

Exploring Self-interruptions as a Strategy for Regaining the Attention of Distracted Users

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 a first exploratory study investigating the effects of a contingently self-interrupting vs non-self-interrupting virtual agent in a smart home environment who transmits information to a human interaction partner. We tested the hypothesis that self-interruptions are a strategy for keeping the user's attention, as measured by post-interaction information recall. Interestingly, our experiment does not allow us to confirm this hypothesis. In fact, users found the self-interruption strategy to be less-likeable. From our observations, we draw suggestions for future implementations of attention-retainment strategies.

CCS Concepts

•Human-centered computing → Human computer interaction (HCI); •Computing methodologies → *Intelligent agents*;

Keywords

incremental processing; multi-modal systems; attention; gaze; attention; dialogue management; human-agent interaction; smart home

1. INTRODUCTION

A strategy that is often recommended to teachers for regaining the attention of distracted students is to just interrupt themselves mid-sentence and only continue speaking when attention has returned (e.g., [20]). Does this also work for embodied agents? We investigated this question in the context of a smart home setting, where control is partially embodied in a virtual agent. Smart homes are among the most emerging interaction scenarios of current and future research. The most obvious challenges of such interactions are (1) the complexity of the technology that the user wants to control, rendering the interaction topic complex, and (2) the embeddedness of the interaction in other activities that

may be more important to the user and lead to distractions [12]. However, interaction with(in) a smart home provides a range of advantages over other interactions: (1) it is massively embodied, i.e., the environment has a range of actuators at its disposal to communicate with the user, and (2) the interaction is much more contextualized, providing important context information, e.g., through knowledge of the time of day and location of the interaction which are powerful cues to predict the user's intention.

One fundamental research question is, therefore, to better understand how an embodied interaction entity, such as a smart apartment, can - through the interface of a humanoid robot avatar - support the guidance of the user's attention during interactions while the user is subject to distractions from the environment. More concretely, we investigate a scenario where a smart home supported by a range of multi-modal, embodied informational cues, about its features is providing verbal explanations via an embodied virtual robot while the user gets distracted through various different events.

In a similar scenario, a car driving situation, [10] have suggested a dedicated attention strategy, that allows the user's attention to focus on a distracting difficult driving maneuver. This is achieved by a self-interrupting dialogue which continues speech production when the distraction is over. It turned out that this approach yielded significantly increased memory performance as opposed to a system that ignored difficult driving situations and continued speaking.

In an exploratory study we wanted to investigate in how far such an attention strategy is transferable to a smart home setting and whether the specific characteristics of the interaction, i.e., the high degree of embodiment, have an effect on the user's memory performance as well as her/his subjective evaluation of the system. In our work we define attention as a state where the user shares the visual focus of attention of the agent or focuses on the agent itself. More specifically, a user is attentive when her/his visual focus of attention lies on the same target as the current focus of dialogue as provided by the dialogue manager or on the interaction partner.

2. STATE OF THE ART

Although there is an increasing amount of 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 task-oriented human-robot interaction (HRI) the eye-gaze of the robot has been shown to convey important in-

formation about its level of understanding and is beneficial in HRI [19]. For example, anticipatory eye-gaze to a target position indicates that the robot has understood what the goal of an action is [22]. At the social level, joint attention indicates engagement in an interaction [15]. Consequently, looking away (if not caused by a reference within the interaction) can be interpreted as leaving the interaction [4]. Using gaze for tracking the user’s level of engagement have thus been proposed as an important feature of HRI [4, 9, 16]. In human-human interaction the important role of eye-gaze is quiet well evaluated, e.g., by [1, 7] – to only mention a few – but also the importance of eye-gaze or head movements as turn-taking or addressee signal in human-agent interaction has been shown in quite a number of interaction scenarios.

Vertegaal *et al.* analyzed different gaze patterns in multi-party conversations and found that “user’s eye gaze can form a reliable source of input for conversational systems that need to establish whether the user is speaking or listening to them” [21]. They also suggest that the “predictive function of gaze might be transformed into a more generic indicator of joint interest: subjects were keeping in touch by looking at the same object” [21]. The rickel gaze model for human-agent interaction consists of different types of gaze – motivated from human-human interaction – to reflect the inner state of the agent’s cognitive processing [11].

