StARe: Gaze-Assisted Face-to-Face Communication in Augmented Reality

Radiah Rivu
Bundeswehr University Munich
Munich, Germany
sheikh.rivu@unibw.de

Yasmeen Abdrabou
Bundeswehr University Munich
Munich, Germany
yasmeen.essam@unibw.de

Ken Pfeuffer
Bundeswehr University Munich
Munich, Germany
ken.pfeuffer@unibw.de

Augusto Esteves
ITI / LARSyS, Instituto Superior Técnico, University of Lisbon
Lisbon, Portugal
augusto.esteves@tecnico.ulisboa.pt

Stefanie Meitner
LMU Munich
Munich, Germany

Florian Alt
Bundeswehr University Munich
Munich, Germany
florian.alt@unibw.de

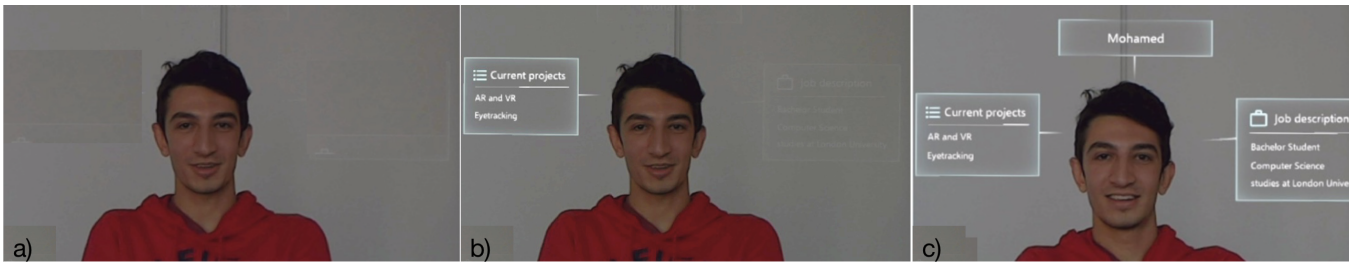

Figure 1: AR devices allow dynamic information overlays. Gaze provides a mean to select information on demand when the user looks at that particular information. This allows the user to keep the UI uncluttered (a), but at anytime make specific information appear (b). Without gaze, showing all information can distract the user (c)

ABSTRACT

This research explores the use of eye-tracking during Augmented Reality (AR) - supported conversations. In this scenario, users can obtain information that supports the conversation, without augmentations distracting the actual conversation.We propose using gaze that allows users to gradually reveal information on demand. Information is indicated around user's head, which becomes fully visible when other's visual attention explicitly falls upon the area. We describe the design of such an AR UI and present an evaluation of the feasibility of the concept. Results show that despite gaze inaccuracies, users were positive about augmenting their conversations with contextual information and gaze interactivity. We provide insights into the trade-offs between focusing on the task at hand (i.e., the conversation), and consuming AR information. These findings are useful for future use cases of eye based AR interactions by contributing to a better understanding of the intricate balance between informative AR and information overload.

ETRA '20, June 2–5, 2020, Stuttgart, Germany

ACM ISBN 978-1-4503-7135-3/20/06…$15.00
https://doi.org/10.1145/3379157.3388930

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; **Human computer interaction (HCI)**; **Mixed / augmented reality**; **Pointing**; **Visualization**.

KEYWORDS

AR; Gaze Interaction; Eye-tracking; Assistive Conversation

ACM Reference Format:
Radiah Rivu, Yasmeen Abdrabou, Ken Pfeuffer, Augusto Esteves, Stefanie Meitner, and Florian Alt. 2020. StARe: Gaze-Assisted Face-to-Face Communication in Augmented Reality. In *Symposium on Eye Tracking Research and Applications (ETRA '20), June 2–5, 2020, Stuttgart, Germany.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3379157.3388930

1 INTRODUCTION

Head mounted devices with augmented reality technology are maturing and can be soon expected to enter the consumer market. These devices can become as pervasive as the smartphone, with promising applications as an assistive technology in everyday life. These devices also integrate new sensors such as eye trackers, that support not only improved graphics rendering (e.g., foveated rendering [18]) but also aid users in user interface controls [14].

We explore the interface for eye-tracked AR devices aimed at assisting users in face-to-face communication. In conversations, AR can provide additional, contextual information to be superimposed onto people. The user can quickly obtain knowledge about the person they are talking with, with such a system being useful for

small talk, interviews, or other relevant situations. However, this also poses new challenges in the design of the user interface. The UI needs to strive a balance between providing as much information as possible to be useful, but at the same time keep the interface as simple as possible to avoid interrupting the conversation flow.

