especially [40], which showed the importance of giving the user the choice of deactivating the shared gaze, as it is not a universally-preferred or always-appropriate feature. Overall, the interview responses support our data showing relatively shorter completion time and higher perceived performance for voice and voice+gaze. As we expected, shared gaze was seen mostly as a supplementary signal to voice and text that affected the collaboration process subtly. It is interesting to note, though, that voice+gaze was the most preferred medium among participants, whose comments emphasize the benefit of shared focus.

The interviews also revealed how shared gaze altered communication differently for voice versus text. Ten pairs mentioned that shared gaze was more helpful with text than with voice for various reasons, one of which was a direct result of the text medium's slower transmission velocity. As one participant explained, "We used [gaze] more during the texting [condition][...] In the gaze + text, it was easier to be like 'look at where I'm looking'." Another reason was that shared gaze filled in the gaps of silence during typing. The shared gaze visualization helped participants differentiate between when their partners were typing and when they were looking over the clues or simply thinking, which allowed them to guess when to anticipate messages. This supports our quantitative finding on the increased awareness of partner's activity from text to text+gaze. It also sheds light on how the added gaze signal seems to increase the synchronicity of text, while for voice, with its fundamental advantage of higher transmission velocity [45], the effect is more subtle. However, three participants also shared their struggle with using the gaze visualization while typing: "I definitely didn't use the gaze as much because I was just looking at the keyboard or reading the chat." We expect that there would be a distinction between users who can and who cannot touch type, although this is an ability that we did not record during our experiment.

5 DISCUSSION

Our findings illustrate the importance of taking into account the nuances of communication media when designing novel ways of interaction to support remote collaborators. By comparing shared gaze when added to two primary forms of signals, voice and text, we find differences in objective and subjective measures of collaboration performance. Below, we provide a larger conversation around our findings, the limitations of our study, and the opportunities for future research.

5.1 Findings

During text-based communication, shared gaze had a positive effect on most objective measures such as correctness, gaze overlap and image referral proportion. Similar improvements in objective measures during voice-based communication agree with prior studies on shared gaze [16, 61] and can all be justified by an increase in the synchronicity of the medium brought by the expanded symbol set [22]. In addition, subjective measures such as perceived joint attention, flow of communication, and awareness of the partner's activity slightly increased during text+gaze. However, shared gaze increased the completion time and negatively impacted the subjective attitudes of users compared to improvements brought by shared gaze for voice-based collaboration.

From these results, we see noticeable differences in how shared gaze affects voice versus text. While almost every objective metric improved from voice to voice+gaze, the addition of shared gaze to text increased both correctness scores and completion times. This interesting finding demonstrates that shared gaze induced qualitative changes in the text-based collaboration process rather than simply facilitating it. One example of such change could be seen in the increased awareness of the partner's activity. We hypothesize this is due to voice being fundamentally an 'eyes-free' communication channel as opposed to text, which requires visual attention to the chat box. Thus, voice inherently enables relatively flexible 'co-expressive' composite interaction with gaze while participants in the text condition cannot read the chat and simultaneously address spatio-visual references with the shared gaze.

One of the strongest motivations behind this work is our expectation that shared gaze is beneficial for collaborators' mutual awareness in the overlooked text-based remote collaborations. When looking at the exchanges between participants during the text+gaze condition, we find that joint attention is indeed increased. First, in text+gaze, participants refer to the shared image of the crime scene in 9.74% of the messages exchanged in contrast to 3.85% of the messages in text. Below, direct quotes extracted from messages exchanged during the text+gaze demonstrate that the participants are not only referring to the shared content but are also aware of what their collaborator is doing:

- P1-A: "i found new phone. see my gaze on the left".
  P1-B: "i see it"
- P12-B: "do you see by the crib it says bad mother?"
  P12-A: "ohh nice yeah"
- P29-A: "look at where I'm looking"
  P29-A: "do you see the words bad mother next to the crib"
  P29-B: "Go on it again"
  P29-A: "I'm on it now"
  P29-B: "I see it"

Similarly, when diving into the post-study interviews, we find examples of the collaborators' awareness of each other's activity when participants explained how they experienced shared gaze in the text+gaze condition. One participant said:

> "I found myself paying more attention to the bubble when it was text... With audio, it was easier to communicate the things we were looking at, it was also a lot quicker. But for text, since it took more time to type out what we wanted to say, I could glance at the bubble to see what the other person was looking at."