Skantze and Gustafson evaluated mutual gaze as an interaction control mechanism in a multi-party human-human-computer dialogue setting [18]. They presented an attention and interaction model that allows the user to switch attention between the system and other humans. Skantze *et al.* monitor users’ attention by tracking their head movements, which directly influences the speech production of the system, i.e., the system stops speaking if it loses the attention of the user. They compared this approach with a push-to-talk scenario but there where no clear consensus which system was the best. Also Yu *et al.* used head movements of the user as attention signal to coordinate the robot’s speech production, but they did not evaluated the effect on the users’ attention or subjective ratings [24].

In a robot teaching context, Palinko *et al.* evaluated 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 [14]. An in-car scenario, Kousidis *et al.* showed that self-interruptions and repetitions of an information giving system leads to increased memory performance [10]. Self-interruptions were initiated in situations where the user was involved in another, potentially dangerous task such as switching lanes or overtaking somebody. While these works focus on self-interrupting agents and attention in task-oriented HAI interaction (e.g., dictating or providing directions) they do not evaluate how this strategy affects the subjective ratings of the agent.

Another strategy for maintaining a user’s attention that has been proposed by Bohus and Horvitz consists of introducing hesitations (or filled pauses). They report on increased engagement levels when using hesitations in a human-robot interaction scenario [5]. 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.

3. ATTENTION MODEL IN OUR SYSTEM

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 belief of the on-going dialogue. The user’s VFoA can be recognized through visual perception of his/her eye-gaze, whereas the FoD is provided by the dialogue management 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 the interaction partner. Figure 1 shows a schematic

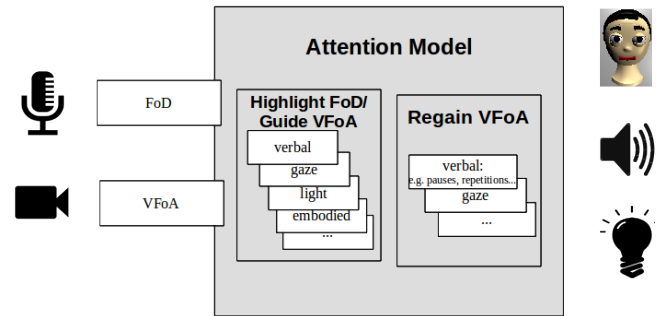


Figure 1: Our attention model defines attention as state where the visual focus of attention (VFoA) is consistent with the focus of discourse (FoD) and triggers different actions. Image ©Birte Carlmeyer

graphical representation of our attention model as part of the overall dialogue management. It consists of two different parts. The first part contains various modalities to highlight the current focus of discourse. If the agent wants to shift the current focus of discourse to another object, this can be done by verbalizing it (e.g., “to your left side”) or looking or pointing at the new focus of discourse. In the context of our smart environment, the agent is also able to use other modalities such as ambient or spot light. The second part of our attention module consists of different repair strategies to recover the visual attention of the interaction partner. If the human interaction partner is attentive, meaning his/her VFoA is consistent with the FoD of the current interaction state, the agent will start or continue with the interaction, i.e. speaking. Otherwise the agent needs to reacquire attention through a dedicated re-acquisition action. One strategy in our model is an immediate break-off of the speech synthesis. If the human interlocutor’s visual focus of attention shifts away from the FoD (i.e., a referenced object or the interaction partner itself) the agent stops speaking. As soon as the interlocutor is attentive again the agent continues speaking. Another strategy belongs to the attention guiding. In case that the human does not react to a change of the discourse – meaning the human does not look (even once) at the new focus of discourse – the agent stops and repeats the last highlighting strategy.

4. STUDY SETUP & METHOD

In this section we describe the study setup and applied evaluation methods. Therefore we first expand on the experimental setup, the different conditions, and then on the data recording and annotation.

4.1 Experimental setup

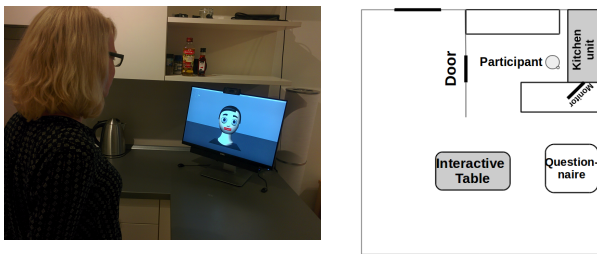


Figure 2: Experimental setup. Left: person interacting with the agent. Right: ground view of the experimental apartment. Image ©Birte Carlmeyer

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 and the smart environment through a sequence of 23 sentences with four different foci of discourse (the agent itself, the kitchen unit, the living room with an interactive table and the ceiling). Accordingly, we defined different areas of visual focus of attention (agent, kitchen, living, ceiling, other).