We propose using gaze interaction in such an interface to dynamically reveal information on demand. Additional user information is only revealed if the other shows interest through visual cues. Particularly, information boxes close to a user's head fade in when the other gazes at them for a prolonged time. If the user decides to ignore those, the information remains in the background. This allows the user to engage and consume information if wanted, but also ignore the information to stay focused on the conversation.

In this paper, we investigate a first prototype of such a gaze-enabled AR interface. We first describe the design and implementation of the system. In contrast to prior interfaces that augment information on people [1, 5], using gaze requires a balance between different design goals, including maximising utility, minimising information overload [2, 6], and avoiding the Midas Touch [10]. Here we propose a multi-stage visualisation of augmented information. As shown in Figure 1, information boxes with different information such as name, personal info. or interests are displayed around the head of each conversational partner. To avoid overwhelming the user by always presenting all this information, it is faded out unless his or hers visual attention focuses on it.

We then evaluate a conversation scenario in which the user is talking to people while superimposed personal information about each of them is displayed. The study participant is wearing the AR device, and converses with two confederates. Our focus is on the user experience with the new user interface. Users experienced the gaze based approach, and another non-gaze variant where the information is constantly visible. We collect qualitative feedback in form of questionnaire inputs, interviews, and observations.

The findings show that while the current hardware poses feasibility challenges for everyday use, users were positive about the concept of dynamically gathering more information, and be in control of it — while also aware the conversation is the primary task and the UI needs to be carefully designed to stay in the background. Thus, our research provides insights into the design of a gaze UI for people and points to promising applications to support users in conversations with people.

2 RELATED WORK

In contrast to visualizations for digital content such as menus or personal desktop UIs, situated visualization [21] or contextual information [1] show information anchored onto the environment, e.g. restaurants, signposts, or people. Azuma et al. hints at the issue of information overload that can overwhelm the user [2]. To avoid user effort of manual filtering, the UI layout can be designed to alleviate this, e.g. by view management techniques, where the system automatically avoids occlusion and groups related information [3]. Knowledge-based AR [6] takes into account knowledge of the environment such as object location and shape, and uses goals and rules to prioritise context relevant information. To optimise data density, Julier et al. consider physical, spatial and hybrid filtering [11]. Keil et al. proposes camera and motion based techniques [12].

Information filtering with gaze was investigated by Peripheral Vision Annotations [8], using gaze tracking to make information appear in the visual periphery to avoid occluding the center of visual attention. Ajanki et al. used gaze as a cue to provide contextual information to augmented objects [1], and Tönnis and Klinker attached information to the user's line of sight [19]. An investigation of early prototypes of gaze based information highlighting showed improvements in performance and positive user feedback [13, 15, 16]. Implicit use of gaze related to non-command interfaces [9, 17], that contrast typical command UI with e.g. the eyes as a convenient, natural, and high-bandwidth source for input [10]. Attentive User Interfaces [20] consider attention as a resource to be carefully managed by the UI to not overwhelm the user, i.e. display only relevant information to make interactions more effective [7].

Gaze-orchestrated Dynamic Windows [4] utilised gaze input to zoom into objects of interest and provide detail-on-demand. Jacob identified the Midas Touch issue of accidental gaze selection, and proposes dwell-time or manual activation to resolve this [10]. In AR, Pinpointing [14] are interaction techniques that leverage combinations of gaze, head and manual input to precisely select targets in AR, and in the paper the time-accuracy trade-off is studied. We extend prior work, and focus on investigating gaze to reveal contextual information about people in AR.

3 STARE USER INTERFACE

We now describe the design and implementation of the gaze based interface for people as illustrated in Figure 2.

3.1 Design Considerations

The premise is a conversation scenario in which the user is talking to people while superimposed personal information about each of them is displayed. These super-impositions can be considered like information boxes that appear around the user's head as the proxy for most of the user's attention, a concept applied to similar people-UIs developed in prior work [1, 5]. In contrast to typical AR UIs, the scenario of a conversation has specific design challenges that need to be considered:

Utility vs. Overload: The overarching trade-off is between the benefit of gaining information and the drawback of interrupting the conversation. How can the UI provide useful information, without overloading users and being detrimental to the main activity?

