Using Virtual Reality to Investigate the Emergence of Gaze Conventions in Interpersonal Coordination

Mills, Gregory; Boschker, Remko

Published in:
Human-Computer Interaction – INTERACT 2023

DOI:
10.1007/978-3-031-42293-5_76

Using Virtual Reality to Investigate the Emergence of Gaze Conventions in Interpersonal Coordination

Gregory Mills[1,2(✉)] and Remko Boschker[2]

[1] School of Computer Science and Mathematics, Kingston University, London, UK
g.mills@kingston.ac.uk

[2] Centre for Language and Cognition, University of Groningen, Oude Kijk in't Jatstraat 26, 9712 EK Groningen, Netherlands

Abstract. Gaze plays a central role in regulating turn-taking, but it is currently unclear whether the turn-taking signals of eye gaze are static and fixed, or whether they can be negotiated by participants during interaction. To address this question, participants play a novel collaborative task, in virtual reality. The task is played by 3 participants, and is inspired by games such as Guitar hero, Rock Band, Beat Saber, and Dance-Dance Revolution. Crucially, the participants are not allowed to use natural language – they may only communicate by looking at each other. Solving the task requires that participants bootstrap a communication system, solely through using their gaze patterns. The results show that participants rapidly conventionalise idiosyncratic routines for coordinating the timing and sequencing of their gaze patterns. This suggests that the turn-taking function of eye-gaze can be flexibly negotiated by interlocutors during interaction.

Keywords: Dialogue · Transformed Social Interaction · Eye-gaze · Turn-taking

1 Introduction

When people speak with each other, they dynamically adapt their language to that of their conversational partner (Pickering and Garrod 2004; Clark 1996). A central finding in dialogue research is that the meanings of words and phrases used are negotiated ad hoc by participants. In addition to natural language expressions, face-to-face conversation is underpinned by myriad non-verbal signals which are used, inter-alia, to regulate procedural coordination in the interaction. For example, speakers tend to look away from their addressee when starting to speak, and then re-establish eye-contact at the end of their turn in order to yield the floor or signal the next speaker (Kendon 1967; Degutyte and Astell 2021). Although research has shown clear cultural differences in such gaze-behaviour (Rossano et al. 2019), it is currently unclear whether the communicative meaning of eye-gaze is static and fixed, or whether, like natural language, it might be dynamically negotiated by participants during interaction.

To address this question, participants play a novel collaborative task within a virtual reality environment which allows for testing whether and how idiosyncratic eye-gaze signals might emerge.

J. Abdelnour Nocera et al. (Eds.): INTERACT 2023, LNCS 14145, pp. 580–584, 2023.
https://doi.org/10.1007/978-3-031-42293-5_76

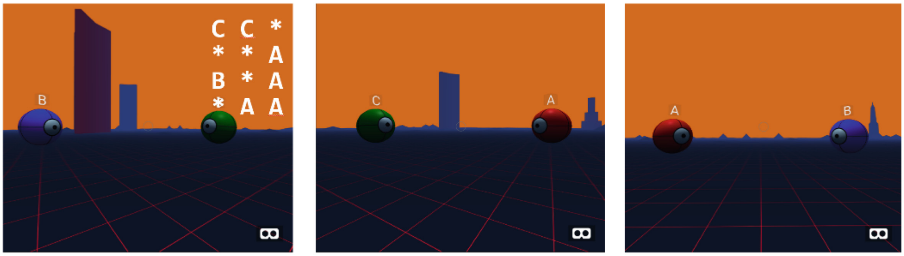

Fig. 1. The view from each of the three participants' headsets (From left to right: Participants A, B, C).

