Eye-Tracking for Avatar Eye-Gaze and Interactional Analysis in Immersive Collaborative Virtual Environments

William Steptoe[*1], **Robin Wolff**[2], **Alessio Murgia**[3], **Estefania Guimaraes**[4],
John Rae[4], **Paul Sharkey**[3], **David Roberts**[2], **Anthony Steed**[1]

[1]Department of Computer Science, University College London   [3]Department of Computer Science, University of Reading
[2]Centre for Virtual Environments, University of Salford   [4]Department of Psychology, Roehampton University

ABSTRACT

Participants' eye-gaze is generally not captured or represented in immersive collaborative virtual environment (ICVE) systems. We present *EyeCVE*, which uses mobile eye-trackers to drive the gaze of each participant's virtual avatar, thus supporting remote mutual eye-contact and awareness of others' gaze in a perceptually unfragmented shared virtual workspace. We detail trials in which participants took part in three-way conferences between remote CAVE[TM] systems linked via *EyeCVE*. Eye-tracking data was recorded and used to evaluate interaction, confirming the system's support for the use of gaze as a communicational and management resource in multiparty conversational scenarios. We point toward subsequent investigation of eye-tracking in ICVEs for enhanced remote social-interaction and analysis.

Author Keywords

Telecommunication, Eye-Gaze, Immersive Collaborative Virtual Environments, Eye-Tracking, Avatars, Social Presence.

ACM Classification Keywords

H.4.3 Communications Applications: Computer conferencing, teleconferencing, and videoconferencing; I.3.7 Three-Dimensional Graphics and Realism: Virtual Reality

INTRODUCTION

In a face-to-face meeting, the surrounding environment is a contiguous space in which participants are able to use a full-range of non-verbal communicational resources: they can move their eyes and head to look at others, change facial expression, gesture, posture, and position themselves as desired. Despite development of highly sophisticated CSCW systems, there is no substitute for such co-located interaction which supports rich non-verbal communication while allowing free movement in a perceptually unfragmented workspace.

Video-mediated communication (VMC) systems add a video channel to audio, and have been found to improve participants' ability to show understanding, forecast responses, give non-verbal information, enhance verbal descriptions,

---
[*]W.Steptoe@cs.ucl.ac.uk

manage pauses and express attitudes [13]. However, VMC compresses the representation of 3D space, constraining rich cues available in co-located collaboration such as depth, resolution and field of view, and thereby limiting awareness and ability to point at and manipulate objects [10].

Immersive Collaborative Virtual Environments (ICVEs) connect remote or co-located users of immersive display systems (such as the CAVE) within a spatial, social and informational context, with the aim of supporting high-quality interaction [19]. ICVEs generally represent participants as avatars (a graphical representation of a human) in a shared computer-generated virtual environment [5]. Head-orientation and pointing gestures are animated by head and hand trackers worn by participants, and significantly contribute toward perception of other's visual attention [14]. However, ICVEs currently fail to track participant gaze. Murray and Roberts determined that augmenting avatar head-orientation with gaze (replayed from pre-recorded eye-tracking data) is of vital importance for observers to be able to correctly identify where an avatar is looking (gaze target) [17]. Uses of gaze which are particularly relevant to interaction in ICVEs include indicating and determining focus of attention, addressing, prompting, and when describing and handling objects.

We addressed the lack of participant gaze capture and representation in ICVEs by developing *EyeCVE*: a novel system with integrated mobile eye-trackers that drive avatar gaze in real-time, thus supporting natural awareness of other's visual attention and allowing eye-contact (mutual gaze) to be established between participants. Although the primary motivation for this work is to accurately *represent* gaze for enhanced remote social-interaction in ICVEs, the use of eye-trackers grant other opportunities directly and through data collection. This paper demonstrates interactional analysis to measure the system's ability to support remote multiparty virtual conferencing scenarios. We cover initial use, evaluation, and potential of *EyeCVE* as opposed to technical implementation described in [26].

EYE-GAZE IN MULTIPARTY MEDIATED COMMUNICATION

Representing gaze information has long been recognised as a requirement for natural communication through visual remote collaboration and conferencing systems [1]. This is a logical extension of Argyle's conviction that gaze is of central importance in social behaviour and non-verbal communication where it is used as a bidirectional channel monitoring initiation, maintenance and termination of messages [2].