We provided three audio-visual external distractions in the apartment in order to distract the user’s attention from the system. The first two distractions were blinking lights and a sound in the apartment, while the third was achieved by the experimenter re-entering the room.

Figure 2 shows the experimental setup. The participants were facing a monitor, which was showing a simulation of the robot platform Flobi[13], an anthropomorphic robot head. Using a camera on top of the monitor, the simulated Flobi was able to detect faces in front of it and focus on them to establish shared attention. The human-agent interaction had three phases of verbal action (monologue) by the agent in both conditions: *Phase 1*: Greeting. *Phase 2*: Information about the agent itself, the kitchen, the living room and the ceiling of the intelligent apartment. *Phase 3*: Request to move on to fill out questionnaire at the computer in the living room to the right of the participant.

Flobi’s verbalizations were predefined and the start of the interaction was triggered from an adjoining room by a wizard who observed the participant through the camera. To allow verbal self-interruptions, we used the incremental speech synthesis module of InproTK[3] and its integration in the PaMini dialogue manager [6] which supports immediate interruption and continuation of speech synthesis. To assess the gaze direction of currently observable persons the gaze detector from Schillingmann *et al.* [17] was used.

The questionnaire consisted of two parts: a memory task and subjective ratings about Flobi. The memory task consisted of ten multiple choice questions about the information previously stated by Flobi during Phase 2. Four of the ten questions addressed information which was not only verbally described by Flobi, but also additionally presented embodied via a corresponding actor in the apartment. In the second part the participants had to provide subjective ratings of the agent through a set of 24 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 [2]).

The experimental procedure was as follows: After signing a consent form, the subjects were led to the experiment room. They entered the room alone, only with the instruction to go into the kitchen, look at the agent, listen carefully and afterwards fill out a questionnaire on the computer. The interaction started as soon as the participant stood in front of Flobi. The disturbances were always triggered at predefined points of the interaction. At the end of the interaction the participants went to the table and filled out the questionnaire on a computer.

4.2 Conditions and Dependent Variables

During all trials the agent highlighted the current focus of discourse verbally, through gaze movements and in case of embodied topics through the corresponding actuator (e.g., ambient light or facial expressions). In the experimental condition the agent reacted *with self-interruptions* when the participant’s VFoA did not match the FoD according to the attention model. The agent would continue speaking exactly at the break-off point of the sentence when the user’s VFoA moved back to the agent. Furthermore, the attention guiding strategy was repeated when necessary. In the *control* group the agent ignored the changes in the participant’s VFoA, and neither stopped speaking nor repeated itself. 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. For the subjective rating we evaluated the answers of the second part of the questionnaire.

4.3 Data Recording and Annotation

All interactions were recorded via two network-enabled Basler cameras and one Rode NT55 omni-directional microphone mounted at the ceiling of the apartment to cover the whole interaction area. Moreover, we collected various system events such as generated dialogue acts and detailed information of the gaze recognition results. For annotation purposes, the two top-down videos, the audio stream and system events were merged into one ELAN[23] file (for further information about this process refer to [8]).

The study has been carried out with German native speakers. In total, we recorded 31 trials with 14 female and 17 male participants in total. The average age was 25.1 with a standard deviation of 4.3. 16 participants were in the condition *with self-interruption* (8f, 8m) and 15 in the *control* group (6f, 9m). Most of them had very little or no experience with robotic systems.

5. RESULTS

In this section we first explore the memory performance of the participants and then the subjective ratings of the agent. For the statistical analysis we chose an alpha level of 0.05.

5.1 Memory Performance

We first explore the memory performance of the participants. No significant effects in the results of a χ^2 test between the two conditions could be found. The overall error rate for the experimental condition *with self-interruption* was 18.1% whereas the subjects in the *control* condition answered in 19.3% of the cases incorrectly. Also, no significant differences between the correct answers to the questions regarding the three utterances during which the disturbances

were given, could be found. Thus, we could not confirm the positive memory effect of self-interruption. In general, the relative low degree of errors indicate that the difficulty of the questions was appropriate.