Participants are rendered as virtual eye-balls. In this example, Participant A is assigned the role of Director, The target sequence of "look events" is displayed as a three-column table in the top-right hand corner of A's display. The table is read from top to bottom. The left-most column cells represent the actions to be performed by A. The middle column represents the actions to be performed by B, and the right-most column represents the actions to be performed by C. Each row describes a gaze configuration that must be achieved simultaneously by the triad. The target sequence represents the following sequence of actions: "*First A and B both need to look at C* (row 1) *Then C needs to look at A.* (row 2) *Then while C looks at A, B needs to look at A.* (row 3) *Then B needs to look at A* (row 4)". The task of the Director is to get the triad to perform this sequence of look events.

2 Methods

2.1 The Task

Groups of 3 participants play a collaborative task[1], in virtual reality, using Oculus Go headsets. Participants, who are rendered as "eye-ball" avatars, are placed equidistantly and facing each other in a virtual environment (see Fig. 1, above). The task is inspired by games such as Guitar Hero, Rock Band, and Dance-Dance Revolution. The three key differences are:

1. Instead of performing target sequences of musical notes or dance moves, each triad needs to perform, together, sequences of gaze events. For example, a typical target sequence might be: "*B must look at C. Then C must look at A. Then, while C continues looking at A, A and B must look at each other*".
2. On each trial, only one participant (the Director) sees the target sequence. This means that in order for the group to complete the target sequence, the Director has to instruct the other participants.
3. Crucially, the participants are not allowed to use natural language to communicate – they may only communicate by looking at each other.

[1] The source-code is available at https://github.com/gjmills/VRLookingGame.

Successfully solving target sequences requires that triads bootstrap an ad hoc communication system (see, e.g., Nölle and Galantucci 2022; Stevens and Roberts 2019) for instructing and taking turns, solely using their gaze patterns[2].

2.2 Manipulation

In order to test whether participants develop idiosyncratic signals for coordinating procedurally, the experiment was divided into a 25 min "training phase" followed by a 5 min "test phase". During the training phase, triads complete the task as described above. At the start of the test-phase, the identities of the participants were swapped: In Participant A's headset, Participant B's physical head movements are mapped onto Participant C's avatar, while Participant C's physical head movements are mapped onto B's avatar. Similarly, for B and C. This manipulation in the test-phase investigates whether participants within the triads develop a different communication system with each partner: participants are unaware that the identities of their partners are swapped, so if they have indeed established different systems, then, on entering the test phase, they will attempt to reuse a convention with the same partner (who is actually the other partner), leading to more errors and less efficient communication.

2.3 Hypotheses

The experiment tested two hypotheses:

1. During the training phase, participants will establish a communication system with each other that will allow them to collaboratively solve the target sequences.
2. In the test phase, the manipulation will cause participants to inadvertently use the wrong signals with each other, causing disruption to task performance.

3 Results

3.1 Training Phase

During the 25-min training phase, triads completed a mean of 20.5 sets (S.D. = 3.45). The most successful triad completed 27 sets. By the end of the training phase, triads were solving sets with a mean of 5.5 target items (S.D. = 1.2). The most successful triad completed sets containing 8 targets (see, e.g., Fig. 2 which shows a target set containing 7 "look events").

3.2 Test Phase

To test the effect of the intervention, we compared participants' performance in the 5 min preceding the swap with their performance during the 5-min test phase. We used two measures of disruption to task performance.

The first measure, task success, was modelled with a mixed binary logistic regression, using the lme4 package, which showed that triads solved significantly fewer games in

[2] See https://youtu.be/ctXXtFBr6Cc for a video of participants playing the game.

the test phase ($b = -0.49$, S.E. = 0.193, $z = -2.54$), $p = 0.0111$). The model predicts that triads successfully solve 66% [95% CI: 0.60, 0.72] of target sets in the training phase and 54% [95% CI: 0.48, 0.61] of target sets in the test phase.