Our tasks have been inspired by previous literature [33, 45, 65], however in these studies, participants were given physical copies of the information they needed to perform the task. When transferring them onto a screen, we chose a design that reflects the nature of the partially-overlapping information. In our setup, only the image with the crime scene is perfectly identical between the two collaborators. In contrast, the police report, suspect profiles, and chat in text-based conditions, are placed in the same locations but their content is not identical. An intuitive expectation is that spatial-based gaze representation necessitates the exact matching of user interfaces ("What You See Is What I See" or WYSIWIS) [60]. Thus, as mentioned in the User Study section, we piloted versions with the shared gaze taking different forms or disappearing altogether when transitioning from WYSIWIS content to the rest of the user interface elements. Instead, our conversations with the pilot users swayed us toward choosing a uniform visualisation of the gaze for the entire screen. Interestingly, none of the 48 participants in the user study found that aspect confusing. Instead, they confirmed that they understood that the content was not identical but they still utilized the information of where their collaborator looked. For example, one participant during the post-study interview stated:

> "[Gaze] helps a lot more with the texting. Because texting sometimes takes a long time, and sometimes you're like 'okay, you're not sending me a message so I don't know what you're doing'. But with the gaze, I know like 'okay, she's looking at the report'."

This quote shows that the gaze location in that case is perceived as a tool that enables the user to know where their collaborator's attention is directed. It is interesting that this was found to be particularly useful once again in the text-based condition.

Although we believe that it is important for additional studies to systematically explore the different affordances of shared gaze in WYSIWIS versus partially-shared layouts, we posit that qualitative feedback like the quote above show that there are interesting lessons to be learned from our setup. Specifically, we envision different forms of collaborative software where collaborators have the same active applications but different content. For example, if a user knows that their collaborator is looking at their common email client, they can understand that the contents of the emails will not be identical while still achieving improved awareness about what their collaborator is doing. Online classrooms where the instructor can see the type but not the exact content that students engage with is another example that such a setup can be useful. Nevertheless, it is important to note that our study was limited to a single active window. If we were to transition to setups where only one active application can be seen at at time, it is reasonable to expect that the visualization of the collaborators' gaze could be confusing as it transitions between different windows. Future research needs to tackle these questions across different setups and diverse tasks that correspond to different collaborative contexts.

Bringing us back to our particular study, similarly to hidden-profiles tasks [33], the logic behind mystery puzzles is that pairs that communicate well will choose the correct suspect while pairs that communicate poorly will not. In this way, researchers are able to isolate the effects of media on communication performance. However, success or failure in typical mystery tasks may actually be a result of information recall or group confirmation bias rather than the result of effective or non-effective communication [44]. We tried to avoid this by pairing participants that did not previously know each other. Lack of familiarity between collaborators and the task can alter the way that they solve problems and resolve disagreements. MST posits that this might push them more toward establishing shared interpretations of the problem. This would favor higher synchronicity media [22] and could justify the preference of participants for voice-based communication.

5.2 Limitations

As with all experiments, there were trade-offs that could compromise its validity and our findings. Many stem from the choices we had to make when considering the duration of the experiment, the size of the population that would be required for a meaningful analysis, and the cost of compensating the participants. For example, we imposed a 10 minute limit per condition for practical reasons. For a complex task with a lot of information to be shared, this restricts the free flow of the communication and could lead to rushed decisions. It could also favor a higher-synchronicity medium. Out of the 96 tasks performed by our 24 pairs, there were 19 tasks during which the participants went overtime and 4 for which they eventually failed to come to an agreement. We also acknowledge that our population sample is limited in terms of age due to our recruitment limitations. A future study with no time limit and broader demographics could further clarify and strengthen our findings.