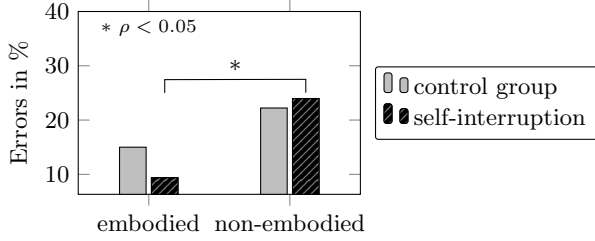


Figure 3: Cognitive performance in memory task for the different kind of questions.

In order to address the more exploratory question, in how far the embodiment of information affected the memory performance, we compared the error rate for each condition for the different kind of memory questions (cf. Figure 3). These results show a significant difference between embodied and non-embodied within the *with self-interruption* condition ($\chi^2=5,503, p=0,019$) and indicates that the embodied information was better recalled than the non-embodied if the agent uses self-interruptions. This would point towards interesting effects regarding the potential of information exchange in a smart home. However, this needs to be investigated further with embodiment as dependent variable using the same questions in both conditions.

5.2 Subjective Ratings

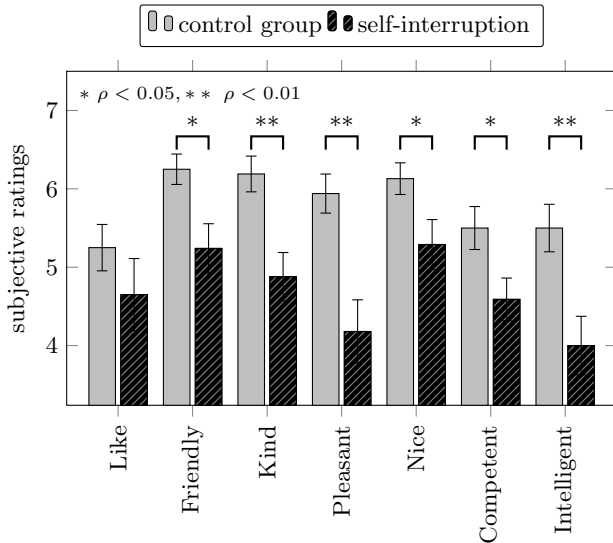


Figure 4: Subjective ratings of the virtual agent. Error bars represent the standard error.

We further wanted to know whether the self-interruptions had any effect on the subjective ratings of Flobi. Figure 4 shows the results of the subjective ratings in the two conditions. The ANOVA about the subjective ratings revealed a significant effect of self-interruption: while in the *control* group Flobi received very high scores for the likeabil-

ity questions (*Like, Friendly, Kind, Pleasant* and *Nice*), the participants in the experimental condition rated the agent significantly less likable (*Friendly* $F=7.296, p=0.011$; *Kind* $F=11.393, p=0.002$; *Pleasant* $F=13.338, p=0.001$; *Nice* $F=4.732, p=0.037$).

We further found a significant effect of condition on the rating of the perceived intelligence of the agent (*Competent* $F=5.573, p=0.025$; *Intelligent* $F=9.582, p=0.004$). While the experimental group rated the intelligence of the agent with 4 the *control* group’s rating was significantly higher (5.5). For other ratings no significant effects were found. Thus, overall, the self-interruption had a very negative effect on the perception of Flobi.

In order to find an explanation for the different judgments we looked more closely at the subjective ratings of the condition *with self-interruption*. During inspection of the videos we observed a strong effect of the attention module in the self-interruption condition: there were two utterances that referred to an object to the left and to the right of the user. If the user was not looking at the object Flobi would repeat the request to look left/right. Flobi would repeat this utterance up to three times. Interestingly, in the *with self-interruption* condition every participant received at least one repetition, some others even two or three repetitions (whereas in the control condition Flobi did not react at all to the user’s gaze and thus did not repeat at all). We therefore divided the *with self-interruption* group into these three (sub-)groups and compared the subjective ratings. Figure 5 shows the

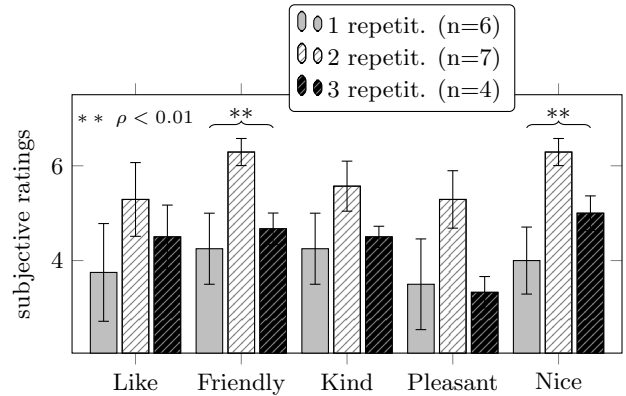


Figure 5: Subjective ratings of the agent for the experimental condition, based on the number of repetitions. Error bars represent the standard error.