When defining the fidelity of a system's gaze awareness conservation, a distinction can be made between what we term *attentional* and *communicational* gaze, which are informed by Vertegaal's requirements for such visual conferencing systems [23]. Attentional gaze infers only focus of attention (i.e. a head-orientation metaphor), and does not support mutual gaze due to for instance, a fragmented workspace or narrow field of view. The higher-fidelity communicational gaze locates attentional properties in a perceptually unified space, allowing fuller gaze awareness and mutual gaze. Previous systems developed to support gaze in multiparty remote interaction generally fall into three categories as discussed below. We cover an archetype VMC system supporting communicational gaze, before focusing on previous work in CVEs, and then addressing hybrid approaches.

Video-Mediated Communication

The MAJIC system [18] defines the standard method for supporting communicational gaze over VMC, in which relative position, head orientation and gaze are preserved during multiparty meetings. Cameras are positioned behind a thin half-transparent curved screen, on which life-size videos of other participants are seamlessly projected to simulate a three-way round-table meeting. Communicational gaze is achieved by positioning cameras behind the screen at the location of each participant's head. Several commercial "telepresence" systems (e.g. Cisco, LifeSize, Polycom, Telris) support communicational gaze similarly, but typically fix cameras above high-definition displays. Despite the offset angle between camera and remote participant display, mutual gaze can be supported by operating within Chen's threshold of $1°$ horizontal and $5°$ vertical [6]. Consequently, a major drawback of VMC is the difficulty to support people moving around the room or even shifting posture without causing parallax between camera position and the on-screen participant display, resulting in loss of gaze awareness.

Collaborative Virtual Environments and Avatars

CVEs generally represent participants in a shared virtual environment using avatars, so their design is a fundamental development issue [5]. Virtual humans are capable of eliciting appropriate responses from people, and it has been observed that unwritten social norms such as proxemics and unease from close-range mutual gaze with unknown partners occur in virtual environments as they do in real-life [4]. Models controlling avatar gaze have been developed [15, 25, 9], adopting a parametric, data-driven or hybrid approach to gaze calculation. Evaluation of Vinayagamoorthy et al.'s model [8] featured a two-way conversational task between participants represented by avatars exhibiting model, random or no gaze behaviour. Findings show that such models can simulate behaviour that significantly enhances perceived quality of communication. This informs a primary hypothesis behind this work: using eye-tracking to drive avatar gaze will lead to more natural social-interaction in ICVEs.

EyeCVE has been developed to operate between immersive CAVE displays, and this paper details a three-way conversational task similar to the dyadic scenario presented in [8]. Hindmarsh and colleagues suggested that some of the collaboration limitations such as lack of information about other's
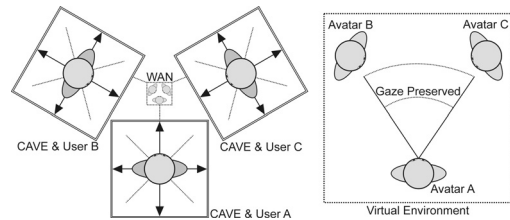

Figure 1. Linked CAVE users not rooted to the spot, with co-present avatars in *EyeCVE*'s shared VE with replicated eye-tracked gaze.

actions that are observed in non-immersive (i.e. desktop) CVE systems could be alleviated by ICVEs due to intuitive body tracking and the larger field-of-view [12]. Later studies have shown this to be the case, but there are still problems due to the representation of avatars: in particular, gaze is identified as an important cue not captured or transmitted faithfully by the medium [11]. Figure 1 illustrates the ability for unconstrained movement in a shared virtual space while maintaining meaningful gaze as supported by *EyeCVE*.

Hybrid Approaches

The non-immersive GAZE Groupware system supports attentional gaze by using photos to represent participants in a shared virtual environment [23]. Eye-trackers measure each person's gaze, and rotate the photos to indicate attention. The system was later enhanced to support communicational gaze (and posture-shift but not free-movement) using video streams and dynamic cameras to avoid parallax [24].