Some participants reported difficulty with touch typing, which potentially impacted their performance during the text-based tasks. We did not distinguish between participants who were able to versus those not able to touch type due to the ambiguous parameters between these two groups. Additionally, only two participants shared that they had a hard time maintaining attention to their screen while typing, during the post-study interviews. Therefore, we are unsure about the real impact of this factor, but we note it as a potential direction for future research.

In an attempt to ensure that the critical clues were not too difficult to identify and that the correct answer was unambiguously derivable with enough information, we did not place many clues on the images of the crimes; they instead assisted with pointing in the right direction. Hence, in terms of the information exchanged, our task can be perceived as mostly limited to reading textual information (rather than interacting with, as in [16, 40]) in the form of the suspect profiles and police reports. Of course, this is different for the chat box. However, five pairs shared in their post-study interviews that they thought shared gaze was more helpful for visual than textual content. With a more visual-centric collaborative task, the effect of shared gaze could potentially be more prominent. We consider our design as being in-between collaborative platforms like Google Docs that favor textual information and Google Slides that include more visual information. In practice, in both of these examples, the shared content is dynamic, with Google Docs being similar to the setup used by Kütt et al [40] in voice-based collaborative writing. We see our simplified static setting as a first step that has allowed us to focus on the differences between the media. Nevertheless, we acknowledge the need for future research that could extend our findings to dynamic settings.

Finally, our work shares the same limitations on scalability with existing work on shared gaze. Most findings on shared gaze so far have been constrained to scenarios with only two collaborators.

Following Siirtola et al. [64], further research is required to understand how our discoveries would apply to multi-user settings where visualizing the gaze of more than two users simultaneously could be detrimental to the quality of collaboration. This could be particularly interesting to study in text-based collaborations as rehearsability and reprocessability of text have been found particularly advantageous for larger teams [34]. This question provides many opportunities for novel research. For example, exploring at-will visualization of the gaze [40] or automated methods that distinguish active roles among collaborators could be promising directions.

5.3 Future Research

In our analysis of eye-tracking data, we solely focused on gaze overlap, the standard measure in gaze sharing studies [16, 40]. However, we note that there are many important lessons to be learned by further diving into the gaze activity of collaborators across the four conditions. For example, the subjective reports on cognitive load could be supplemented with objective measures such as fixations, saccades, and pupil dilation. Scanpaths across the different components of the shared screen could also facilitate a better understanding of the differences between the four media. We propose these exciting questions around gaze activity as future research directions.

Our study examined voice-based and text-based communication separately and in combination with shared gaze. Nevertheless, it is worth considering how we can join the strengths of both media in novel applications of shared gaze. For example, an application could support both voice- and text-based communication along with shared gaze. Participants could mostly use voice when looking at shared information and thus take advantage of the lower mental demand of processing their collaborator's gaze. At the same time, voice-to-text technology could be used to automatically send that information via chat. Collaborators would then have the ability to review and reflect on the exchanged written messages which could lead to more informed decisions and higher performance. Such a scenario would take advantage of both the high transmission velocity of voice and the rehearsability and reprocessability of text-based communication along with all aforementioned improvements that shared gaze brings to both. We also imagine that such an application could contain an offline component of reviewing voice recordings and transcripts of the messages along with the corresponding replays of the gaze activity, similar to [36].

As promising as shared gaze may be, it has so far been explored without taking into account ability-diverse teams and the complex ways they collaborate [20]. Going beyond mainstream collaborative platforms, we consider a few ways that shared gaze in combination with traditional communication media could alleviate some accessibility challenges. For example, individuals with speech impairments could benefit from a platform that supports gaze-augmented text-based communication, as in our text+gaze condition. Nevertheless, shared gaze should not remain accessible only to sighted users. For example, instead of being visualized on the screen, it could be translated into an auditory signal incorporated into screen readers used by blind and low-vision collaborators. We believe that there are a lot of possibilities for novel research around shared gaze that can lead to more inclusive and therefore more productive collaborations.

