A Model of Joint Attention for Humans and Machines

Nadine Pfeiffer-Leßmann (nlessman@techfak.uni-bielefeld.de),
Thies Pfeiffer, Ipke Wachsmuth

Collaborative Research Centre 673, “Alignment in Communication”,
A.I. Group, Faculty of Technology, Bielefeld University

Abstract

Joint attention is the simultaneous allocation of attention to a target as a consequence of attending to each other’s attentional states. It is an important prerequisite for successful interaction and supports the grounding of future actions.

Cognitive modeling of joint attention in a virtual agent requires an *operational model* for behavior recognition and production. To this end, we created a declarative four-phase-model (*initiate* / *respond* / *feedback* / *focus*) of joint attention based on a literature review. We applied this model to gaze communication and implemented it in the cognitive architecture of our virtual agent Max. To substantiate the model regarding the natural timing of gaze behavior, we conducted a study on human-agent interactions in immersive virtual reality. The results show that participants preferred the agent to exhibit a timing behavior similar to their own.

Building on these insights, we now aim at a process model of joint attention. We are interested in patterns of joint attention emerging in natural interactions. In the preliminary results of a human-human study, we find patterns of fixation targets and fixation durations that allow us to identify the four phases and infer the current state of joint attention.

Method

In the human-machine study we focused on the timing during individual phases and therefore prescribed the procedure to establish joint attention. The second study now focuses on the sequence of phases. In this human-human study, we still concentrate on gaze communication and thus constrained participants to use this modality only. In each trial, two participants are asked to identify one of 23 Lego figures based on the experimenter’s description (e.g. “the figure has blue trousers and spectacles”). Each participant is faced by only half of the figures and thus only one of the two is able to fully identify the target. This practical consequence is not made explicit beforehand. As a joint goal, the participants have to agree upon the target non-verbally and each of them has to write down the answer individually. Eye movements of one participant from each trial are tracked and the interaction showing both participants is video-taped.

Preliminary Results

So far, 6 participants took part in our study and altogether 3690 fixations were recorded and annotated. A total of 20 participants is planned. Figure 1 shows the durations of fixations as a function of time differenced regarding the fixation targets: partner and target. The most interesting part of the interaction is the time before the interlocutors establish joint attention on the presumed target. The origin of time 0 is thus aligned to the event marking this decision and all previous events are thus ordered along a negative time scale. As an unambiguous event marking the final decision, we chose the first contact of the pen with the paper when the participants write down their decision.

The left side of Figure 1 depicts the graph for the participants that had all necessary information to identify the target and thus had to initiate joint attention with the interlocutor. As expected, we find increased fixation durations towards the target object, especially in the last 10 seconds before the decision. The time course of the participants that needed to engage in joint attention to identify the correct target is shown in the right part of Figure 1 and complements these findings. Participants showed longer fixation durations towards the interlocutor more than 10 seconds before the decision (waiting for *initiate*). In the last 10 seconds, the target object receives longer fixations (*respond*). In the last three seconds, the participants again exhibited longer fixations on the partner (*feedback*) before they wrote down their decision (*focus*).

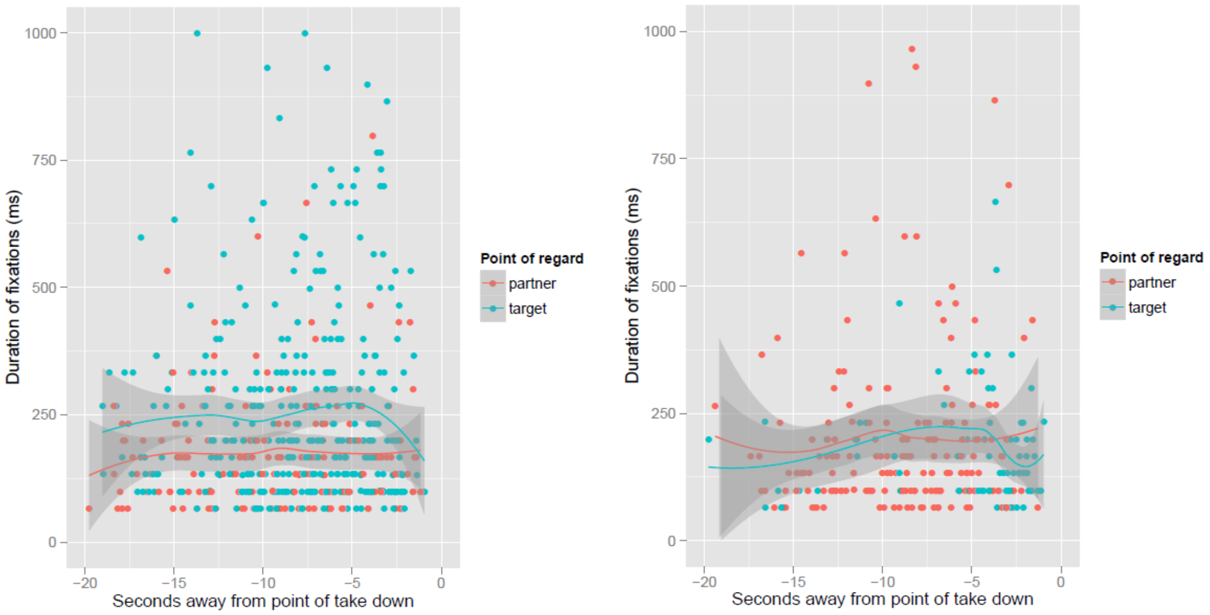

Figure 1: The diagram shows the time course of fixation durations during the last 20 seconds before participants write down their decision. *Left*: the participant can identify the target and has to communicate the reference. *Right*: the participant has to rely on joint attention to identify the target.

Conclusion

While our first study was targeted at the timing of the individual phases of joint attention, this second study is focusing on their time course and sequence. The preliminary results show that fixation durations are actively used to establish joint attention and are thus a plausible index to identify the relevance of the fixation target for the current phase in the joint attention model. However, in natural interactions an individual phase is not realized by a single clear-cut fixation exchange (one look at the partner, one extended look at the object, one look at the partner), but can be an iteration of multiple gaze exchanges. We are now collecting and annotating more data to provide grounds for a more precise modeling of this joint attention process. Data collection will be completed in May 2013. This will finally allow us to monitor the progress of joint attention in human-agent interaction.