"Look at Me!" – Self-Interruptions as Attention Booster?

Birte Carlmeyer
CITEC - Bielefeld University
Bielefeld, Germany
bcarlmey@techfak.uni-bielefeld.de

David Schlangen
CITEC - Bielefeld University
Bielefeld, Germany
david.schlangen@uni-bielefeld.de

Britta Wrede
CITEC - Bielefeld University
Bielefeld, Germany
bwrede@techfak.uni-bielefeld.de

ABSTRACT

In this paper we present results of an exploratory experiment investigating the effects of a contingently self-interrupting vs non-self-interrupting virtual agent who transmits information to a human interaction partner. In the experimental condition self-interruptions of the agent were triggered by an external event whereas in the control group the agent did not react to this event. We measured the effect of the agent's self-interruptions on human attention, memory performance and subjective ratings. In this paper we discuss the results with respect to the design of incremental human-agent dialogue modeling.

ACM Classification Keywords

H.5.2 Information interfaces and presentation: User Interfaces; I.2.7 Artificial intelligence: Natural Language Processing

Author Keywords

Dialogue management; incr. processing; multi-modal systems

INTRODUCTION

Smart home environments provide a range of powerful automation capabilities. However, so far no convincing concept of a smart and easy to follow interface has been proposed. Consequently, many functionalities remain unused. In our project we envision interaction with the environment via a virtual or robotic agent who provides help and information on request. However, information about such complex functionalities tend to be large and may lead to lengthy monologues. Alternatively, they may be chunked into smaller pieces with explicit requests for continuation. Both strategies yield cumbersome interactions leaving the user with the wish to interrupt or simply leave. We therefore propose an incremental dialogue model that enables interruptions of the system at any time. Additionally, the system should also be able to interrupt the ongoing interaction when the user looses interest or disengages due to distractions in the home environment. It is thus important to monitor a user's attention in order to avoid disruptions of the interaction or inattentive system behavior. Models of keeping track of the user's level of engagement have been proposed as an important feature of human-agent interaction [3, 6]. At the social level, joint attention indicates engagement in an interaction [10]. Consequently, looking away (if not caused by a reference within the interaction) can be interpreted as leaving the interaction [3]. While monitoring the user's attention level is an important step for modeling human-agent interaction, it still neglects the question how to reacquire a user's attention when it has moved away?

STATE OF THE ART

Although there is an increasing amount on studies in the area of attention and the use of eye-gaze, still relatively few studies explicitly address on how to best apply these findings to human-agent interaction (HAI) and to implement them in incremental dialogue systems. In [10] the authors report that the initial 5 seconds of an interaction correlate with the user's following engagement level: if the robot provides a contingent looking strategy, including looking-away if the user did not look (at the beginning of the interaction) users in the museum setting were more likely to remain longer in interaction with the robot as opposed to a non-contingent strategy. Thus, interruptions in the agent's gazing behavior seem to have an effect on the user's attention. [12] analyzed different gaze patterns in multi-party conversations and found that gaze can be used as a predictor of attention in conversations.

An in-car scenario [7] showed that self-interruptions of an information giving system leads to increased memory performance. Self-interruptions were initiated in situations where the user was involved in another, potentially dangerous task such as switching lanes or overtaking somebody. In a robot teaching context [9] evaluate the role of gaze as implicit signal for turn-taking in a dictating scenario and showed that gaze as synchronization cue has an impact on task performance in a two-party setting. Thus, self-interruptions in a task-oriented interaction can lead to increased task performance. While these works focus on task-oriented interaction they do not evaluate how this strategy affects subjective ratings.