At the same time, shared gaze does not have to be restricted only to cooperative and work-related settings. For example, competitive games are an area where including the gaze of a team has been shown to allow inference of the team player's attention and intent [51, 52]. We imagine that there are numerous possibilities for novel games that can emerge in the presence of shared gaze. For example, a player could strategically coordinate with their teammates and manipulate where they look in order to deceive their opponents. This also provides ample opportunities for researchers studying human behavior in more traditional psychological and game theoretical settings [67]. Similarly, shared gaze could be applied in mixed reality and collaborative virtual environments, where collaborators might be facing the same scene but from a different perspective [57, 58]. We consider our work as a step toward understanding how we can better connect remote collaborators by taking advantage of the powerful social properties of gaze and acknowledge the large number of opportunities for exciting future research.

6 CONCLUSION

In this paper, we studied the effect of shared gaze on remote collaborations in which text-based communication was examined for the first time. We conducted a within-subjects study with 24 pairs that collaborated over voice- and text-based media with and without the presence of a real-time shared gaze visualization.

This work presents novel findings on the effect of gaze on text-based communication. Adding the real-time visualization of the partner's gaze to chat improved collaboration in terms of task correctness and led participants to talk about and look more often at shared content. In addition, collaborators reported higher levels of awareness of their partner's activity, as well as increased perceived performance, joint attention, and smooth flow of communication. However, shared gaze negatively impacted the effort and mental demand required from participants, who overall preferred gaze-augmented voice-based communication. Contrary to text-based communication, collaboration over audio additionally benefited from shared gaze through reductions in completion time and cognitive workload. The comparison of results between gaze-augmented voice and text demonstrate that in text-based collaboration processes, shared gaze led to qualitative changes, rather than simply facilitating it.

Our study highlights the differences in how shared gaze impacts audio- versus text-based communication. While the results for voice mostly agree with prior literature, our novel findings demonstrate the potential benefits of real-time shared gaze visualization to text-based collaborations. Although chat as a baseline medium could be seen as inferior to voice due to its inherently lower transmission velocity, the notable improvements in several objective and subjective measures show that further research is warranted to better understand the nuances of shared gaze in the context of text-based communication. Thus, we believe this work presents a first step towards adding shared gaze as a feature to novel settings. In addition to implications for accessibility and recreational settings, our findings on gaze-augmented text communication are relevant for larger teams. The lack of familiarity between participants, the type of the task, and the time pressure during the experiment could all be factors that made participants favor higher-synchronicity, voice-based media. This shows that user satisfaction needs to be further investigated to match the objective improvements brought by shared gaze to text-based communication. Ultimately, developers of novel gaze-sharing tools and interaction techniques should take into account the capabilities of the communication medium to better support remote collaborators.

ACKNOWLEDGMENTS

This research is supported by NSF grant IIS-1948517. We would like to thank Lemuel Lan and Matthew Paik for their early literature review of communication theories.

REFERENCES

[1] Deepak Akkil, Jobin Mathew James, Poika Isokoski, and Jari Kangas. 2016. GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1151–1158. https://doi.org/10.1145/2851581.2892459

[2] Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2017. Looking Coordinated: Bidirectional Gaze Mechanisms for Collaborative Interaction with Virtual Characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2571–2582. https://doi.org/10.1145/3025453.3026033

[3] Francisco Jonathan Araújo de Oliveira, Ligia Samara Abrantes de Lima, Pedro Ivo Martins Cidade, Camila Bezerra Nobre, and Modesto Leite Rolim Neto. 2020. Impact Of Sars-Cov-2 And Its Reverberation In Global Higher Education And Mental Health. *Psychiatry Research* 288 (2020), 112977. https://doi.org/10.1016/j.psychres.2020.112977

