Gazed and Confused: Understanding and Designing Shared Gaze for Remote Collaboration

Sarah D'Angelo and Darren Gergle
Northwestern University
2240 Campus Dr., Evanston, Illinois 60208, USA
sdangelo@u.northwestern.edu; dgergle@northwestern.edu

ABSTRACT
People utilize eye gaze as an important cue for monitoring attention and coordinating awareness. This study investigates how remote pairs make use of a graphical representation of their partner's eye-gaze during a tightly-coupled collaborative task. Our results suggest that reproducing shared gaze in a remote collaboration setting makes pairs more accurate when referring to linguistically complex objects by facilitating the production of efficient forms of deictic references. We discuss how the availability of gaze influences coordination strategies and implications for the design of shared gaze in remote collaboration systems.

Author Keywords
Eye-tracking; computer supported collaborative work

ACM Classification Keywords
H5.3. Information interfaces and presentation (e.g., HCI): Group and organizational interfaces – collaborative computing, computer-supported collaborative work

INTRODUCTION
It is increasingly common to use Internet-based technologies such as video-mediated communication systems, shared display groupware, and telepresence systems for remote collaboration. Popular accounts of these systems often portend the death of distance and promote notions of equitable access to resources such as online education or remote medical assessments. However, current scholarship paints a more complex picture. Challenges arise when collaborative tools are developed without a thorough understanding of the ways in which groups coordinate their activities. For example, failure to consider the visual context of an interaction can result in systems that impair a pair's ability to establish mutual understanding and effectively collaborate [8].

In this paper we examine shared gaze and its importance for coordinating interaction. We explore the role it plays in attention and collaborative work, and examine its potential as a design feature for remote collaboration systems.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CHI'16, May 07-12, 2016, San Jose, CA, USA
© 2016 ACM. ISBN 978-1-4503-3362-7/16/05⋯$15.00
DOI: http://dx.doi.org/10.1145/2858036.2858499

BACKGROUND
When pairs collaborate in face-to-face settings they rely upon views of others' actions to coordinate their interactions [13,14]. Access to shared visual information is particularly useful for establishing situation awareness [7] (i.e., assessing the current state of the task and planning future actions) and conversational grounding (i.e., supporting the conversation around a joint activity and providing evidence of mutual understanding) [5,10]. In other words, pairs better coordinate because they can monitor task state, deliver timely statements and clarifications, and efficiently and unambiguously refer to task objects.

Shared gaze, or the ability to see where a partner is looking, is a critical feature for coordination because it provides evidence of an individual's allocation of attention. Monk and Gale [15] demonstrated that seeing where a partner was looking was more effective than simply seeing a partner's face. Vertegaal and colleagues [19] showed how gaze behavior could be used to better understand attention during group conversations [cf., 17]. Pairs also use their partner's gaze to help disambiguate similar objects [11].

In response to these findings, researchers have begun to develop dual eye-tracking techniques that track and display both partners' eye gaze [3,4,12]. Studies have shown that pairs are more efficient and faster at finding a target when their partner's gaze cursor is displayed on their workspace [2,16]. Shared gaze can also increase learning gains for students discussing complex diagrams [18], and it has been shown to serve as an effective referential pointer [1].

THE CURRENT STUDY
In this paper we are interested in understanding how shared gaze can be used as a collaborative resource, and we examine this in three different settings: co-located, remote with shared gaze, and remote without shared gaze. The co-located condition serves as a baseline for performance comparison with the remote setups that examine a shared visual space with and without shared gaze. Prior work suggests that the benefits of shared visual feedback depend upon task and object features [10]. Therefore, we also manipulate the lexical complexity and discriminability of the objects because it has been shown that visual feedback is particularly useful when referential grounding is difficult.

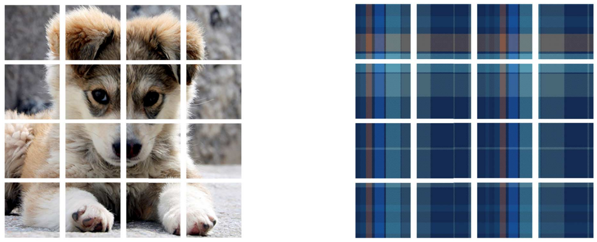

Figure 1: Example simple (left) and complex puzzles (right).