Voluntary vs. Involuntary Activation: The information boxes need to be activated by gaze. It can be considered a selection mechanism that is hindered by the Midas Touch problem [10], i.e. it does not distinguishes whether a user's gaze is intended to select or is only looking. Dwell-time can be used to alleviate the issue, by using a set time required to look at the box to activate it. This introduces another trade-off: a too long time to activate can become annoying for users, while a too short time can involuntary activate it.

Primary vs. Secondary Activity: In this context, the primary activity is the conversation between users. The secondary activity the use of the AR interface to obtain more information. As a secondary activity, it should minimise the effort required to complete the information task, so that users can keep focused on conversing.

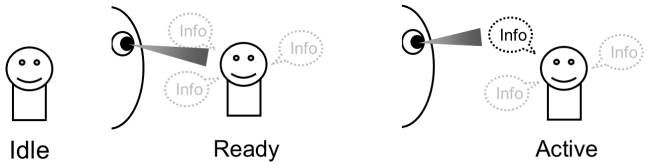

Figure 2: Design considerations in a conversation scenario

Information on Demand vs. Interface Cluttering: Considering the display of information in an AR head mounted device, ideally it should consume the least amount of space from the user's field of view (FoV). The idea is to keep the user's FoV as clear as possible with minimal clutter. With gaze, information can dynamically unfold to only augment the UI on demand.

3.2 Technical Design

As shown in Figure 1, three information boxes with different information such as name, personal info., or interests were displayed around the head of each conversation partner. In order to avoid overwhelming the user by always presenting all information, we have decided to subtract the information of all people until they are explicitly highlighted by the user's gaze.

Especially when we have a conversation with people, retrieving contextual information about each person may be relevant, such as talking about current research activities in university and being able to see which topics each person is currently researching about.

Since the main task is the conversation, filtering the information should only be performed as a secondary task. We have therefore considered a concept of highlighting the information as subtly as possible. The information boxes can assume three different states:

- *Idle*: Information boxes are not visible when the user is not visually attending to the conversation partner.
- *Ready*: Information boxes are ready to be consumed, and are slightly visible (5% opacity) when the user attends to the conversant without focusing on the boxes.
- *Active*: When the user gazes at an information box, it becomes fully visible (100% opacity).

As soon as the user looks at a person, all information boxes of that person change their state from *Idle* to *Ready*. Here, the user can already recognize the outline of the boxes, however, to be able to read the information, the user needs to look at one of the boxes in order to make it *Active*. The transition from one state to another is performed by a smooth fade in and fade out animation. Once the user has turned his gaze away from one information box, it is set to *Ready* again. All boxes are set to their default state of *Idle* in case no person is being looked at.

4 EVALUATION

Our goal is to understand (1) how the gaze-based information is perceived by the users, (2) how this gaze-based interaction and the visualization of the information affects the conversation, (3) whether the content of the overlaid information was useful during the conversation, and (4) in which other scenarios such information might be relevant.

Figure 3: Study setup of a three person conversation scenario

4.1 Task

We scripted a conversation of three people of which one was the actual participant and the two others confederates. Since users sat in fixed locations, we displayed the augmented information at absolute positions in the vicinity of the confederates's heads. The task was to imagine a first day at the university and getting to know new colleagues. The study ended with a short questionnaire and interview session.

4.2 Study Design

The study consisted of each user conducting two conversations, each 3-5 minutes with different topics (order counterbalanced). Conversation topics included *work* (the boxes showed information about the job position, research field, and research projects), *social life* (personal information and hobbies were presented). The confederates' names was displayed in both scenarios.

4.3 Apparatus and Participants

12 participants (6 female) aged between 21-30 years (M = 25, SD = 2.99) took part in the study, mostly students from a technical discipline. We used a Microsoft HoloLens (1st gen, 1268 px × 720 px per eye, 60 Hz, 30° × 17.5° FoV). The HoloLens was extended with a Binocular eye tracking add-on from Pupil Labs to record eye data (200 Hz, 9-point calibrations).

4.4 Study Procedure

As participants arrived, we explained the study and had them fill in a consent form and a demographics questionnaire. Participants were asked to carry out conversations for 3-5 minutes with the given topics. After the study we asked them to fill in an additional questionnaire for user feedback. We concluded with a semi-structured interview to collect qualitative feedback in which we asked them to reflect on the experience about the conducted conversations. A setup of the study is displayed in Figure 3.

5 RESULTS

We first discuss general feedback about the UI, and then specifics about the trade-offs.