Another strategy for maintaining a user's attention that has been proposed consists of introducing hesitations (or filled pauses). [4] report on increased engagement levels when providing hesitations in a human-robot interaction scenario. While focusing on a task-oriented interaction (i.e. providing directions) it was not evaluated how this strategy affected cognitive performance. Also, its effect on subjective ratings was not assessed.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

Considering the results of the presented literature, one important feature for re-acquiring a user's attention seems to be by contingent self-interruptions. We therefore propose a strategy that provides self-interruptions of the agent in situations where the user is distracted. We further define the following hypotheses: (1) Self-interruptions of an agent will increase cognitive performance (better post-interaction information recall) of the human interaction partner (2) Self-interruptions will reacquire attention as measured by gazing behavior and (3) influence subjective ratings of the agent.

ATTENTION MODEL

In our model we define attention while the system or agent is speaking as a state where the human interlocutor's visual focus of attention (VFoA) is consistent with the focus of discourse (FoD) as determined by the system's interpretation of the ongoing dialogue. The user's VFoA can be recognized through visual perception of his/her head pose [11], whereas the FoD is provided by the dialogue management (DM) and defined as the physical reference of the topic that is currently being talked about (e.g. a referenced object in the environment or direction) or - in absence of this - the interaction partner. Figure 1 shows a schematic graphical representation of our model. If

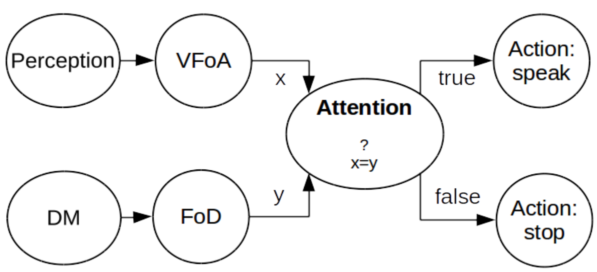

Figure 1. Attention Model: define attention as state where the visual focus of attention (VFoA) is consistent with the focus of discourse (FoD) and triggers different actions. ***Image ©Birte Carlmeyer***

the human interaction partner is attentive, meaning his/her VFoA is consistent with the the FoD of the current interaction state, the agent will start or continue with the interaction, i.e. speaking. Otherwise the agent has to reacquire attention through a dedicated reacquisition action which in our model is defined as an immediate break-off of the speech synthesis. To simplify the evaluation of the effect of a self-interrupting agent, we chose a topic where the FoD is on the agent itself (in this case information about the agent).

METHOD

We evaluated the effect of a verbally self-interrupting agent in a human-agent interaction in a smart home environment. The agent was providing information about itself through a sequence of 6 sentences. In both conditions we provided an external distraction in the apartment to the right side of the participant at an angle of about 90 degrees in order to withdraw the user's VFoA from the system. In the experimental condition this triggered the self-interruption behavior of the system (triggered by a Wizard through pressing a button upon perceiving the user's VFoA shifting away). The agent would directly stop speaking and continue exactly at the break-off point when the user's VFoA moved back to the agent. In the control condition the agent simply continued speaking throughout the distraction. The distraction was achieved by the experimenter reentering the room pretending to bring in some missing documents for the experiment, issuing a brief verbal apology with explanation and leaving.

Experimental Setup

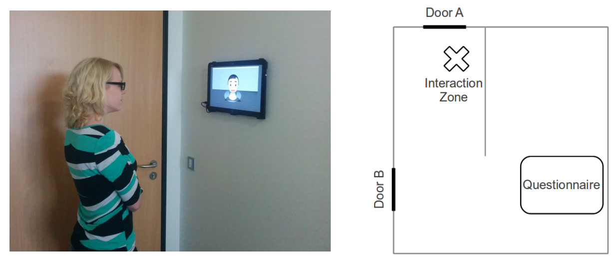

Figure 2. Experimental setup. Left: person interacting with the agent. Right: ground view of the apartment. ***Image ©Birte Carlmeyer***