The second measure recorded the number of "look events" per game, i.e., the number of times a participant selected a target. All things being equal, if participants are encountering more difficulties coordinating with each other, this will lead to them having to make more selections, i.e., expend more effort, to solve a set. A linear mixed model using the lme4 package showed that triads produced significantly more look events in the 5-min test phase than in the last 5 min of the training phase ($b = 10.4$, S.E. = 2.98, $t = 3.5$, $p < 0.001$). The model predicts 40 [95%CI: 36.2, 43.8] look events per game in the training phase, and 50.4 [95% CI: 45.5, 55.4] look events in the test phase.

4 Discussion

The results provide support for both hypotheses. The average sequence length at the end of the training phase suggests that the participants were solving the sets by communicating with each other, as opposed to solving via individual trial and error.

Moreover, the increased number of timeouts and look events in the test phase suggest that the manipulation disrupted participants' coordination. A plausible explanation for this pattern is that many participants communicated differently with each partner. This was confirmed by the participants themselves. On debriefing, we asked participants about the communication system they had developed. Some participants explicitly stated that they noticed that their partners communicated differently (e.g., using different signals for the same actions, or communicated faster/slower), which they had attempted to accommodate.

These findings are subject to a couple of important caveats: First, the participants' movements are severely constrained. The Oculus Go headsets only capture rotations around the x, y, z axes, but do not capture any change in location: throughout the experiment, the avatars are anchored at a fixed location. Second, the setup conflates "head gaze" and "eye gaze", as participants' head-movements are mapped onto their virtual eye-ball (see, e.g., Špakov et al. 2019).

Nonetheless, these findings suggest that the interactive signals that participants use to attract and direct another's visual attention can be flexibly negotiated during an interaction.

To conclude, these findings are of central importance for theories of Human-Computer Interaction. Research on dialogue has shown that in order for systems to converse naturalistically with humans, they must be able to dynamically adapt their vocabularies, ontologies, and emotional signals to their conversational partner (Healey et al. 2021; Mills 2014; Mills et al. 2021). The findings from the current experiment suggest that, in addition, technologies such as avatars, dialogue systems, as well as self-driving cars when communicating with pedestrians (Habibovic et al. 2018), need to be able to flexibly adapt their non-verbal and turn-taking signals to those of the user.

References

Clark, H.: Using language. Cambridge University Press, Cambridge (1996)

Degutyte, Z., Astell, A.: The role of eye gaze in regulating turn taking in conversations: a systematized review of methods and findings. Front. Psychol. **12**, 616471 (2021)

Habibovic, A., et al.: Communicating intent of automated vehicles to pedestrians. Front. Psychol. **9**, 1336 (2018)

Healey, P., Muggleton, S., Chater, N.: Human-Like Communication. Oxford University Press, Oxford, England (2021)

Kendon, A.: Some functions of gaze-direction in social interaction. Acta Physiol. (Oxf) **26**, 22–63 (1967)

Mills, G.J.: Dialogue in joint activity: complementarity, convergence and conventionalization. New Ideas Psychol. **32**, 158–173 (2014)

Mills, G., Gregoromichelaki, E., Howes, C., Maraev, V.: Influencing laughter with AI-mediated communication. Interact. Stud. **22**(3), 416–463 (2021)

Nölle, J., Galantucci, B.: Experimental semiotics: past, present and future. In Garcia & Ibanez. Routledge Handbook of Neurosemiotics (2022)

Pickering, M.J., Garrod, S.: Toward a mechanistic psychology of dialogue. Behav. Brain Sci. **27**(2), 169–190 (2004)

Rossano, F., Brown, P., Levinson, S.C.: Gaze, questioning and culture. Convers. Anal. **27**, 187–249 (2009). https://doi.org/10.1017/CBO9780511635670.008

Špakov, O., Istance, H., Räihä, K.J., Viitanen, T., Siirtola, H.: Eye gaze and head gaze in collaborative games. In: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, pp. 1–9 (2019)

Stevens, J.S., Roberts, G.: Noise, economy, and the emergence of information structure in a laboratory language. Cogn. Sci. **43**(2), e12717 (2019)