Our task requires participants to jointly assemble a 4 x 4 puzzle (figure 1). The puzzle pieces are randomly placed in a fixed position on the screen. To move a piece both participants must simultaneously select two independent pieces. Both mouse cursors are displayed in the co-located condition; however, they are not shared across participants in the remote conditions. Dragging the pieces to within 10 pixels of their matching edges allows them to be combined.

Participants perform the task in all three gaze configurations. They are either co-located, seated side-by-side and working together on the same display, or pseudo remote and working on individual but identical displays separated by a visual barrier (figure 2). In the co-located condition, physical gesturing was not explicitly forbidden; however, the task design encouraged pointing via the mouse cursor. Each trial is comprised of four puzzles, which are composed of either simple objects or visually complex patterns (figure 1 and 3). Score and time are displayed on screen to encourage both fast and accurate performance. We employ a 3 × 2 within-subjects design with degree of shared gaze (co-located, remote with shared gaze, remote without shared gaze) and linguistic complexity (simple and complex) as the factors. The experimental trials are fully counterbalanced and pairs are randomly assigned (without replacement) to each set of task orderings.

METHOD

Participants

Thirty-six students and staff members from a Midwestern U.S. University participated in the study. Fifty-five percent of the pairs were female-female (6% male-male, 39% mixed). Sixty-one percent of participants were Asian/Pacific Islander (28% Caucasian, 3% African-American, 8% Mixed Race / Other). Participants ranged in age from 18 to 40 with 53% percent in the 18-21 range. Individuals with colorblindness were excluded from the study. Pairs had no prior exposure to one another and no prior experience with the shared gaze display. All participants were consented and received $10 or course credit for their participation.

Procedure

Participants are first asked to complete a colorblindness test and demographic survey. They are then given instructions and encouraged to communicate freely. Participants are informed they will be timed for each puzzle and receive a score, and then they perform a simple 3 x 3 practice trial with the experimenter present to answer any questions. In

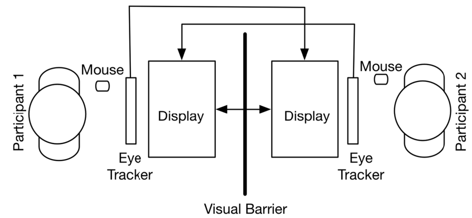

Figure 2: Dual eye-tracking configuration.

the remote conditions the participants' eye gaze is calibrated using a 9-point calibration. Participants were recalibrated when the tracker became inaccurate (required once each for 7 of 18 pairs).

Measures

Accuracy Score captures the extent to which the pairs successfully coordinated on piece selection and assembly. The pairs receive 10 points for correctly combining two pieces and they lose 10 points for incorrect combinations.

Completion Time is recorded as the time (in milliseconds) from when the puzzle first appears on the screen until all pieces are combined. If participants are recalibrated, the recalibration time is subtracted from overall time.

Our analysis uses a mixed model regression with shared gaze condition (co-located, remote with shared gaze, remote without shared gaze), linguistic complexity (simple, complex), and trial (1-4) as within-pair factors, and pair is modeled as a random effect. Additional analyses on deictic references and acknowledgements uses a mixed model regression with shared gaze and linguistic complexity as within-pair factors, and pair is modeled as a random effect.

Conversational Coding is used to help us understand the ways in which the pairs make use of the shared gaze indicator. The spoken corpus contained a total of 84,394 words, with an average of 4,689 words per pair.

Two levels of coding are performed. The first is a general classification of utterances related to the shared gaze display or in which participants use the gaze cursor as verified by videos. The following behaviors emerged as common patterns (and are further discussed in the results section): directed use of gaze (e.g., pointing, comparing, confirming, and clarifying) and problems with gaze (e.g., misalignment, confusion, and distractions). Additionally, we use a targeted coding scheme to capture instances of deictic references and acknowledgements (table 1). These capture linguistic

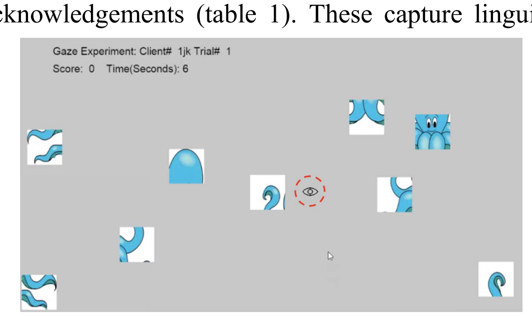

Figure 3: Partner's gaze representation highlighted in red.

expressions that change depending on the visual context in which they are used. For example, when pairs discuss objects in a shared space they often shift from using longer noun phrases to shorter deictic pronouns [9].