results of the three groups. The ANOVA about the subjective ratings for the three groups shows a significant effect of condition on the ratings for *Friendly* ($F=7.018, p=0.008$) and *Nice* ($F=7.350, p=0.007$) between the groups, with the two-repetition group receiving the most positive ratings. For other ratings no significant effects were found.

6. IMPLICATIONS

Although [10] could show that self-interruptions can help to increase cognitive performance, our experiment could not confirm this hypothesis in the interaction with an embodied agent within a smart environment. However, a closer look at the results revealed a range of factors that play an important role and should be included in the attention model.

Embodied Information. The results showed a significant effect of embodiment of information on the memory performance in the *self-interruption* but not in the *control* condition. This is plausible, as in the embodied information case the VFoA is an important cue to gather relevant information whereas in the non-embodied information case the VFoA does not necessarily indicate if the participant's attention is on the (verbally) given information. This has implications on the attention model which needs to distinguish between embodied information and non-embodied information.

Insistence. The results (cf. figure 5) show that the attention guiding strategy for a new focus of discourse has a big impact on the subjective rating of the agent's likeability. Especially, if the participant does not look – or even want to look – at the new focus of discourse and the agent tries to enforce the visual attention repeatedly, this leads to a significant increase in likeability if the participant complies (after 2 repetitions) vs a decrease in likeability if the participant does not want to comply (3 repetition case). The consequence for the attention model is that it should include a threshold for insistence on achieving the desired user attention. Note that this threshold can be learned, and can be dependent on user preferences as well as on information type and relevance.

Speech Synthesis. In the questionnaires we received comments about the perceived monotony of the synthesized voice indicating the importance to provide a more variable prosody to counteract a perceived rudeness of self-interruption. We therefore 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 [5] or react with other subtle signals such as facial expressions. As the perceived rudeness will correlate with the system's insistence, the hesitation should be included as a further factor into the attention system to ameliorate negative impacts on the subjective evaluations.

Validity. The questionnaires as well as the other observations indicate that the verbal attention guiding in combination with the looking behavior of the agent has worked very well, as in the self-interruption condition each participant did indeed look once at each focus of discourse. An intensive analysis of the looking behavior of the participants will follow.

However, it remains unclear if the positive memory effect of self-interruption in [10] occurred due to self-interruption alone or was also caused by the fact that a part of the utterance was repeated at the re-start of the interrupted utterance. In our experiment the novelty effect of the virtual agent in the smart environment may cover a less strong effect, rendering both participant groups equally attentive. We will explore in further studies how the positive effect on the memory performance can be replicated in our setting.

7. CONCLUSION

We presented a first exploratory human-agent interaction experiment investigating the effect of a self-interrupting agent in a smart home environment on the memory performance of interaction partner and the subjective ratings of the agent. While we were not able to replicate the results from [10] we found a range of factors that need to be adapted in our attention model due to differences in the human-agent situation.

8. 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).