MOTIVATION

Evidence from referenced studies suggests that capturing and representing participant gaze will facilitate more natural social interaction in ICVEs. Although not specifically investigated in this experiment, the use of mobile head-mounted eye-trackers allow free-movement within the CAVE while maintaining communicational gaze between multiple participants without the parallax problematic to VMC systems. Integrating eye-tracking and ICVEs also invites exploration of other possible uses as illustrated in Figure 2.
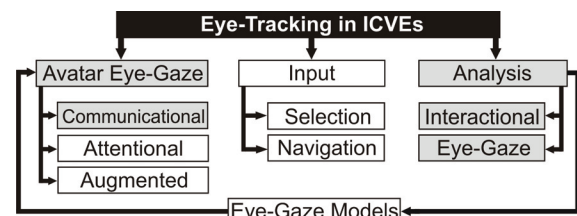

Figure 2. Applications of eye-tracking in ICVEs.

As indicated by Figure 2's shaded boxes, this paper focuses on communicational avatar gaze and analysis, while the unshaded boxes will be addressed in future work. Here, we differentiate *communicational* from *attentional* gaze by classifying the former as behaviour from raw eye-tracker data, while the latter will apply filters to the data to recognise eye-fixations and saccades in order to selectively represent focus of attention in a more iconic manner. Inferred from Hindmarsh et al. [12] and Garau et al. [8] respectively, we predict that varying gaze fidelity will benefit interaction depending on task (i.e. conversational or object-focused) and avatar

quality. Augmented gaze refers to communicating information not usually available in face-to-face interaction, such as indicating gaze target or manipulating perceived attention as seen in Bailenson et al. [3]. Advanced gaze models informed by multiparty analysis will subsequently be used to drive avatar gaze, allowing performance comparison with tracked gaze. Natural object selection and navigation in 3D space is also granted by using eye-trackers as an input device.

USER TRIALS

We used an informal interview scenario to collect data to evaluate *EyeCVE*'s support for three-way conferencing. Five participants were recruited from students at UCL (all male, one had previous CAVE experience) to be interviewed by confederates at the Salford and Reading Universities. Questions were designed to be engaging but not personal, and related to academic background. To minimise distractions that may influence participant gaze, the virtual environment depicted a white meeting room containing only a round-table. All parties sat on chairs placed in the centre of each CAVE, and wore stereo-glasses with integrated head and eye-trackers and held the hand-tracker in their right hand. Thus, each avatar's head, eyes, and right arm were animated based on natural movement. Audio communication was held over a *Skype^{TM}* conference call, and played through loud-speakers. Sessions were captured from each eye-tracker's scene-camera (wearer's perspective) with overlaid 2D gaze target (see Figure 3). Log-files were recorded for replay and analysis of the virtual scene [16].

Figure 3. *EyeCVE* session and scene-camera video with 2D gaze target.

ANALYSIS

In combination with gaze data from the three eye-trackers, our qualitative analysis of participants' interactional behaviour draws primarily on *conversation analysis*: an inductive process for analysing how human interaction is organised into sequences of action or systematic practices [20]. We limit analysis to the interviewees, although it will also be necessary to consider some of the interviewers' behaviour. Due to space limitations, we can only show one data extract, but full transcripts and videos will be made available online. In the extract, reference numbers relate to speech from the adjacent name. Various symbols and spellings are used to capture aspects of the sound of the talk. The interviewee's gaze is indicated by labeled lines below and parallel to the related transcribed talk. We consider the exchange of information and involvement of participants, including the ability to interpret attention from gaze as well as maintaining mutual gaze with others during conversation.

Supporting Eye-Gaze Practices

Certain gaze practices observed in co-located interaction are supported by *EyeCVE*: generally, participants look at the

speaker while being asked a question, look away when starting their response, and return gaze to the questioner as their answer reaches completion [7]. In Figure 4, as Rob asks a question, Ric gazes directly at him, and brings his gaze briefly to the other interviewer (Paul) as Rob's question comes to an end. During the brief silence before staring his response, and during the "Er" with which it begins, Ric's gaze is directed at the central space between the two interviewers' avatars. Ric then alternates his gaze to both avatars before returning gaze to the questioner (Rob) as his response ends.

```
#1.2 CAVE#06 (Simplified)
04  Rob:    Okay thanks er a what kind of person:
            Rob
05          (.) would you refuse to work with
            Rob                              Paul
06          (.)
            Centre
07  Ric:    Er:: (.) someone stubborn (.) that doesn't
            Centre        Rob      Paul
08          listen to others   (2.0)
            Paul          Rob
```