Deictic references phrases that use the deictic terms "this," "that," "there," or other related terms to describe a piece.	*"You grab that piece, and I'll grab the corresponding piece."*
Acknowledgment phrases such as "okay," "yeah," and/or repetition of the previous speaker's words.	*S1: "This one?"* *S2: "Yeah"*

Table 1. Coding scheme and examples.

Apparatus

We use two Eyetribe remote eye trackers and two 20" Dell monitors with wireless mice. The eye trackers capture the gaze patterns of each participant and our software sends the coordinates to the partner's display to be visually represented with an eye cursor (figure 3). We sample at a rate of 30Hz and participants are calibrated to an accuracy between .5 and 1 degree of visual angle. The task is locally networked to mirror all actions on each display.

Natural eye movements and fixations are sporadic, and without correction can lead to overly jittery shared gaze streams. We implemented a real-time smoothing function to reduce the distracting local eye movements that occur around fixations. We iteratively applied a velocity based saccade detection method to determine an appropriate threshold for the smoothing function (following [6]). For every new gaze coordinate, the function first calculates the speed of gaze movement (distance from current gaze to the previous one divided by time). If the speed of the gaze is smaller than 1.5 pixels per second, the cursor will not move from the current location. If the speed of the gaze is larger than 1.5 pixels per second, the cursor will move to the new coordinates. The smoothing function is as follows:

$$if \; \frac{\sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}}{time} \geq 1.5 \; pixels/s$$

RESULTS

The degree of shared gaze did not exhibit a main effect on the accuracy score, $F(2, 32.7) = 0.88$, $p = .43$; however, this was masked by a significant interaction between the degree of shared gaze and linguistic complexity, $F(2, 34.1) = 3.35$, $p = .047$, (figure 4). The interaction reveals little difference in score when the pieces are easy to describe with language (all p's > .05). However, when the task objects are linguistically complex, the pairs perform better in the co-located than the no shared gaze condition ($F(1, 56)=5.02$, $p = .029$). The shared gaze condition was not found to be different from either the co-located or the no shared gaze conditions (both p's > .05). To summarize, when the task objects are easy to describe with language, collocation and shared gaze make little difference. However, when the task objects are linguistically ambiguous, co-location performs better than no shared gaze, while the results for shared gaze are inconclusive.

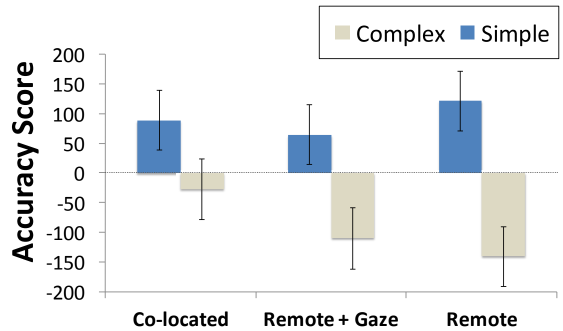

Figure 4. The effect of shared gaze and linguistic complexity on accuracy (error bars represent +/- SE).

We also examined completion time, which is important for tasks that may exhibit a speed-accuracy tradeoff. The degree of shared gaze had an effect whereby the pairs were fastest to complete the puzzle when they were in the co-located condition ($M = 176.4$s $SE = 11.3$) compared to both the shared gaze condition ($M = 231.7$s $SE = 11.4$; for the comparison $F(1, 30.55) = 24.6$, $p < .001$) and the no shared gaze condition ($M = 222.8$s $SE = 11.1$; $F(1, 28.6) = 18.1$, $p < .001$). There were no detectable differences in the pairs' performance in the shared gaze and no shared gaze setups, $F(1, 29.4) = 0.65$, $p = .43$.

Content Analysis

Participants used significantly more deictic references when talking about pieces in the shared gaze condition ($M = 23$ $SE = 2.95$) compared to the no shared gaze condition ($M = 17.13$ $SE = 2.95$); for the comparison $F(1, 42) = 8.53$, $p = .006$. Additionally, participants produced more acknowledgments in the shared gaze ($M = 24.27$ $SE = 1.69$) compared to the no shared gaze condition ($M = 19.4$ $SE = 1.69$; $F(1, 42) = 5.55$, $p = .023$); however, there was no significant difference in the number of questions asked between the two remote conditions.