5.1 General Feedback and Use Cases

Our observations, questionnaire and interview reveal that participants welcome the idea of having gaze-revealing information during conversations. We can categorise our insights as follow:

The participants described the UI as useful as it allows them to better remember information, especially when one has to remember names of people (P4, P8, P9) and "it was nice to see information about someone before they themselves say it" (P8). We also learned that to few participants this design seemed to be useful only in specific applications such as a training or teaching scenario rather than everyday conversations (P5).

Few participants partly felt that from a social perspective, it can be overwhelming to see information in advance, especially given they are unfamiliar with this interface design. We observed that many participants were unsure how to integrate the displayed information into the conversation. Participants who shared similarities such as hobbies or same course of study found the conversation became easier and more natural (for a total of 7 participants).

Most participants suggested that it would be helpful for meeting new people (P2, P3, P4, P9) and well suited for a job interview (P3, P6, P9, P11), as we could gather information about the background of the interviewer and thus be "more confident and open in the interview" (P3). Providing additional information would also be conceivable at events such as a conference or meeting where many people come together and useful for people "with poor memory or even with dementia" (P7).

5.2 Utility vs. Overload Trade-Offs

Overall we found that users preferred the gaze based information overlay concept due to this being less distracting (P1, P3, P7, P8, P11), in comparison to showing all information constantly. This indicates that the information overload can occur during conversations, and a more selective approach using gaze can alleviate the issue.

Yet, there were a few participants who preferred all information for various reasons. One user stated that "having to select a text box by looking at it required more attention" (P10). As the used eye tracker of our system was inaccurate in some cases, the actual selection with gaze became erroneous which reduced selection quality. This also affected the actual conversation, as the additional selection effort made it "harder to concentrate on the conversation" (P1). This shows that, if the system is not accurate enough, it can quickly become overloading to the user; however otherwise the majority of users clearly see the utility of the approach.

Thus, we find that critical aspects relating to human factors affected the system and thus lead to decreased usability. This included (1) gaze inaccuracy induced by rapid and multiple head movements that caused the headset to slip and resulted in a calibration offset, as well as (2) eye-tracking errors caused by laughing of the participants, whereas the cheeks lifting may narrow the eye slightly resulting in the eyes being often undetectable. A further reason for inconsistencies might be performance issues, which partly caused a delay in the presentation and, therefore, lead to inconsistent fade in and out animation behaviour of the information boxes. However, the current trend in head mounted devices with integrated eye tracking such as the Hololens 2 or Vive Eye Pro, we are sure most of these issues will be addressed soon.

6 DISCUSSION

We presented a study to evaluate the concept of gaze-assisted face-to-face communication in AR, and provide detailed insights into the possibilities and challenges of conducting conversations using AR. Based on our findings, we believe gaze based AR can become an important aspect of future head mounted devices, with the main advantages including 1) minimising information overload, 2) providing user semi-explicit control over information, and 3) possibility to gradually explore multiple information types about people.

Our primary observation is that interactions with people need to be carefully augmented with information, as the primary task is the conversation and augmentations are secondary. This is in line with the challenge of concentrating on the task of pointing to the information with the gaze, while reading the information and at the same time following the conversation. The pointing task needs to work flawlessly, to avoid the user attention shifting to the secondary pointing activity.

The small field of view of the headset is currently a limiting factor. If there were several people in the field of view of the user, a gaze-based information display could possibly be useful displaying all information of only the person on whom the attention is directed to. In future work, we believe the use of AR devices with larger field of view will likely extend the utility of the gaze AR UI. Also full field of view devices may induce a much larger effect of information overload, that can be better modulated when the user's visual interest is better integrated for augmented face-to-face conversations. We also wish to work further on the information display so that the positions are not absolute rather, enables free movement of the intended user and information are represented at a relative position on the screen. Another aspect of future work is to explore various use case applications of our proposed prototype, for example, enhanced conversations providing memory support to a certain sector of the population.

7 CONCLUSION

In this paper, we presented the design of an experimental interface for revealing information about people during a conversational task. User tests set within a real conversation, showed that it can both hinder as well as support the conversation. The possibility to have real time information within the current activity context opens questions about human behaviour, how to integrate such information, and potentially alter the flow of human-human interaction. Results show that users perceive our application as a useful additional feature during conversations, and reveal the trade-offs between information overload and utility.

ACKNOWLEDGMENTS

This research was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant agreement no. 316457582 and the Studienstiftung des deutschen Volkes ("German Academic Scholarship Foundation").

REFERENCES

[1] Antti Ajanki, Mark Billinghurst, Hannes Gamper, Toni Järvenpää, Melih Kandemir, Samuel Kaski, Markus Koskela, Mikko Kurimo, Jorma Laaksonen, Kai Puolamäki, et al. 2011. An augmented reality interface to contextual information. *Virtual reality* 15, 2-3 (2011), 161–173.