[4] Frank Biocca, Chad Harms, and Judee K. Burgoon. 2003. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments* 12, 5 (2003), 456–480. https://doi.org/10.1162/105474603322761270 arXiv:https://doi.org/10.1162/105474603322761270

[5] Susan E. Brennan, Xin Chen, Christopher A. Dickinson, Mark B. Neider, and Gregory J. Zelinsky. 2008. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition* 106, 3 (2008), 1465 – 1477. https://doi.org/10.1016/j.cognition.2007.05.012

[6] Erik Brynjolfsson, John Horton, Adam Ozimek, Daniel Rock, Garima Sharma, and Hong Yi Tu Ye. 2020. COVID-19 and Remote Work: An Early Look at US Data. *unpublished* (2020).

[7] Goran Bubaš. 2001. Computer mediated communication theories and phenomena: Factors that influence collaboration over the Internet. In *3rd CARNet users conference, Zagreb, Hungary*. Citeseer.

[8] Judee Burgoon, Joseph Bonito, Bjorn Bengtsson, Artemio Ramirez, Norah Dunbar, and Nathan Miczo. 1999. Testing the Interactivity Model: Communication Processes, Partner Assessments, and the Quality of Collaborative Work. *Journal of Management Information Systems* 16 (12 1999), 33–56. https://doi.org/10.1080/07421222.1999.11518255

[9] John R Carlson and Robert W Zmud. 1999. Channel expansion theory and the experiential nature of media richness perceptions. *Academy of management journal* 42, 2 (1999), 153–170.

[10] Barbara L. Chalfonte, Robert S. Fish, and Robert E. Kraut. 1991. Expressive Richness: A Comparison of Speech and Text as Media for Revision. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '91)*. Association for Computing Machinery, New York, NY, USA, 21–26. https://doi.org/10.1145/108844.108848

[11] Mauro Cherubini, Marc-Antoine Nüssli, and Pierre Dillenbourg. 2008. Deixis and Gaze in Collaborative Work at a Distance (over a Shared Map): A Computational Model to Detect Misunderstandings. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications (ETRA '08)*. ACM, New York, NY, USA, 173–180. https://doi.org/10.1145/1344471.1344515

[12] Herbert H Clark, Susan E Brennan, et al. 1991. Grounding in communication. *Perspectives on socially shared cognition* 13, 1991 (1991), 127–149.

[13] Catherine Durnell Cramton. 2001. The Mutual Knowledge Problem and Its Consequences for Dispersed Collaboration. *Organization Science* 12, 3 (2001), 346–371. https://doi.org/10.1287/orsc.12.3.346.10098 arXiv:https://doi.org/10.1287/orsc.12.3.346.10098

[14] Michael G Cruz, David Dryden Henningsen, and Brian A Smith. 1999. The Impact of Directive Leadership on Group Information Sampling, Decisions, and Perceptions of the Leader. *Communication Research* 26, 3 (1999), 349–369. https://doi.org/10.1177/009365099026003004 arXiv:https://doi.org/10.1177/009365099026003004

[15] Richard L Daft and Robert H Lengel. 1986. Organizational information requirements, media richness and structural design. *Management science* 32, 5 (1986), 554–571.

[16] Sarah D'Angelo and Andrew Begel. 2017. Improving Communication Between Pair Programmers Using Shared Gaze Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6245–6290. https://doi.org/10.1145/3025453.3025573

[17] Sarah D'Angelo, Jeff Brewer, and Darren Gergle. 2019. Iris: A Tool for Designing Contextually Relevant Gaze Visualizations. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications (ETRA '19)*. ACM, New York, NY, USA, Article 79, 5 pages. https://doi.org/10.1145/3317958.3318228

[18] 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 (CHI '16)*. ACM, New York, NY, USA, 2492–2496. https://doi.org/10.1145/2858036.2858499

[19] Sarah D'Angelo and Darren Gergle. 2018. An Eye For Design: Gaze Visualizations for Remote Collaborative Work. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 349, 12 pages. https://doi.org/10.1145/3173574.3173923

[20] Maitraye Das, Darren Gergle, and Anne Marie Piper. 2019. "It Doesn't Win You Friends": Understanding Accessibility in Collaborative Writing for People with Vision Impairments. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 191 (Nov. 2019), 26 pages. https://doi.org/10.1145/3359293

[21] Fred D. Davis. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13, 3 (1989), 319–340.