Further analysis of the conversations suggests that participants used the gaze cursor to attract attention to a specific piece or compare multiple pieces without needing to use descriptive language. For example, in figure 5, Subject A uses their gaze to indicate two pieces (line 1-2), and Subject B follows their gaze and is able to correctly identify and

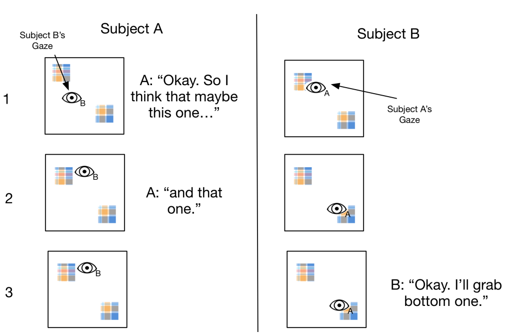

Figure 5: Pointing and comparing (all illustrations are simplified by removing non-relevant pieces).

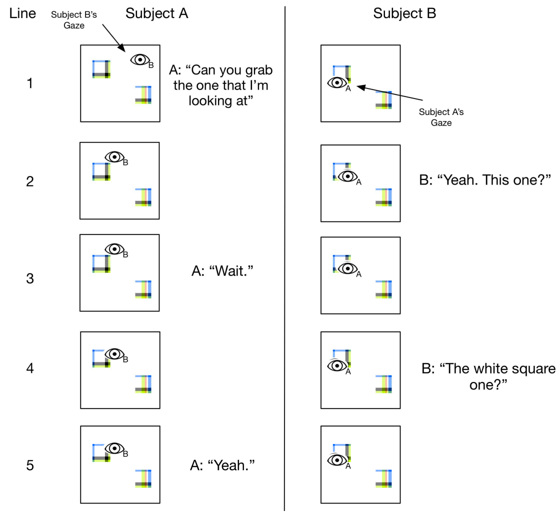

Figure 6: Confirming and clarifying.

combine the two pieces (line 3). The use of pointing and comparing allowed participants to bypass grounding on pieces and instead use deictic references.

Participants were also able to coordinate by indicating with both gaze and language to ground on the pieces. For example, in figure 6, Subject B uses their gaze cursor to signal back to Subject A and confirm that they are looking at the same piece (line 2). This requires both participants to explicitly mention and acknowledge each other's gaze.

Participants also used the gaze cursor to circumscribe the referential domain [14]. The gaze cursor allows participants to highlight a particular sub-region of the screen and then use more general features to describe a specific piece. For example in figure 6, Subject B, maintains their gaze to indicate the piece and uses generic features to clarify that they are looking at the same piece (line 4).

Problems with Gaze

Shared gaze was not entirely beneficial. In fact, it becomes problematic when the eye-tracking is even slightly misaligned. When this occurred participants sent mixed signals to their partner by describing one piece while appearing to be looking at a different piece. For example, in figure 7, Subject A tries to communicate with Subject B about two pieces but their gaze cursor is misaligned and the pair are not aware they coordinated on different pieces.

Participants were also reluctant to use the gaze cursor because of the nature of gaze itself. Sporadic eye movement can be difficult to interpret and potentially distracting. In the following excerpt, Subject A tries to use the gaze cursor to indicate a piece but Subject B cannot interpret the signal.

Subject A*: Okay... Do you see the ones with, like the orange stripes? The one I'm looking at now? And then this other one that I'm looking at.*

Subject B*: The top... Your eyes are very... You're moving a lot.*

Subject A*: Well, they're like one of the "L" parts.*

In this case the gaze cursor is a distracting feature that requires more effort than it is worth.

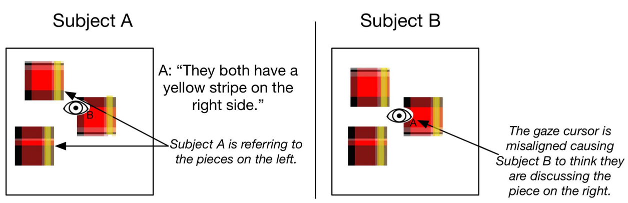

Figure 7: Misalignment.