Figure 4. Conversation transcript illustrating gaze behaviour

To support the conversation analysis, we performed a quantitative measure by asking seven academics (from psychology and computer science) to classify participants' gaze targets while being asked questions (56 over the five sessions). To ensure inter-rater reliability, we calculated Fleiss' kappa coefficient of 0.67, constituting "substantial agreement" between response sets. Table 1 shows that participants' gaze targeted the questioner's avatar (as opposed to the other or neither avatar) in 66.7% of cases. "Ambiguous" refers to cases in which gaze was deemed too erratic for broad categorisation, often due to noise in the eye-tracker data.

Table 1. Participant gaze target while being asked a question.

Speaker	Other	Neither	Ambiguous
66.7%	5.6%	16.6%	11.1%

Management of speaker transition

Interaction and conversation management ran smoothly despite current limitations of the system such as no avatar lip-movement and non-spatialised audio. To compensate, interviewers adopted a naming and gesturing protocol to clarify questioner transition. This gestural element of turn-taking was often followed by the interviewee's gaze. Although somewhat artificial in comparison with co-located turn-allocation [20], their success highlights the importance of the shared and perceptually unified environment, which allowed gesture to be used as a communicational resource.

DISCUSSION

Avatar gaze mimicked participant gaze, which was generally found to follow standard gaze practices. Therefore, *EyeCVE* allowed participants to use gaze for bidirectional feedback and signaling similarly to face-to-face interaction. While being asked a question, participants gazed at the questioner in 66.7% of cases. This figure seems to have been deflated due to eye-tracker noise, which contributed to the 11.1% of ambiguous cases in which there was no agreement between raters as to where to the participant was looking. In any case, this frequency is comparable to Argyle's time-range of 70-75% that listeners gaze at speakers during dyadic face-to-face conversation [2].

Although our quantitative analysis did not expose obvious within-session effects, conversation analysis revealed behavioural changes in participants who initially gazed *generally* towards other avatars, but began to consistently target fellow avatar's eyes as the interview progressed. This suggests that some participants realised the significance of gaze information, and responded with interactionally relevant behaviour, despite not having been told that gaze was being captured for real-time representation. Gesture was important for conversation management and speaker transition, and was employed by interviewers to compensate for the lack of avatar lip-movement and spatialised audio. Participant gaze was sensitive to gesture, and conversation flowed smoothly. *EyeCVE*'s support for such observed emergent learning and communicative behaviour will be explored in future work.

CONCLUSIONS AND FUTURE WORK

We presented *EyeCVE*, our multiparty CAVE-based system using mobile eye-trackers to drive avatar gaze, thus supporting communicational gaze in an unfragmented workspace. The capture of nonverbal behaviour is considered essential in order to support collaborative interactions that are more similar to those in the real world [21], and the system augments standard ICVE replication of head and hand movement with gaze. While further investigation and addressing current shortfalls is required, our system enables communicational gaze and free movement in an unfragmented workspace, in which multiple participants can be co-present together with virtual objects and documents, thus supporting remote cooperative working.

Future work will focus on content outlined in Figure 2, and will be evaluated in dynamic and object-focused tasks common to CSCW. We will enhance our software framework and robustness of eye-tracking in the CAVE to reduce gaze target ambiguities. High-fidelity avatars as detailed in [22] featuring facial expression, lip-movement and accurate eye-animation will be integrated, allowing investigation into behavioural and representational fidelity. Finally, we will explore the impact of gaze communication in ICVEs, establishing situations where it is critical for remote collaboration and its use as a tool for interactional analysis.

ACKNOWLEDGMENTS

REFERENCES

1. S. Acker and S. Levitt. Designing Videoconference Facilities for Improved Eye Contact. *J. Broad. & Elec. Media*, 31:181, 1987.

2. M. Argyle and M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press Cambridge, 1976.

3. J. Bailenson, A. Beall, J. Loomis, J. Blascovich, and M. Turk. Transformed Social Interaction: Decoupling Representation from Behavior and Form in Collaborative Virtual Environments. *Presence: Teleoperators & Virtual Environments*, 13(4):428–441, 2004.