Figure 2 shows the experimental setup. The participants were facing a tablet, which was showing a simulation of the robot platform Flobi[8], an anthropomorphic robot head. Through the tablet camera, the simulated Flobi is able to detect faces in front of it and focus on them, thus establishing shared attention. The human-agent interaction had three phases of verbal interaction (monologue by the agent) in both conditions: *Phase 1*: Greeting. *Phase 2*: Information about the system (6 sentences). The distraction was initiated after the first sentence. *Phase 3*: Request to move on to fill out the questionnaire at the computer in the room to the right of the participant. Flobi's verbalizations were predefined and triggered from an adjoining room by a wizard who observed the participant through the tablet's camera. To allow verbal self-interruptions, we used the incremental speech synthesis module of InproTK[2] and its integration in the PaMini dialogue manager [5] which supports immediate interruption and resuming of the speech synthesis.

The questionnaire consisted of two parts: a memory task and subjective ratings about Flobi. The memory task consisted of six statements for which the participants had to decide whether or not this was a statement that had been made by the agent during Phase 2. In the second part the participants had to provide subjective ratings of the agent through a set of adjectives on a Likert scale ranging from 1 to 7 to evaluate five key concepts in human-robot interaction: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety (based on [1]).

The experimental procedure was as follows. After signing a consent form, the subjects were led to the experiment room. They entered the room alone through *Door A*, only with the instruction to look at the tablet on the left wall and to fill out a questionnaire on the computer after the interaction. The wizard started the interaction as soon as the participants stood in front of the tablet facing it. The study assistants disturbed the interaction always after the first sentence of Part 2 of the interaction was finished, by entering the the experiment room through *Door B*. At the end of the interaction the participants went to the table and filled out the questionnaire on a computer.

Conditions and Dependent Variables
We compared two conditions. In the first condition the agent reacted *with self-interruptions* (cf. Experimental Set-up). In the *control* group the agent did not react to the distractions and kept on speaking. In order to assess the memory performance of the participants we counted the number of correct answers to the content-related questions of the questionnaire. To obtain a measure for the attention we manually annotated the head position of the user and measured the number and duration participants looked away from the agent during the interaction (i.e. the number and duration where VFoA was different to FoD). For the subjective ratings we evaluated the answers of the second part of the questionnaire.

RESULTS
In total 27 subjects (9 female, 18 male, aged 21-51) took part in the study. The average age was 27.2 with a standard deviation of 5.3. 13 participants were in the condition *with self-interruption* and 14 in the *control* group. The study assistants disturbed the human-agent interaction in the experimental condition *with self-interruptions* 10.47 second in average and in the *control* condition 10.25 seconds. For the statistical analysis we chose an alpha level of 0.05.

Memory Performance
At first we want to explore the memory performance. Note that all questions were yes/no questions. The percentages of correct answers for each condition for the different memory questions are shown in Figure 3. No significant effects in the

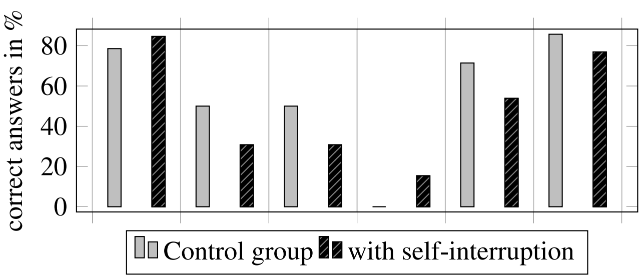

Figure 3. Performance in memory task for the different questions.

results between the two conditions were found. The overall percentages of correct answers for the experimental condition *with self-interruption* is 48.7% whereas the subjects in the *control* condition answered 56.0% correct. This difference is not large enough for statistical significance. These results indicate participants simply guessed in both conditions. We were thus not able to confirm our first hypothesis (1).

