Designing the Impression of Social Agents' Real-time Interruption Handling

Patrick Gebhard
Tanja Schneeberger
firstname.lastname@dfki.de
German Research Center for Artificial Intelligence
Saarbrücken, Germany

Gregor Mehlmann, Tobias Baur
Elisabeth André
lastname@hcm-lab.de
Augsburg University
Augsburg, Germany

ABSTRACT

Human interaction partners can deal with interruptions and then resume the interaction. This ability should be emulated by social agents. How fast interruptions are handled might influence the overall impression of an agent. In this paper, we present the results of a user study on how a human dialog partner perceives the behavior of a virtual agent handling verbal user interruptions with different reaction times. The study goes beyond typical perception experiments by preserving the real-time interaction experience. For the evaluation, we rely on a parametrizable parallelized computational model that represents dialog flow, overlap detection, conflict recognition, and conflict handling in real-time. The evaluation results show that the timing of the agent's interruption handling in interactive human-agent dialogues is related to different interpersonal attitudes.

KEYWORDS

Model of conversational behavior, Verbal and nonverbal coordination, Evaluation methodologies and user studies

ACM Reference Format:
Patrick Gebhard, Tanja Schneeberger, Gregor Mehlmann, Tobias Baur, and Elisabeth André. 2019. Designing the Impression of Social Agents' Real-time Interruption Handling. In *ACM International Conference on Intelligent Virtual Agents (IVA '19), July 2--5, 2019, PARIS, France.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3308532.3329435

1 INTRODUCTION

Central to the interactive agent research is the understanding of human social communication behavior. Within that context, it is a primary aim to create social agents that show a believable social communication behavior [5]. A particular challenge is the detection of verbal interruptions by users and the creation of adequate agent reactions [16]. The timing of interruptions handling seems to be an important aspect. If the agent responds too late or not at all to a user barge-in, i.e., simply continues speaking, it might be perceived unfriendly, dominant or unresponsive.

For the first time with a real-time agent system, this work investigates the effect the agent's interruption handling time has on the user perception of the agent's behavior. In our study, we examine how a human dialog partner perceives our virtual agent's interpersonal attitudes dominance, friendliness, and closeness during an ongoing interaction.

IVA '19, July 2--5, 2019, PARIS, France

ACM ISBN 978-1-4503-6672-4/19/07.
https://doi.org/10.1145/3308532.3329435

2 THEORETICAL BACKGROUND

Dyadic conversations ideally follow a protocol that defines that the speaker sends nonverbal and/or verbal signals to the addressee when a turn change is approriate [14] while the listener waits for these signals to take the turn. An infringement of that protocol is an *interruption* of this ideal conversation structure of well-defined turn-changes. For this work, we focus on verbal interruptions by a user in an interaction with an agent. We investigate situations where a user interrupts an agent resulting in an overlap of speech.

Research in the area of virtual agents addresses the question of how interrupting agents are perceived by a human user engaged in a dialogue with the agent [15] or by an external human observer listening to a dialogue between two agents [3]. However, none of the agents are acting autonomously. In the first case, the agent is controlled by a human wizard. In the second case, pre-scripted videos are shown to the user. Recently, a decision-theoretic approach has been presented to automatically create dialogues between two conversing agents that portray a particular personality through their turn-taking and interruption behavior [10]. However, the approach has not yet been tested in an interactive setting with a human user.

2.1 Dominance, friendliness, and closeness

The conscious or unconscious evaluation of others' behavior is related to an assessment of dominance, friendliness, and closeness (among other attitudes) one person has regarding another [1, p. 1 ff]. One of the most widely used models of interpersonal attitude is the one of Argyle [2]. It is based on two dimensions, derived from the two main aspects of interpersonal behavior [8]: 1) *affiliation* ranges from unfriendly to friendly, and 2) *status* ranges from submissive to dominant. A dominant person has the disposition to control others [4]; a highly affiliated person is interested in high friendliness and closeness level between him and the interaction partner [11].

