Eye gaze in interaction: towards an annotation scheme for dialogue

Vidya Somashekarappa, Christine Howes and Asad Sayeed
Centre for Linguistic Theory and Studies in Probability (CLASP)
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg
{vidya.somashekarappa,christine.howes,asad.sayeed}@gu.se

Abstract

This paper proposes an approach to annotating eye gaze in natural dialogue which takes into account both social and referential functions of eye gaze and how they interact. The goal of this research is to provide a basis for robot or avatar models which communicate with humans using multimodal natural dialogue.

1 Introduction

Linguists and psychologists have shown a long standing interest in non-verbal communication relating to speech and gesture, including eye-gaze, which is the focus of this work (Kendon, 1967; Argyle and Cook, 1976; Goodwin, 1980, 1981).

1.1 Social functions of eye gaze in dialogue

Argyle and Cook (1976) showed that listeners display longer sequences of uninterrupted gaze towards the speaker, while speakers tended to shift their gaze towards and away from the listener quite often. Later work has refined this observation, with, for instance (Rossano, 2012) noting that this distributional pattern is dependent on the specific interactional activities of the participants, for example, a more sustained gaze is necessary in activities such as questions and stories, since gaze is viewed as a display of attention and engagement. (Brône et al., 2017) also found that different dialogue acts typically display specific gaze events, from both speakers' and hearers' perspectives.

Unaddressed participants also display interesting gaze behaviour showing that they anticipate turn shifts between primary participants by looking towards the projected next speaker before the completion of the ongoing turn (Holler and Kendrick, 2015). This may be because gaze has a 'floor apportionment' function, where gaze aversion can be observed in a speaker briefly after taking their turn before returning gaze to their primary recipient closer to turn completion (Kendon, 1967; Brône et al., 2017).

1.2 Referential functions of eye gaze in dialogue

Previous studies have tended to focus on either social functions of gaze (e.g., turn-taking or other interaction management) or how gaze is used in reference resolution, with few researchers combining these.

The process of identifying application-specific entities which are referred to by linguistic expressions is reference resolution. One example is identifying an image on a display by referring to "the painting of a night sky". One area in which multimodal reference resolution has been previously studied is in the context of sentence processing and workload. For example, Sekicki and Staudte (2018) showed that referential gaze cues reduce linguistic cognitive load. Earlier work (e.g., Hanna and Brennan, 2007) showed that gaze acts as an early disambiguator of referring expressions in language.

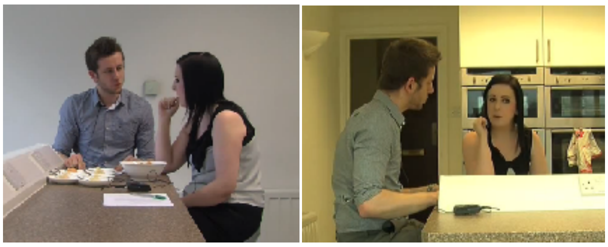
Campana et al. (2002) proposed to combine reference resolution component of a simulated robot with eye tracking information; they intended to deploy this on the International Space Station. Unfortunately, eye movements' integration with speech was not addressed. Also, eye gaze information was used only in case of inability to identify unique referenced objects. Zhang et al. (2004) implemented reference resolution by integrating a probabilistic framework with speech and eye gaze; results showed an increase in performance. They also found that reference resolution of eye gaze could also compensate for lack of domain modelling.

2 Research questions

- Annotation – is it feasible (and can we automate some or all of it using machine learning techniques)?
- Can we classify elements of the dialogue based only on gaze behaviours? (Dialogue acts? Turn-taking? Reference objects?)
- Can we come up with an implementable model of gaze in dialogue for a conversational robot or avatar to interpret human gaze behaviour and produce human-like gaze behaviour?

3 Data

The data used in this pilot come from the case study reported in Lavia et al. (2018). Data consists of videos of pairs of participants (staff at the Good Housekeeping Institute) taste-testing eight different kinds of hummus. Participants are seated at a right-angle to each other, with separate cameras and radio microphones capturing each participant (see figure 1), providing a clear recording of eye movements, facial expressions, gestures and speech.