Visual Attention
In order to assess the participants' visual attention we measured the number of shifts of VFoA away from the agent during Part 2 of the interaction. In the experimental condition *with self-interruptions* most participants looked away only once (9/13). One participant did not get distracted at all. Only three subjects looked away more than once. In contrast in the *control* group more than half of the participants (8/14) looked away more than once, even while there were no more distractions. Five participants looked away two times and 3 subjects even three times. However, these differences did not reach a significant level. Also for the time of the first "look away" after the student assistant's distraction, no significant effects of the mean time were found. Figure 4 shows the overall time of participants looking away during phase two of the interaction. We tested significance of the results using a generalized linear mixed model and found a significant effect between the two conditions (F=4.386, $p = 0.047$). The participants in

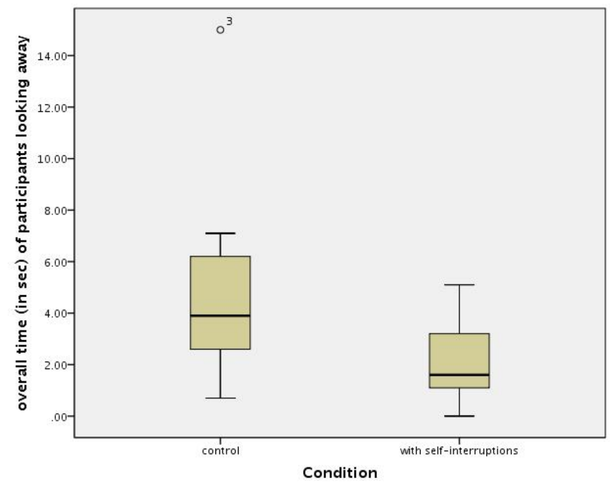

Figure 4. Overall time (in sec) of participants looking away during the interaction.

the *control* group looked away longer than the participants in the experimental condition *with self-interruption* during phase two of the interaction, thus confirming our hypothesis (2).

Subjective Ratings
Next we want to explore the subjective ratings of the agent. The MANOVA over all subjective ratings revealed a marginal multivariate effect (F=9.718, p=0.97) between the two conditions. More specifically, participants in the experimental group *with self-interruption* rated the agent significantly less likable than the *control* group (F=6.588, p=0.017). While the experimental group rated the likability of the agent with 4.8 the *control* group's rating was 5.6. For other ratings no significant effects were found.

DISCUSSION AND OUTLOOK
The results of the looking behavior measurement suggest that the self-interrupting of the agent has a significant effect on the visual focus of attention of the human interaction partner. This effect manifests in the overall time participants looked away from the agent and indicates that self-interruptions are an effective intervention strategy to regain the attention of the interaction partner. Interestingly, in one case, the self-interrupting behavior had precisely the contrary effect as it lead the participant to repeatedly look away in order to test the agent's capabilities. While this is clearly a novelty effect that is likely to disappear in further interactions it indicates that the highly contingent self-interrupting behavior has a very

powerful effect on the interaction partner's perception of the agent and deserves further investigations as detailed below.

While in the presented experiment the FoD was on the agent itself we will move on to scenarios where the FoD will change, as is typically the case in joint task situations. Consider, for example, a scenario where a robot or virtual agent gives assistance in a smart home. In such a task-oriented interaction the FoD (and thus the VFoA) will shift away from the agent itself towards appliances that are being explained or discussed by the agent. In such more task-oriented interactions that require even more cognitive involvement of the user - as the explanations may become complicated - the resuming of the interaction will become more important. As shown by [7] self-interruptions can help to increase cognitive performance. However, it remains unclear if this effect is due to the self-interruptions alone or also caused by the repetition of parts of the utterances. In our next study we will target this question. The fact that our results did not show a significant effect of self-interruptions on memory performance may be due to two different factors: on the one hand, the users were not provided with repetitions of the utterance from just before the interruption, on the other hand the questions might simply have been too difficult. We will explore in further studies how the positive effect on the memory performance can be replicated in our setting.

Although the self-interruptions had a positive guiding effect on the VFoA of the participants, they rated the self-interrupting agent significantly less likable. To prevent or at least ameliorate this effect we plan to integrate a more adaptable speech synthesis. For example the agent could not only repeat the last few utterances but also produce hesitations as proposed by [4].