3 RELATED WORK

The investigation of agent user interruption handling has a long tradition. Many work focus on the need to handle low-latency timing requirements when responding to user barge-ins. The REA agent is one of the first virtual agent systems that can handle interruptions by users [5]. REA releases the turn to the user as soon s/he starts speaking or signals a turn taking by a gesture. That

is the user always succeeds in interrupting the agent no matter how long the speech overlap is. A more fine-tuned approach to handle verbal interruptions by users has been proposed for the virtual HWYD? ("How Was Your Day?") agent [7]. It comes with a intensity model to distinguish true interruptions from backchannel signals. Examples of newer systems are DiscoRT [13]. As the REA agent, DiscoRT agents stop immediately talking as soon as a barge-in event from the speech perceptor has been detected. The simon agent, an upper-torso humanoid robot, realizes collaborative behavior towards a user [6] by allowing users to interrupt robot actions in a shared physical workplace. For all systems it remains unclear how long it takes the agent to stop speaking after being interrupted and how this is perceived by users. Indeed, a small delay might have a negative effect on the agent's believability and other attitudes dominance, friendliness, and closeness.

Cafaro et al. [3] have investigated the effect of different voice overlaps in a dialogue between a speaker agent and an interrupting agent on a human observer in an offline study. The agents are displayed as opposing silhouettes in a picture giving a visual hint that two entities are talking to each other. We extend this work by moving beyond videos and involving the user as an interlocutor in a real-time dialogue.

4 TECHNICAL FRAMEWORK

Our interactive social agent Tom consists of software components for user input recognition, agent behavior modeling, virtual agent and TTS rendering, and automatic event annotation as described in [9]. For this work, we focus on the *StopSpeakingCommand*. It takes 400*ms* including all communication and processing delays. In addition to this time, the dry run time (0*ms* - 374*ms*) of the audio buffer has to be added (Fig. 1).

5 HYPOTHESES

The hypotheses address how Tom's interruption handling time affects its perceived dominance, friendliness, closeness. Cafaro *et al.* [3] found that observers rated the interruptee as more dominant when the vocal overlap time with the interrupter is increased. Furthermore, the assessment of the friendliness is decreased. Therefore, we expect that the agent is rated more dominant in the conditions with longer interruption handling times. Furthermore, users should evaluate the agent as more friendly and more close when the interruption handling time is shorter.

6 METHOD

6.1 Participants and Design

Data of 42 participants (25 female) aged between 18 and 44 years (M = 24.16 years, SD = 4.47 years) was examined. Participants were recruited via flyers and mailing lists at a German University on condition that they were fluent in German. All participants were rewarded with a cafeteria voucher for 20 minutes of participation.

To investigate how our agent is perceived by a user when reacting differently on natural interruptions, we realized a within-participants design with the agent's interruption handling time as a within factor. The three (Overlap Minimal *OM*, Overlap Short *OS*, Overlap Long *OL*) conditions (Fig. 1) were counterbalanced across the participants.

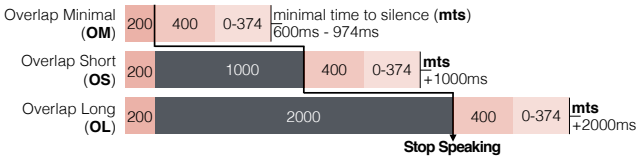

Figure 1: Conditions influence agent's speaking end.

6.2 Scenario and Experimental Setup

Participants were instructed to interrupt Tom three times in a relaxed conversation: 1) was assuming a wrong field of study of his dialog partner, 2) was talking about a topic that should be avoided, and 3) was fallen into a lengthy monolog.

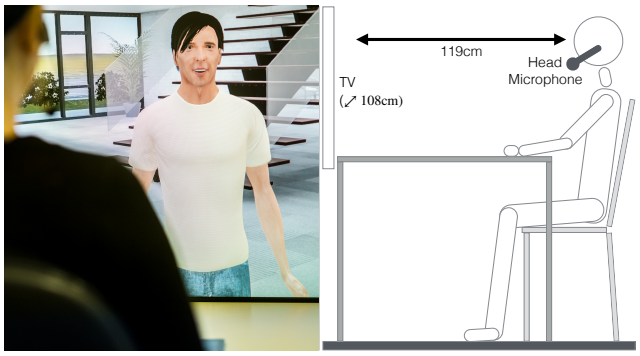

Figure 2: Study participant interacting with Tom.