[2] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. Recent Advances in Augmented Reality. *IEEE Comput. Graph. Appl.* 21, 6 (Nov. 2001), 34–47. https://doi.org/10.1109/38.963459

[3] Blaine Bell, Steven Feiner, and Tobias Höllerer. 2001. View Management for Virtual and Augmented Reality *(UIST '01)*. ACM, New York, NY, USA, 101–110. https://doi.org/10.1145/502348.502363

[4] Richard A Bolt. 1981. Gaze-orchestrated dynamic windows. In *ACM SIGGRAPH Computer Graphics*, Vol. 15. ACM, 109–119.

[5] Rakesh D Desale and Vandana S Ahire. 2013. A study on wearable gestural interface–a sixthsense technology. *IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN* (2013), 2278–0661.

[6] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the World: 2D Windows for 3D Augmented Reality *(UIST '93)*. ACM, New York, NY, USA, 145–155. https://doi.org/10.1145/168642.168657

[7] Bernardo A Huberman and Fang Wu. 2007. The economics of attention: maximizing user value in information-rich environments. In *Proceedings of the 1st international workshop on Data mining and audience intelligence for advertising*. ACM, 16–20.

[8] Yoshio Ishiguro and Jun Rekimoto. 2011. Peripheral Vision Annotation: Noninterference Information Presentation Method for Mobile Augmented Reality *(AH '11)*. ACM, New York, NY, USA, Article 8, 5 pages. https://doi.org/10.1145/1959826.1959834

[9] Robert JK Jacob. 1993. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. *Advances in human-computer interaction* 4 (1993), 151–190.

[10] Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-based Interaction Techniques *(CHI '90)*. ACM, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246

[11] Simon Julier, Yohan Baillot, Dennis Brown, and Marco Lanzagorta. 2002. Information Filtering for Mobile Augmented Reality. *IEEE Comput. Graph. Appl.* 22, 5 (Sept. 2002), 12–15. https://doi.org/10.1109/MCG.2002.1028721

[12] Jens Keil, Michael Zoellner, Timo Engelke, Folker Wientapper, and Michael Schmitt. 2013. Controlling and filtering information density with spatial interaction techniques via handheld augmented reality. In *International Conference on Virtual, Augmented and Mixed Reality*. Springer, 49–57.

[13] Mirae Kim, Min Kyung Lee, and Laura Dabbish. 2015. Shop-i: Gaze based Interaction in the Physical World for In-Store Social Shopping Experience. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. Association for Computing Machinery, Seoul, Republic of Korea, 1253–1258. https://doi.org/10.1145/2702613.2732797

[14] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality *(CHI '18)*. ACM, New York, NY, USA, Article 81, 14 pages. https://doi.org/10.1145/3173574.3173655

[15] Ann McNamara, Katherine Boyd, David Oh, Ryan Sharpe, and Annie Suther. 2018. Using Eye Tracking to Improve Information Retrieval in Virtual Reality. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 242–243. https://doi.org/10.1109/ISMAR-Adjunct.2018.00076 ISSN: null.

[16] Ann McNamara and Chethna Kabeerdoss. 2016. Mobile Augmented Reality: Placing Labels Based on Gaze Position. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 36–37. https://doi.org/10.1109/ISMAR-Adjunct.2016.0033 ISSN: null.

[17] Jakob Nielsen. 1993. Noncommand User Interfaces. *Commun. ACM* 36, 4 (April 1993), 83–99. https://doi.org/10.1145/255950.153582

[18] Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards Foveated Rendering for Gaze-Tracked Virtual Reality. *ACM Trans. Graph.* 35, 6, Article Article 179 (Nov. 2016), 12 pages. https://doi.org/10.1145/2980179.2980246

[19] Marcus Tönnis and Gudrun Klinker. 2014. Boundary Conditions for Information Visualization with Respect to the User's Gaze *(AH '14)*. ACM, New York, NY, USA, Article 44, 8 pages. https://doi.org/10.1145/2582051.2582095

[20] Roel Vertegaal et al. 2003. Attentive user interfaces. *Commun. ACM* 46, 3 (2003), 30–33.

[21] Sean White and Steven Feiner. 2009. SiteLens: Situated Visualization Techniques for Urban Site Visits *(CHI '09)*. ACM, New York, NY, USA, 1117–1120. https://doi.org/10.1145/1518701.1518871