CONCLUSIONS

We have presented an human-agent interaction experiment investigating the effect of a verbal self-interrupting agent on human attention, memory performance and subjective ratings showing that self-interruptions are effective in re-acquiring VFoA which is in line with [4]. We furthermore showed that this positive effect is achieved at the cost of less positive subjective ratings and proposed to adapt the speech synthesis to ameliorate or compensate this effect. Additionally, we discussed potential positive implications of this behavior on the user's memory performance. For the further optimization process of our model we will take all three dimensions (attention, memory performance, subjective ratings) into account.

ACKNOWLEDGMENTS

This work was funded as part of the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277), Bielefeld University, and by the German Federal Ministry of Education and Research (BMBF) via the KogniHome project (project number: 16SV7054K).

REFERENCES

1. Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1, 1 (2009), 71–81.
2. Timo Baumann and David Schlangen. 2012. The InproTK 2012 release. In *Proc. of the NAACL-HLT Workshop on Future directions and needs in the Spoken Dialog Community: Tools and Data*. ACL, 29–32.
3. Dan Bohus and Eric Horvitz. 2009. Models for Multiparty Engagement in Open-world Dialog. In *Proc. of the SIGDIAL 2009 Conference*. ACL, Stroudsburg, PA, USA, 225–234.
4. Dan Bohus and Eric Horvitz. 2014. Managing Human-Robot Engagement with Forecasts and... Um... Hesitations. In *Proc. of the 16th International Conference on Multimodal Interaction*. ACM, New York, USA, 2–9.
5. Birte Carlmeyer, David Schlangen, and Britta Wrede. 2014. Towards Closed Feedback Loops in HRI: Integrating InproTK and PaMini. In *Proc. of the 2014 Workshop on Multimodal, Multi-Party, Real-World Human-Robot Interaction*. ACM, 1–6.
6. David Klotz, Johannes Wienke, Julia Peltason, Britta Wrede, Sebastian Wrede, Vasil Khalidov, and Jean-Marc Odobez. 2011. Engagement-based Multi-party Dialog with a Humanoid Robot. In *Proc. of the SIGDIAL 2011 Conference*. ACL, 341–343.
7. Spyridon Kousidis, Casey Kennington, Timo Baumann, Hendrik Buschmeier, Stefan Kopp, and David Schlangen. 2014. Situationally Aware In-Car Information Presentation Using Incremental Speech Generation: Safer, and More Effective. In *Proc. of the EACL 2014 Workshop on Dialogue in Motion*. 68–72.
8. Ingo Lütkebohle, Frank Hegel, Simon Schulz, Matthias Hackel, Britta Wrede, Sven Wachsmuth, and Gerhard Sagerer. 2010. The Bielefeld Anthropomorphic Robot Head ''Flobi''. In *2010 IEEE International Conference on Robotics and Automation*. IEEE, 3384–3391.
9. Oskar Palinko, Alessandra Sciutti, Lars Schillingmann, Francesco Rea, Yukie Nagai, and Giulio Sandini. 2015. Gaze contingency in turn-taking for human robot interaction: Advantages and drawbacks. In *IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 369–374.
10. Karola Pitsch, Hideaki Kuzuoka, Yuya Suzuki, Luise Süssenbach, Paul Luff, and Christian Heath. 2009. The first five seconds: Contingent stepwise entry into an interaction as a means to secure sustained engagement. In *IEEE International Symposium on Robot and Human Interactive Communication*. 985–991.
11. K. Smith, S. O. Ba, J. M. Odobez, and D. Gatica-Perez. 2008. Tracking the Visual Focus of Attention for a Varying Number of Wandering People. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 7 (July 2008), 1212–1229.
12. Roel Vertegaal, Robert Slagter, Gerrit van der Veer, and Anton Nijholt. 2001. Eye Gaze Patterns in Conversations: There is More to Conversational Agents Than Meets the Eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 301–308.