DISCUSSION

The addition of shared gaze in remote collaborative work allows pairs to coordinate and use strategies similar to co-located practices. In this study we see that pairs are more accurate with linguistically complex objects when they are co-located as compared to when they are remote without shared gaze, while the results for the shared gaze condition are inconclusive. However, a detailed content analysis reveals that pairs use efficient deictic references and acknowledgements of behavior when gaze is shared compared to when it is not. The most explicit benefits of the gaze cursor occur when it is used as a pointer to highlight areas of interest or compare pieces, especially when the visual space or objects being discussed are lexically complex. Additionally, gaze information was used to signal between participants to confirm and clarify points of confusion and facilitate agreement before piece selection.

Designing shared gaze for remote tasks has a number of constraints resulting from the fact that eye trackers do not always provide a reliable track and as a result can misrepresent where people are looking. However, attempting to maintain accuracy by providing feedback to the gaze sender is not effective because it can produce a feedback loop that causes people to follow their own cursor.

Another challenge exists in that gaze is typically a subtle cue in co-located work, while in contrast, when gaze is shared digitally the constant movement and lack of intentional signaling can be disruptive and confusing. This may be a result of our task design; in search tasks [2] gaze information has a defined role that has been shown to facilitate coordination when shown continuously. However, in a highly interdependent task without a leader - follower distinction [18], attending to gaze was primarily beneficial when it was used purposefully.

Finally, future avenues for the design of shared gaze should consider how often gaze information should be displayed and when. For example, shared gaze was primarily useful when participants made directed intentional eye movements by staring at a region of interest for an extended period of time. A form of selective shared gaze could reduce some of the distracting aspects of a continuous gaze stream [cf., 17].

ACKNOWLEDGMENTS

We would like to thank Jerry Li and Mia Manavalan and acknowledge the support of NSF grant #0953943 and NSF GRFP #DGE-1324585.

REFERENCES

1. Ellen Gurman Bard, Robin L. Hill, Mary Ellen Foster, and Manabu Arai. 2014. Tuning accessibility of referring expressions in situated dialogue. *Language, Cognition and Neuroscience* 29, 8: 928–949.

2. 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: 1465–1477.

3. Jean Carletta, Robin L. Hill, Craig Nicol, Tim Taylor, Jan Peter De Ruiter, and Ellen Gurman Bard. 2010. Eyetracking for two-person tasks with manipulation of a virtual world. *Behavior Research Methods* 42, 1: 254–265.

4. Alan T. Clark and Darren Gergle. 2011. Mobile dual eye-tracking methods: challenges and opportunities. *Proc. of International Workshop on Dual Eye Tracking*.

5. Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. *Perspectives on socially shared cognition* 13, 1991: 127–149.

6. Andrew Duchowski. 2007. *Eye tracking methodology: Theory and practice*. Springer Science & Business Media.

7. Mica R. Endsley. 1995. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1: 32–64.

8. Susan R. Fussell, Robert E. Kraut, and Jane Siegel. 2000. Coordination of communication: Effects of shared visual context on collaborative work. *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, ACM, 21–30.

9. Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam DI Kramer. 2004. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction* 19, 3: 273–309.

10. 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: 1–39.

11. Joy E. Hanna and Susan E. Brennan. 2007. Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language* 57, 4: 596–615.

12. Patrick Jermann, Darren Gergle, Roman Bednarik, and Susan Brennan. 2012. Duet 2012: dual eye tracking in CSCW. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*, ACM, 23–24.

13. Robert E. Kraut, Susan R. Fussell, and Jane Siegel. 2003. Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction* 18, 1: 13–49.

14. Robert E. Kraut, Darren Gergle, and Susan R. Fussell. 2002. The use of visual information in shared visual spaces: Informing the development of virtual copresence. *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, ACM, 31–40.

15. Andrew F. Monk and Caroline Gale. 2002. A look is worth a thousand words: Full gaze awareness in video-mediated conversation. *Discourse Processes* 33, 3: 257–278.

16. Mark B. Neider, Xin Chen, Christopher A. Dickinson, Susan E. Brennan, and Gregory J. Zelinsky. 2010. Coordinating spatial referencing using shared gaze. *Psychonomic bulletin & review* 17, 5: 718–724.

17. Jiazhi Ou, Lui Min Oh, Susan R. Fussell, Tal Blum, and Jie Yang. 2008. Predicting visual focus of attention from intention in remote collaborative tasks. *Multimedia, IEEE Transactions on* 10, 6: 1034–1045.

18. 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: 375–397.

19. Roel Vertegaal, Robert Slagter, Gerrit Van der Veer, and Anton Nijholt. 2001. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM,301–308.