[22] Alan R. Dennis, Robert M. Fuller, and Joseph S. Valacich. 2008. Media, Tasks, and Communication Processes: A Theory of Media Synchronicity. *MIS Q.* 32, 3 (Sept. 2008), 575–600. http://dl.acm.org/citation.cfm?id=2017388.2017395

[23] Alan R Dennis and Joseph S Valacich. 1999. Rethinking media richness: Towards a theory of media synchronicity. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*. IEEE, 10–pp.

[24] Alan R Dennis, Joseph S Valacich, Cheri Speier, and Michael G Morris. 1998. Beyond media richness: An empirical test of media synchronicity theory. In *Proceedings of the Thirty-First Hawaii International Conference on System Sciences*, Vol. 1. IEEE, 48–57.

[25] Nadeem Ehsan, Ebtisam Mirza, and Muhammad Ahmad. 2008. Impact of computer-mediated communication on virtual Teams' Performance: An empirical study. In *2008 International Symposium on Information Technology*, Vol. 3. 1–8. https://doi.org/10.1109/ITSIM.2008.4632068

[26] Thomas A. Finholt and Gary M. Olson. 1997. From Laboratories to Collaboratories: A New Organizational Form for Scientific Collaboration. *Psychological Science* 8, 1 (1997), 28–36. https://doi.org/10.1111/j.1467-9280.1997.tb00540.x arXiv:https://doi.org/10.1111/j.1467-9280.1997.tb00540.x

[27] Janet Fulk. 1993. Social Construction of Communication Technology. *Academy of Management Journal* 36 (10 1993). https://doi.org/10.2307/256641

[28] Darren Gergle and Alan T. Clark. 2011. See What I'M Saying?: Using Dyadic Mobile Eye Tracking to Study Collaborative Reference. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW '11)*. ACM, New York, NY, USA, 435–444. https://doi.org/10.1145/1958824.1958892

[29] Darren Gergle, Robert E. Kraut, and Susan R. Fussell. 2013. Using visual information for grounding and awareness in collaborative tasks. *Human-Computer Interaction* 28, 1 (1 1 2013), 1–39. https://doi.org/10.1080/07370024.2012.678246

[30] International Workplace Group. 2019. The IWG Global workspace survey. http://assets.regus.com/pdfs/iwg-workplace-survey/iwg-workplace-survey-2019.pdf.

[31] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage Publications Sage CA: Los Angeles, CA, 904–908.

[32] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[33] Martin Hassell and Moez Limayem. 2011. A portfolio of media: Effects of media synchronicity on communication performance. *International Conference on Information Systems 2011* 2 (2011), 1672–1682.

[34] Ben Hogan. 2006. Lessons learned from an extremely distributed project. In *AGILE 2006 (AGILE'06)*. 6 pp.–326. https://doi.org/10.1109/AGILE.2006.37

[35] Patrick Jermann, Darren Gergle, Roman Bednarik, and Susan Brennan. 2012. Duet 2012: Dual Eye Tracking in CSCW. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion (CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 23–24. https://doi.org/10.1145/2141512.2141525

[36] Patrick Jermann, Marc-Antoine Nüssli, and Weifeng Li. 2010. Using Dual Eye-tracking to Unveil Coordination and Expertise in Collaborative Tetris. In *Proceedings of the 24th BCS Interaction Specialist Group Conference (BCS '10)*. British Computer Society, Swinton, UK, UK, 36–44. http://dl.acm.org/citation.cfm?id=2146303.2146309

