

Turn-taking patterns in self-disclosure interactions with Virtual Agents

Sin-Hwa Kang¹, Candy Sidner², Jonathan Gratch¹ and Louis-Philippe Morency¹

¹ Institute for Creative Technologies, University of Southern California, USA

²Dept of Computer Science, Worcester Polytechnic Institute, USA

{kang, gratch, morency}@ict.usc.edu, sidner@wpi.edu

In typical communication situations, it is desirable to avoid any type of simultaneous talking due to lack of coordination between communicators, as it is not easy to maintain sufficient mutual clarity over conversation at the same time. Researchers have long commented on the lack of coordination in the turn-taking of conversation partners [2]. In our previous study [5] we mainly investigated people's self-disclosure in the interview interaction with real human videos and virtual agents. Bavelas et al. [1] demonstrated that collaboration between a speaker's acts and listeners' responses were coordinated by speaker gaze. Exline et al. [3] described the head direction, specifically looking at or away from another, as a natural gesture in which communicators normally perceive the other's intentions based on the fixations or avoidances of gaze. Therefore, we speculated that it would be crucial to study the timely exchange of speaking turns coordinated by interviewee gaze in the self-disclosure interview interaction, where participants were human interviewees and the interviewers were either virtual agents or humans represented by a modified or unmodified video avatar.

The basic experimental design is a 3 condition between-subjects experiment. The 3 conditions involved three levels of avatar realism – an unmodified human video avatar, a modified human video avatar which was displayed using the outline of a human figure, and a virtual agent (Rapport Agent created by Gratch et al. [4]). One hundred and eight subjects were randomly assigned to one of three experimental conditions. Subjects were asked to complete a pre- and a post-questionnaire. In the interaction sessions, the confederate or the Rapport Agent was an interviewer who asked ten questions requiring self-disclosure of the subject. The typical conversation was allowed to last about thirty minutes. To generate the avatar movement of the human video avatars, a web-cam (Logitech QuickCam Orbit MP) captured subjects' facial expressions. To allow video chat conversation, video conference software (Skype) was used. For the virtual agents, the subjects were informed that the Rapport Agent was an avatar controlled by another subject. To analyze turn-taking patterns, a coder annotated interactants' verbal responses and coded matched turn-taking behaviors on the responses. These included head direction, gaze, head nods and head wags, smiles, utterances and utterance gaps.

We analyzed the data of twelve participants (four participants in each condition) to find interactants' turn-taking patterns. We found that participants looked at the virtual agent with sufficient time at the end of their turn to indicate their attention to the virtual agent. That is, the participant turned their gaze to the agent on average 1.6 seconds before the end of their turn. The duration of their turning gaze helped the agent notice that the participants were finishing their turn. This is supported by finding that the participants also generally provided a prosodic drop when they gazed at

the interviewer at the end of utterance that ended their turn. We note that we do not yet know the minimum duration of gaze to the next speaker for smooth turn-takings, so this could be another subject for future studies. We also found that both types of human interviewers produced head nods during the interviewee's turn (backchannels) as well as at the end of the turn. The end of a turn nod seemed to signal understanding or empathy with the interviewee and his/her comments. The end of turn nods preceded looking down at the question sheet which the interviewers used. The virtual agent was pre-programmed to produce these same behaviors.

One finding about ends of turns and the modified avatar surprised us. When they interacted with the modified avatar human interviewer, interviewees sometimes did not look at their partner until after giving up the turn, which provided little non-verbal notification of the end of their turn. Specifically, we observed that interviewees with the modified avatar produced ten instances total, spread across all four pairs, of looking at their partner only after they ended their turn; the range of times was .4 to 1.2 seconds after the end of speaking. We observed five instances of this behavior in two of the four normal human - human pairs (range .7 to 1.1 seconds), and none in the virtual agent-human pairs. The modified avatar provided less social cues than the normal human interviewer while of the three conditions, but the virtual agent presented the least social cues as it did not show facial expressions. Since the virtual agent, like the both human interviewers, produced timely non-verbal feedback in the form of head nods, we speculate that the modified avatar interviewer on the whole provided less visual information than the virtual agent, which led the interviewee to direct less attention to it.

We also looked at additional differences in patterns between the human interviewer and the agent interviewer so that we can apply the appropriate turn-taking patterns of human interviewers for designing future agents. The human interviewers often started their turns with "okay" in a separate phrase, and they used smiles or sometimes head wags to indicate their responses to what the interviewee said both as backchannel feedback and near the end of the interviewee's turn. We speculate that these gestures represented to the interviewee that interviewers were paying attention to what the interviewee said as well as non-verbally commenting on the interviewee statements.

Further data analysis on the turn-taking patterns is in progress. Thus, we would like to hold our conclusion until we complete our data analysis.

References

- [1] Bavelas, J. B., Coates, L., & Johnson, T. Listener Responses as a Collaborative Process: The Role of Gaze. *Journal of Communication*, 52(3), 566-580, 2002.
- [2] Duncan, S. Some Signals and Rules for Taking Speaking Turns in Conversations. *Journal of Personality and Social Psychology*, 23(2), 283-292, 1972.
- [3] Exline, R., Gray, D. & Schuette, D. Visual Behavior in a Dyad as Affected by Interview Content and Sex of Respondent. *Journal of Personality and Social Psychology*, 1(3), 201-209, 1965.
- [4] Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. Creating rapport with Virtual Agents. *Proc. 7th International Conference on Intelligent Virtual Agents*, 2007.
- [5] Kang, S. & Gratch, J. Virtual Humans Elicit Socially Anxious Interactants' Verbal Self-Disclosure. *Journal of Computer Animation and Virtual Worlds*, 21(3-4), 473-482, 2010.

Acknowledgements. This work was sponsored by the U.S. Army Research, Development, and Engineering Command and the National Science Foundation under grant # HS-0713603. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.