9. REFERENCES

- [1] M. Argyle and M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press, London, 1976.
- [2] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1):71–81, 2009.
- [3] T. Baumann and D. Schlangen. The inprotk 2012 release. Proc. of the NAACL-HLT Workshop on Future directions and needs in the Spoken Dialog Community: Tools and Data (SDCTD 2012), pages 29–32. ACL, 2012.
- [4] D. Bohus and E. Horvitz. Models for multiparty engagement in open-world dialog. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, SIGDIAL '09, pages 225–234, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.
- [5] D. Bohus and E. Horvitz. Managing human-robot engagement with forecasts and... um... hesitations. In *Proceedings of the 16th International Conference on Multimodal Interaction*, ICMI '14, pages 2–9, New York, NY, USA, 2014. ACM.
- [6] B. Carlmeyer, D. Schlangen, and B. Wrede. Towards closed feedback loops in hri: Integrating inprotk and pamini. In *Proceedings of the 2014 Workshop on Multimodal, Multi-Party, Real-World Human-Robot Interaction*, MMRWHRI '14, pages 1–6, New York, NY, USA, 2014. ACM.
- [7] N. George and L. Conty. Facing the gaze of others. *Neurophysiologie Clinique/Clinical Neurophysiology*, 38(3):197 – 207, 2008.
- [8] P. Holthaus, C. Leichsenring, J. Bernotat, V. Richter, M. Pohling, B. Carlmeyer, N. Köster, S. M. zu Borgsen, R. Zorn, B. Schiffhauer, K. F. Engelmann, F. Lier, S. Schulz, P. Cimiano, F. Eyssel, T. Hermann, F. Kummert, D. Schlangen, S. Wachsmuth, P. Wagner, B. Wrede, and S. Wrede. How to address smart homes with a social robot? a multi-modal corpus of user interactions with an intelligent environment. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, May 2016. European Language Resources Association.
- [9] D. Klotz, J. Wienke, J. Peltason, B. Wrede, S. Wrede, V. Khalidov, and J.-M. Odobez. Engagement-based multi-party dialog with a humanoid robot. Proc. of the SIGDIAL 2011 Conference, pages 341–343. Association for Computational Linguistics, 2011.
- [10] S. Kousidis, C. Kennington, T. Baumann, H. Buschmeier, S. Kopp, and D. Schlangen. Situationally Aware In-Car Information Presentation Using Incremental Speech Generation: Safer, and

- More Effective. In *Proceedings of the EACL 2014 Workshop on Dialogue in Motion*, pages 68–72, 2014.
- [11] J. Lee, S. Marsella, D. Traum, J. Gratch, and B. Lance. The rickel gaze model: A window on the mind of a virtual human. In *International Workshop on Intelligent Virtual Agents*, pages 296–303. Springer, 2007.
- [12] G. Leitner, M. Hitz, A. J. Fercher, and J. N. A. Brown. Aspekte der human computer interaction im smart home. *HMD Praxis der Wirtschaftsinformatik*, 50(6):37–47, 2013.
- [13] I. Lütkebohle, F. Hegel, S. Schulz, M. Hackel, B. Wrede, S. Wachsmuth, and G. Sagerer. The bieiefeld anthropomorphic robot head "flobi". In *2010 IEEE International Conference on Robotics and Automation*, pages 3384–3391. IEEE, May 2010.
- [14] O. Palinko, A. Sciutti, L. Schillingmann, F. Rea, Y. Nagai, and G. Sandini. Gaze contingency in turn-taking for human robot interaction: Advantages and drawbacks. In *24th IEEE International Symposium on Robot and Human Interactive Communication*, pages 369–374. IEEE, 2015.
- [15] K. Pitsch, H. Kuzuoka, Y. Suzuki, L. Sussenbach, P. Luff, and C. Heath. "the first five seconds": Contingent stepwise entry into an interaction as a means to secure sustained engagement in hri. In *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 985–991, Sept 2009.
- [16] C. Rich, B. Ponsler, A. Holroyd, and C. L. Sidner. Recognizing engagement in human-robot interaction. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 375–382, March 2010.
- [17] L. Schillingmann and Y. Nagai. Yet another gaze detector: An embodied calibration free system for the icub robot. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pages 8–13, Nov 2015.
- [18] G. Skantze and J. Gustafson. Attention and interaction control in a human-human-computer dialogue setting. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, SIGDIAL '09, pages 310–313, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.
- [19] M. Staudte and M. W. Crocker. Visual attention in spoken human-robot interaction. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, HRI '09, pages 77–84, New York, NY, USA, 2009. ACM.
- [20] C. Vandewater. 6 ways to get (and keep) students' attention, March 2015. Retrieved May 13, 2016 from <http://teach.com/teach100-mentor/classroom-management>.
- [21] R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt. 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*, CHI '01, pages 301–308, New York, NY, USA, 2001. ACM.
- [22] A.-L. Vollmer, M. Mühlig, J. J. Steil, K. Pitsch, J. Fritsch, K. J. Rohlfing, and B. Wrede. Robots show us how to teach them: Feedback from robots shapes tutoring behavior during action learning. *PLoS ONE*, 9(3):1–12, 03 2014.
- [23] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes. Elan: a professional framework for multimodality research. In *Proceedings of LREC*, volume 2006, page 5th, 2006.
- [24] Z. Yu, D. Bohus, and E. Horvitz. Incremental coordination: Attention-centric speech production in a physically situated conversational agent. In *16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, page 402, 2015.