Technically, the experimental set-up consisted of a PC running MS Windows 10^{TM} (Intel Core i7 CPU@3.5GHZ, 16GB Memory, NVIDIA GTX 990 graphics cards) connected to an LCD TV screen (108cm diagonal) showing the agent at a realistic size in a 3d environment. Each participant was seated at a table in front of the display at a distance of 119 cm (Fig. 2). Each participant was asked to wear a head mounted microphone (Sure SM35) in order to cancel any environmental sounds. The microphone was connected via a USB (version 2) interface to the PC (Tascam US-2x2).

6.3 Measurements

Agent's Dominance. This scale consisted of the three items "*Tom tries to control the interaction*", "*Tom dominates the interaction*", and "*Tom talks at rather than with me*".
Agent's Friendliness. Like in [3], friendliness was judged with two the items "*Tom seems likable*" and "*Tom expresses hostility*".
Agent's Closeness. Four items were used to build the scale closeness. Those were "*Tom seems to like me*", "*Tom expresses sympathy toward me*", "*Tom Expresses warmth*", and "*Tom keeps me at a distance, avoids development of any sort of interpersonal relationship*".

6.4 Procedure

After the experimenter explained the procedure, participants filled in demographics and the informed consent. Hereafter, the they were equipped with a head-mounted microphone and entered the room where the conversation with the virtual agent took place alone. After they had plugged in the microphone, the tutorial to learn how to interrupt Tom started. Then, the participants had

the three conversations in the three conditions with Tom, each followed by the questionnaire to assess Tom and the interaction with him. At the end, participants left the lab and were debriefed by the experimenter. The whole procedure took around 20 minutes.

7 RESULTS

Agent's Dominance. We found a significant main effect ($F(2,82) = 5.75$, $p < .005$). The longer the interruption handling time the more dominant the agent was perceived ($M = 3.04$; $SD = 0.65$ in the condition overlap minimal, $M = 3.27$; $SD = 0.62$ in the condition overlap short, $M = 3.39$; $SD = 0.50$ in the condition overlap long).
Agent's Friendliness. For friendliness, the univariate analyses revealed a marginal effect ($F(2,82) = 2.21$, $p = .06$, $M = 4.10$; $SD = 0.50$ in the condition overlap minimal, $M = 3.98$; $SD = 0.64$ in the condition overlap short, $M = 3.85$; $SD = 0.76$ in the condition overlap long). The longer the interruption handling time, the less friendly the agent was perceived. Helmert contrasts showed that there is a significant difference between the minimal overlap condition and the two other conditions overlap short and overlap long ($F(1,41) = 3.61$, $p < .05$). The agent was perceived more friendly in the minimal overlap condition compared to the two other longer conditions.
Agent's Closeness. We discovered a main effect for the perceived closeness of the agent ($F(2,82) = 2.82$, $p < .05$). With an increasing interruption handling time the closeness of the agent is perceived significantly lower ($M = 3.63$; $SD = 0.63$ in the condition overlap minimal, $M = 3.51$; $SD = 0.69$ in the condition overlap short, $M = 3.39$; $SD = 0.67$ in the condition overlap long).

8 DISCUSSION

Regarding the perceived agent's dominance, we could confirm our hypotheses. As expected, the agent was assessed more dominant the longer the interruption handling time was. Those results are in line with the finding that as the amount of overlap increased, the perceived interruptee's dominance increased as well [3]. For the perceived friendliness, we found that compared to the minimal overlap, the agent was evaluated less friendly in the two conditions with more overlap. This result confirms again what the work of Cafaro *et al.* revealed. With increasing overlap time, the perceived closeness of the agent decreased which confirms our hypotheses regarding the interpersonal attitude closeness. Participants assessed the agent's closeness as higher the shorter the overlap time was. In sum, the agent that stopped speaking immediately was evaluated less dominant, more friendly and more close.

9 CONCLUSION

In this paper, we investigated for the first time with a real-time interactive agent system how different interruption handling times affect the perception of our Social Agent. The study revealed that users assess the agent as less dominant, more friendly and closer when the agent's interruption handling time is short. Based on those results, we would recommend designing interactive agents with a short interruption handling time if a positive attitude regarding the agent is aimed.

The exact adjustment of the interruption handling time was realized with a multi-layer interruption handling model. It enables the agent with a parameterizable timing of interruption handling and allows the agent to react according to the meaning of the user's verbal interruption. Such a model that realizes the reciprocal meshing of concurrent processes for multimodal behavior recognition, knowledge reasoning, and multimodal behavior generation is central to highly responsive interactive agents [12].

