

INTRODUCTION

This study aims at developing and implementing the “observer gaze” methodology in order to study the structural organization of multimodal, face-to-face conversations. Tracking the gaze of a third-party human observer can be used to learn about human real-time expectations of the dialogue and provide an online and implicit method for investigating the dynamics of interactions and the relevant features that capture the observer attention according to the amount of information the observer has available and her/his ability to understand the language. The way an outsider visually perceives the interactional exchanges may also suggest a better modelling of human-human communication for human-machine communication systems. Proto-type versions of the method have proven promising, but the techniques are still rough and in need of development, quantification, and validation. The current study focusses on this and similar, relevant methods of acquiring annotation and labels of recordings of human interaction.

METHODOLOGY

Participants

- 10 participants
- Native speakers of Persian

Task

- Participants (observers) were asked to watch a short clip of a conversation (around 2 mins long) while their gaze was being recorded.
- To keep the attention of the individuals, they were told that after the conversation they will be given a questionnaire to fill about the conversation

Setting

- The trials took place in a soundproof room while it was administered from a separate room (next to it)
- First the participants eye gaze was calibrated and later using this calibration their gaze was recorded during the experiment
- They watched the clip in two settings: sound-on and sound-off
- The same clip was used for everyone to keep the experimental setting constant and to avoid influential factors



Fig 1: The conversational scene observed by the participants



REFERENCES

- De Ruiter, J.P., Mitterer, H., and Enfield. N.J. 2006. Projecting the end of a speaker’s turn: A cognitive cornerstone of conversation. *Language*, 82(3):515.
- Edlund J, Alexandersson S, Beskow J, Gustavsson L, Heldner M, Hjalmarsson A, et al. 3rd party observer gaze as a continuous measure of dialogue flow. In: *Proceedings LREC 2012*. Istanbul, Turkey: LREC; 2012.
- The eye gaze of 3rd party observers reflects turn-end boundary projection. *SemDial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue*, pages 204–205. Los Angeles, California, 21–23 September 2011.
- Tice, M., & Henetz, T. (2011). Turn-boundary projection: Looking ahead. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 838-843). Austin, TX: Cognitive Science Society

RESULT AND CONCLUSION

General overview of gaze fixation time

Figure 2 illustrates the visual attention in each experimental setup. In both conditions, the faces (R1 and R3) of the conversants were of the most interest to the observers. They spend much less time in regions two and four (R2 and R4) which corresponded to the conversants’ hand gestures and body movements.

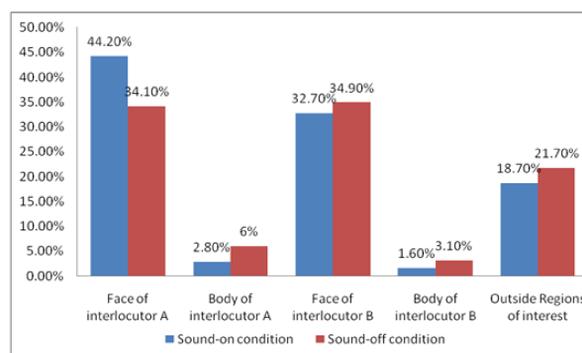


Fig 2: Average percentage of gaze fixation time distributed over conditions and ROIs.

Trajectory of gaze

The most frequently observed sequential pattern of gaze trajectory was R1 – R2 both in audio-visual and visual-only setup meaning that a gaze hit in region one was followed by a gaze hit in region two. Such pattern shows that the observers were mostly screening interlocutor A by consecutively looking at his face followed by his body. This might be explained by the fact that interlocutor A was the more dynamic participant, both verbally and non-verbally, in the conversation.

Gazing behavior at turn-transition points

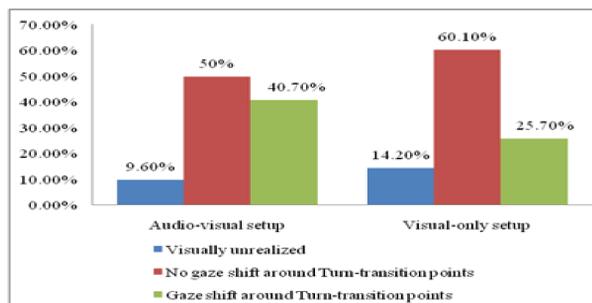


Fig. 3: Gazing behavior of observers around turn transition points across conditions

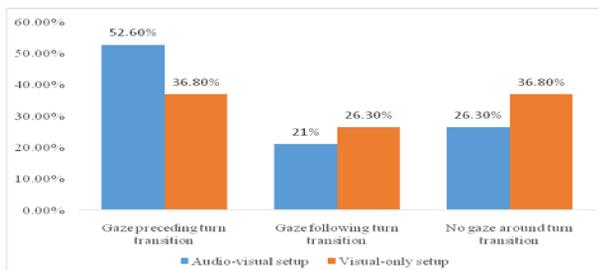


Fig. 4: Gaze anticipating or following turn-transitions

The observers were more visually involved with turn transition in the audio-visual setup presenting the highest number of gaze shifts during the transitions. However, the data shows that regardless of the conditions, the observers more likely anticipated the turn transitions rather than following them. The highest rate of anticipation belonged to audio-visual setup, highlighting the important role of lexio-syntactic information for projecting turn transition points.