(a) View from camera 1 (b) View from camera 2

Figure 1: The two camera views

3.1 Annotation

Multimodal video annotation software ELAN will be used for manual analysis. Each of the annotations are entered in tiers and are assigned to a particular time interval. Speech of each participant will be annotated in different tiers as Speech1 and Speech2 (contains transcription of speech and laughter). This will be followed by four additional tiers focusing on the eye gaze. A joint attention tier displays the information of participants looking at a particular object/place at the same time and what exactly they are paying attention to. Mutual gaze tiers records the eye gaze of participant 1 (P1) looking at participant 2 (P2) and vise versa.

The final two tiers are dedicated to random eye gaze information of each participant when they are not involved in Mutual gaze or Joint attention (Random1 and Random2 for participant 1 and participant 2 respectively).

Annotating the speech along with exclusive eye gaze data would help in understanding the dialogue acts elementary to the non verbal yet-obvious interpretations of speech such as referencing, providing subtle cues to organise and control communication, conveying feedback and coordinating turn taking behaviours during speech overlaps. It is also interesting to look into the influence of disagreement in the rating which is persist over the entire conversation influencing fairness and measure how much of this capitulate behaviour is observed through eye gaze. This could help us understand much more about coordinated opinions and gaze switching soon after joint attention.

The task in the video as mentioned earlier is to rate the various hummus. The eye gaze information linked with emotion driven attention could help explore more of the constantly changing opinion of a participant to go along with the partner's stronger perspective. Also, what are the eye movement patterns during such situations and how does it affect the entirety of rating.

4 Discussion

Looking at all the different forms of non verbal communication, eye gaze is very powerful, but even so, we are rarely consciously aware of it. But we are at the verge of breakthroughs in building virtual human avatars, and now, more than ever, it is important to have them behave in more natural ways. Another application example of where this might help is in the area of virtual teleconferencing, by using user gaze information to enhance participant interaction through the conferencing software user interface. As we have discussed above, there is still a need to expand the part of the multimodal dialogue systems literature that focuses on building effective computational models on how people make use of gaze in ordinary conversations.

References

Michael Argyle and Mark Cook. 1976. *Gaze and mutual gaze*. Cambridge University Press.

Geert Brône, Bert Oben, Annelies Jehoul, Jelena Vranjes, and Kurt Feyaerts. 2017. Eye gaze and viewpoint in multimodal interaction management. *Cognitive Linguistics*, 28(3):449–483.

Ellen Campana, Beth Ann Hockey, Jason Baldridge, Roger Remington, John Dowding, and Leland S. Stone. 2002. Using eye movements to determine referents in a spoken dialogue system. *Proceedings of the 2001 Workshop on Perceptive User Interfaces*.

C. Goodwin. 1981. *Conversational organization: Interaction between speakers and hearers*. Academic Press, New York.

Charles Goodwin. 1980. Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociological inquiry*, 50(3-4):272–302.

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. Language-Vision Interaction.

Judith Holler and Kobin H Kendrick. 2015. Unaddressed participants gaze in multi-person interaction: optimizing recipiency. *Frontiers in psychology*, 6(98):1–14.

Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26:22–63.

Lisa Lavia, Harry J. Witchel, Francesco Aletta, Jochen Steffens, André Fiebig, Jian Kang, Christine Howes, and Patrick G. T. Healey. 2018. Non-participant observation methods for soundscape design and urban planning. In Francesco Aletta and Jieling Xiao, editors, *Handbook of Research on Perception-Driven Approaches to Urban Assessment and Design*. IGI Global.

Federico Rossano. 2012. *Gaze behavior in face-to-face interaction*. Ph.D. thesis, Radboud University Nijmegen Nijmegen.

Mirjana Sekicki and Maria Staudte. 2018. Eye'll help you out! How the gaze cue reduces the cognitive load required for reference processing. *Cognitive Science*, 42(8):2418–2458.

Qiaohui Zhang, Atsumi Imamiya, Kentaro Go, and Xiaoyang Mao. 2004. Overriding errors in a speech and gaze multimodal architecture. pages 346–348.