[37] Heather Johnson. 2018. Slack. *Journal of the Medical Library Association* 106, 1 (2018), 148–151. https://doi.org/10.5195/jmla.2018.315

[38] Growth From Knowledge. 2016. Smartphone Users Spend as Much Time on Entertainment as Texting - GfK MRI Study. https://www.gfk.com/en-us/insights/press-release/smartphone-users-spend-as-much-time-on-entertainment-as-texting-gfk-mri-study/

[39] Robert E. Kraut, Susan R. Fussell, Susan E. Brennan, and Jane Siegel. 2002. Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. In *In P. Hinds & S. Kiesler (Eds.), Distributed work*. MIT Press, 137–162.

[40] Grete Helena Kütt, Kevin Lee, Ethan Hardacre, and Alexandra Papoutsaki. 2019. Eye-Write: Gaze Sharing for Collaborative Writing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 497, 12 pages. https://doi.org/10.1145/3290605.3300727

[41] Michael Lankes, Bernhard Maurer, and Barbara Stiglbauer. 2016. An Eye for an Eye: Gaze Input in Competitive Online Games and Its Effects on Social Presence. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology (ACE '16)*. ACM, New York, NY, USA, Article 17, 9 pages. https://doi.org/10.1145/3001773.3001774

[42] Michael Lankes, Matej Rajtár, Oleg Denisov, and Bernhard Maurer. 2018. Socialeyes: Social Gaze in Collaborative 3D Games. In *Proceedings of the 13th International Conference on the Foundations of Digital Games (FDG '18)*. ACM, New York, NY, USA, Article 3, 10 pages. https://doi.org/10.1145/3235765.3235766

[43] Michael Lankes, Daniel Rammer, and Bernhard Maurer. 2017. Eye Contact: Gaze as a Connector Between Spectators and Players in Online Games. In *Entertainment Computing – ICEC 2017*, Nagisa Munekata, Itsuki Kunita, and Junichi Hoshino (Eds.). Springer International Publishing, Cham, 310–321.

[44] John P Lightle, John H Kagel, and Hal R Arkes. 2009. Information exchange in group decision making: The hidden profile problem reconsidered. *Management Science* 55, 4 (2009), 568–581.

[45] Andreas Löber, Sibylle Grimm, and Gerhard Schwabe. 2006. Audio vs chat: Can media speed explain the differences in productivity?. In *ECIS*. 2172–2183.

[46] Andreas Löber, Gerhard Schwabe, and Sibylle Grimm. 2007. Audio vs. chat: The effects of group size on media choice. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. IEEE, 41–41.

[47] Bernhard Maurer, Michael Lankes, Barbara Stiglbauer, and Manfred Tscheligi. 2016. EyeCo: Effects of Shared Gaze on Social Presence in an Online Cooperative Game. In *Entertainment Computing - ICEC 2016*, Günter Wallner, Simone Kriglstein, Helmut Hlavacs, Rainer Malaka, Artur Lugmayr, and Hyun-Seung Yang (Eds.). Springer International Publishing, Cham, 102–114.

[48] Anne Meier, Hans Spada, and Nikol Rummel. 2007. A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning* 2, 1 (01 Mar 2007), 63–86. https://doi.org/10.1007/s11412-006-9005-x

[49] Stefan Münzer and Torsten Holmer. 2009. Bridging the Gap Between Media Synchronicity and Task Performance: Effects of Media Characteristics on Process Variables and Task Performance Indicators in an Information Pooling Task. *Communication Research* 36, 1 (2009), 76–103. https://doi.org/10.1177/0093650208326464 arXiv:https://doi.org/10.1177/0093650208326464

[50] Bonnie A. Nardi, Steve Whittaker, and Erin Bradner. 2000. Interaction and Outeraction: Instant Messaging in Action. (2000), 79âĂŞ88. https://doi.org/10.1145/358916.358975

[51] Joshua Newn, Fraser Allison, Eduardo Velloso, and Frank Vetere. 2018. Looks Can Be Deceiving: Using Gaze Visualisation to Predict and Mislead Opponents in Strategic Gameplay. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 261, 12 pages. https://doi.org/10.1145/3173574.3173835