ACKNOWLEDGMENTS

This work is partially funded by the German Ministry of Education and Research (BMBF) within the EmpaT project (16SV7229K). We thank Christian Dold (Charamel GmbH) for realizing our requirements with regard to stop TTS output in a virtual agent environment and for his detailed technical explanations.

REFERENCES

[1] Icek Ajzen. 2005. *Attitudes, personality, and behavior* (2nd ed.). Open University Press, New York, NY, USA.
[2] Michael Argyle. 2007. *Bodily Communication* (2nd ed.). Routledge, East Sussex.
[3] Angelo Cafaro, Nadine Glas, and Catherine Pelachaud. 2016. The Effects of Interrupting Behavior on Interpersonal Attitude and Engagement in Dyadic Interactions. In *Proceedings of the International Conference on Autonomous Agents & Multiagent Systems (AAMAS '16)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 911--920.
[4] Elizabeth Cashdan. 1998. Smiles, speech, and body posture: How women and men display sociometric status and power. *Journal of Nonverbal Behavior* 22, 4 (1998), 209--228.
[5] Justine Cassell, Tim Bickmore, Lee Campbell, Hannes Vilhjalmsson, and Hao Yan. 2001. More than just a pretty face: conversational protocols and the affordances of embodiment. *Knowledge-based systems* 14, 1 (2001), 55--64.
[6] Crystal Chao. 2012. Timing Multimodal Turn-Taking for Human-robot Cooperation. In *Proceedings of the 14th International Conference on Multimodal Interaction (ICMI '12)*. ACM, New York, NY, USA, 309--312.
[7] Nigel Crook, Debora Field, Cameron Smith, Sue Harding, Stephen Pulman, Marc Cavazza, Daniel Charlton, Roger Moore, and Johan Boye. 2012. Generating context-sensitive ECA responses to user barge-in interruptions. *Journal on Multimodal User Interfaces* 6, 1-2 (2012), 13--25.
[8] Uriel G. Foa. 1961. Convergences in the analysis of the structure of interpersonal behavior. *Psychological Review* 68, 5 (1961), 341.
[9] Patrick Gebhard, Tanja Schneeberger, Elisabeth André, Tobias Baur, Ionut Damian, Gregor Mehlmann, Cornelius König, and Markus Langer. 2018. Serious Games for Training Social Skills in Job Interviews. *IEEE Transactions on Games* (2018).
[10] Kathrin Janowski and Elisabeth André. 2019. What If I Speak Now? A Decision-Theoretic Approach to Personality-Based Turn-Taking. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2019, Montréal, May 13-17, 2019*.
[11] Zerrin Kasap, Maher Ben Moussa, Parag Chaudhuri, and Nadia Magnenat-Thalmann. 2009. Making Them Remember? Emotional Virtual Characters With Memory. *IEEE Computer Graphics and Applications* 29, 2 (2009), 20--29.
[12] Gregor Mehlmann, Kathrin Janowski, and Elisabeth André. 2016. Modeling Grounding for Interactive Social Companions. *Special Issue of the German Journal of Artificial Intelligence: Social Companion Technologies* 30, 1 (2016), 45--52.
[13] Bahador Nooraei, Charles Rich, and Candace L. Sidner. 2014. A Real-Time Architecture for Embodied Conversational Agents: Beyond Turn-Taking. In *Proceedings of the 7th International Conference on Advances in Computer-Human Interactions (ACHI '14)*. IARIA, Wilmington, DE, USA, 381--388.
[14] Harvey Sacks, Emanuel Schlegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking. *Journal of Language* 50, 4 (1974), 696--735.
[15] Mark Ter Maat, Khiet P. Truong, and Dirk Heylen. 2010. How Turn-Taking Strategies Influence Users' Impressions of an Agent. In *Proceedings of the 11th International Conference on Intelligent Virtual Agents (Lecture Notes in Computer Science)*, Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova (Eds.), Vol. 6356. Springer, Berlin, Heidelberg, Cham, 441--453.
[16] Nigel G. Ward and David DeVault. 2016. Ten Challenges in Highly-Interactive Dialog Systems. *AI Magazine* 37, 4 (2016), 7--18.