4. J. Bailenson, J. Blascovich, A. Beall, and J. Loomis. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence: Teleoperators & Virtual Environments*, 10(6):583–598, 2001.

5. S. Benford, J. Bowers, L. Fahlén, C. Greenhalgh, and D. Snowdon. User embodiment in collaborative virtual environments. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 242–249, 1995.

6. M. Chen. Leveraging the asymmetric sensitivity of eye contact for videoconference. *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, pages 49–56, 2002.

7. S. Duncan. Some signals and rules for taking speaking turns in conversations. *Nonverbal Communication: Readings with Commentary*, 1974.

8. M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. *Proceedings of the conference on Human factors in computing systems*, pages 529–536, 2003.

9. E. Gu and N. Badler. Visual Attention and Eye Gaze During Multiparty Conversations with Distractions. *Intelligent Virtual Agents*, 2006.

10. J. Hauber, H. Regenbrecht, M. Billinghurst, and A. Cockburn. Spatiality in videoconferencing: trade-offs between efficiency and social presence. *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 413–422, 2006.

11. I. Heldal, A. Steed, M. Spante, et al. Successes and Failures in Co-Present Situations. *Presence: Teleoperators & Virtual Environments*, 14(5):563–579, 2005.

12. J. Hindmarsh, M. Fraser, C. Heath, S. Benford, and C. Greenhalgh. Object-focused interaction in collaborative virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(4):477–509, 2000.

13. E. Isaacs and J. Tang. What video can and cannot do for collaboration: A case study. *Multimedia Systems*, 2(2):63–73, 1994.

14. S. Langton, R. Watt, and V. Bruce. Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, 4(2):50–59, 2000.

15. S. Lee, J. Badler, and N. Badler. Eyes alive. *ACM Transactions on Graphics (TOG)*, 21(3):637–644, 2002.

16. A. Murgia, R. Wolff, W. Steptoe, et al. A Tool For Replay And Analysis of Gaze-Enhanced Multiparty Sessions Captured in Immersive Collaborative Environments. *ACM DS-RT'08*, 2008.

17. N. Murray and D. Roberts. Comparison of head gaze and head and eye gaze within an immersive environment. *the 10th IEEE International Symposium on Distributed Simulation and Real Time Applications. IEEE Computer Society, Los Alamitos, CA*, 2006.

18. K. Okada, F. Maeda, Y. Ichikawaa, and Y. Matsushita. Multiparty videoconferencing at virtual social distance: MAJIC design. *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 385–393, 1994.

19. D. Roberts, R. Wolff, O. Otto, and A. Steed. Constructing a Gazebo: Supporting Teamwork in a Tightly Coupled, Distributed Task in Virtual Reality. *Presence: Teleoperators & Virtual Environments*, 12(6):644–657, 2003.

20. H. Sacks. *Lectures on Conversation*. Blackwell, 1995.

21. A. Steed, D. Roberts, R. Schroeder, and I. Heldal. Interaction between Users of Immersion Projection Technology Systems. *HCI International 2005, the 11th International Conference on Human Computer Interaction*, pages 22–27, 2005.

22. W. Steptoe and A. Steed. High-Fidelity Avatar Eye-Representation. *Virtual Reality Conference, 2008. VR'08. IEEE*, pages 111–114, 2008.

23. R. Vertegaal. The GAZE groupware system: mediating joint attention in multiparty communication and collaboration. *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, pages 294–301, 1999.

24. R. Vertegaal, I. Weevers, C. Sohn, and C. Cheung. GAZE-2: conveying eye contact in group video conferencing using eye-controlled camera direction. *Proceedings of the conference on Human factors in computing systems*, pages 521–528, 2003.

25. V. Vinayagamoorthy, M. Garau, A. Steed, and M. Slater. An Eye Gaze Model for Dyadic Interaction in an Immersive Virtual Environment: Practice and Experience. *Computer Graphics Forum*, 23(1):1–11, 2004.

26. R. Wolff, D.Roberts, A. Murgia, et al. Communicating Eye Gaze across a Distance without Rooting Participants to the Spot. *ACM DS-RT'08*, 2008.