[52] Joshua Newn, Eduardo Velloso, Fraser Allison, Yomna Abdelrahman, and Frank Vetere. 2017. Evaluating Real-Time Gaze Representations to Infer Intentions in Competitive Turn-Based Strategy Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17)*. ACM, New York, NY, USA, 541–552. https://doi.org/10.1145/3116595.3116624

[53] Frank Newport. 2014. The New Era of Communication Among Americans. https://news.gallup.com/poll/179288/new-era-communication-americans.aspx

[54] Marc-Antoine Nüssli. 2011. Dual Eye-Tracking Methods for the Study of Remote Collaborative Problem Solving. (2011), 229.

[55] Marc-Antoine Nüssli, Patrick Jermann, Mirweis Sangin, and Pierre Dillenbourg. 2009. Collaboration and abstract representations: towards predictive models based on raw speech and eye-tracking data. In *Proceedings of the 9th international conference on Computer supported collaborative learning-Volume 1*. International Society of the Learning Sciences, 78–82.

[56] Gary M. Olson and Judith S. Olson. 2000. Distance Matters. *Hum.-Comput. Interact.* 15, 2 (Sept. 2000), 139–178. https://doi.org/10.1207/S15327051HCI1523_4

[57] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. 2019. The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI* 6, 5 (2019), 02.

[58] Michael Prilla. 2019. "I Simply Watched Where She Was Looking at": Coordination in Short-Term Synchronous Cooperative Mixed Reality. *Proc. ACM Hum.-Comput. Interact.* 3, GROUP, Article 246 (Dec. 2019), 21 pages. https://doi.org/10.1145/3361127

[59] Ronald E. Rice. 1993. Media Appropriateness: Using social presence theory to compare traditional and new organizational media. *Human Communication Research* 19, 4 (1993), 451–484. https://doi.org/10.1111/j.1468-2958.1993.tb00309.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2958.1993.tb00309.x

[60] Christian Schlösser. 2018. Towards Concise Gaze Sharing. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*. ACM, New York, NY, USA, Article 78, 3 pages. https://doi.org/10.1145/3204493.3207416

[61] Bertrand Schneider and Roy Pea. 2013. Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning* 8, 4 (01 Dec 2013), 375–397. https://doi.org/10.1007/s11412-013-9181-4

[62] Ulrike Schultz and Betty Vandenbosch. 1998. Information Overload in a Groupware Environment: Now you see it, now you don't. *Journal of Organizational Computing and Electronic Commerce* 8, 2 (1998), 127–148.

[63] John Short, Ederyn Williams, and Bruce Christie. 1976. *The Social Psychology of Telecommunications*. John Wiley & Sons.

[64] Harri Siirtola, Oleg Špakov, Howell Istance, and Kari-Jouko Räihä. 2019. Shared gaze in collaborative visual search. *International Journal of Human–Computer Interaction* (2019), 1–13.

[65] Garold Stasser and Dennis Stewart. 1992. Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of Personality and Social Psychology Quarterly* (1992), 67–78.

[66] Oleg Špakov, Howell Istance, Kari-Jouko Räihä, Tiia Viitanen, and Harri Siirtola. 2019. Eye Gaze and Head Gaze in Collaborative Games. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications (ETRA '19)*.

ACM, New York, NY, USA, Article 85, 9 pages. https://doi.org/10.1145/3317959.3321489

[67] Joseph Tao-yi Wang, Michael Spezio, and Colin F. Camerer. 2010. Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games. *American Economic Review* 100, 3 (June 2010), 984–1007. https://doi.org/10.1257/aer.100.3.984

[68] Nancy Yao, Jeff Brewer, Sarah D'Angelo, Mike Horn, and Darren Gergle. 2018. 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 EA '18)*. ACM, New York, NY, USA, Article LBW051, 6 pages. https://doi.org/10.1145/3170427.3188453