Maike Paetzel

Uppsala University
Box 337
SE-751 05 Uppsala
Sweden

maike.paetzel@it.uu.se
http://www.maike-paetzel.de/

1 Research Interests

Conversational agents in social contexts need more than optimized task performance skills: In order to keep the human dialogue partner engaged in the conversation, agents need to find a good **balance between task oriented and social behavior**. Different dialogue policies and interaction strategies can highly influence the user's perception of a dialogue system. During my Master thesis project, I was mainly looking into **different dialogue strategies in incremental vs. non-incremental dialogue systems** and how they can influence both task performance and user satisfaction with the system.

My PhD in social robotics shifted the focus more towards interaction strategies for embodied conversational agents. The goal is to investigate how **different embodiments, appearances and interaction strategies can influence the perception of the agent** in a human interaction partner. Mainly, I want to investigate what in dialogue policies can cause a strange or uncanny feeling towards the conversational agent, which ultimately undermines the goal of the conversational agent to act as a social companion.

1.1 Incremental Dialogue Systems and Optimized Policies for Rapid Conversations

As part of the NSF funded project "Incremental Speech Processing for Rapid Dialogue" we developed a conversational agent capable of playing a rapid object matching game, which could demonstrate the values of a fully incremental system architecture and a dialogue policy optimized towards the incremental nature of the system.

The Domain To enforce a conversation with fast turn-taking behavior between dialogue partners, we developed the Rapid Dialogue Game Image (RDG-Image) (Paetzel et al., 2014). In this game, two players need to cooperate in order to score as high as possible in a given time limit. Both see the same set of eight images on their screen in shuffled order. One of the players acts as a director, who sees the target image highlighted on the screen. The director now describes the image verbally to the matcher, who needs to select the same image on the matcher's own screen. Once the matcher is confident about the selection, the director can request the next target image.

Corpora Two corpora were collected including two human partners playing the game. The RDG-Image Lab corpus was recorded on-site at the Institute for Creative Technologies, University of Southern California, and includes 32 pairs of gameplay (Paetzel et al., 2014). As this data collection was cost and time intensive, Manuvinakurike and DeVault (Manuvinakurike and DeVault, 2015) developed a web framework to crowdsource the data collection. On Amazon Mechanical Turk, 98 human-human game plays could be recorded for the RDG-Image Web corpus.

The Eve Agent During my Master thesis project, we developed an automated agent called **Eve** which is able to autonomously play the role of the matcher in the RDG-Image game. We designed a dialogue system from scratch which allows to easily switch between incremental and non-incremental processing in different modules of the system. By testing a fully incremental version of the system against a semi-incremental and non-incremental version, we could highlight the value full incremental processing adds both to the game performance and user satisfaction (Paetzel et al., 2015)(Manuvinakurike et al., 2015b). In addition, we could show that a dialogue policy optimization which is fitted towards the incremental nature of the architecture is crucial to use the full possibilities of the dialogue system (Paetzel, 2015).

1.2 Interaction Strategies for Embodied Conversational Agents

In my PhD, the focus of my research shifted towards interaction with embodied agents. One main goal of my thesis is to investigate how different interaction and dialogue strategies can cause or prevent an uncanny or eerie feeling in humans. During my first year, I conducted a series of pilot studies to understand the perception of the robot platform Furhat in general and the perception of gender in particular. Furhat is a back-projected robot head, which allows to model facial features and expressions in an easy and cost-effective manner. In a first pilot study, we could show that the robot platform is indeed suitable for research on uncanny feelings and multimodal interactions (Paetzel et al., 2016). Especially, visual incongruent gender cues raised an uncanny feeling towards

the robot in adults. In a second study, we used incongruent multimodal gender cues in a short interaction with children. Here, for various reasons, we did not find a link between incongruence and uncanniness. However, we observed that children rely more on auditory than on visual cues when judging the perception of a robot, which supports future plans of investigating different dialogue strategies and their effect towards the feeling of uncanniness.

2 Future of Spoken Dialog Research

Up to date, most end users still prefer commanding their computer devices by keyboard, mouse or touch input, even though the usability of state-of-the-art dialogue systems, for example in mobile applications, have improved during the last years. However, since users have a different option of interaction which they are familiar with for years, the scepticism towards language interaction with computers is still high. With the raise of robots in domains like household and health care, the demand for natural interaction with computer systems will likely increase in the next five to ten years, as robots don't offer another option of interaction than spoken language, mimic and gesture.

This development can be boon and bane at the same time: When dialogue research becomes much more popular, a more natural interaction with machines might be achieved quicker. On the contrary, interaction with robots adds huge challenges to the field of dialogue systems:

- Instead of dialogue narrowed down to a specific use-case or domain, open-domain dialogue abilities will be required.
- While many current systems use massive computation powers on server farms to process spoken input, people might have privacy concerns if their robot is sending all microphone input from a home environment to external processing. In addition, good internet connections are still not available everywhere. For this reasons, the trend might move back towards better offline algorithms with limited resources.
- Personalized individual robots will require personalized dialogue behaviour, with the ability to memorize preferences of multiple people over the period of years.
- As natural interactions prohibits the use of headsets, dialogue systems will need to become much more robust towards background noise.

3 Suggestions for Discussion

As my main focus is in multimodal interaction with embodied agents, I would be interested in discussing the following questions:

- The influence of embodiment: How does the type of embodiment (2D vs 3D) and the appearance influence the language and dialogue strategies humans use in their conversation with the dialogue system?
- Evaluation: How can we find standardized metrics to evaluate conversational agents across domains and embodiments?
- How can incremental processing be scaled up to processing and generating multimodal interactions?

References

Ramesh Manuvinakurike and David DeVault. 2015. *Pair Me Up: A Web Framework for Crowd-Sourced Spoken Dialogue Collection*. Natural Language Dialog Systems and Intelligent Assistants.

Ramesh Manuvinakurike, Maike Paetzel and David DeVault. 2015. *Reducing the Cost of Dialogue System Training and Evaluation with Online, Crowd-Sourced Dialogue Data Collection*. The 19th workshop on the semantics and pragmatics of dialogue (SemDial).

Maike Paetzel, David Nicolas Racca and David DeVault. 2014. *A Multimodal Corpus of Rapid Dialogue Games*. Language Resources and Evaluation Conference (LREC).

Maike Paetzel, Ramesh Manuvinakurike and David DeVault. 2015. *"So, which one is it?" The effect of alternative incremental architectures in a high-performance game-playing agent*. Special Interest Group on Discourse and Dialogue (SIGDIAL).

Maike Paetzel. 2015. *Exploring the effect of incremental speech processing on dialogue policy performance in a game-playing agent*. Master thesis.

Maike Paetzel, Christopher Peters, Ingela Nystrm and Ginevra Castellano. 2016. *Preliminary results from using a back-projected robot head in uncanny valley research*. Extended abstract at the RO-MAN Interactive Session.

Biographical Sketch

Maike Paetzel just finished her first year as a PhD student at Uppsala University, Sweden, under the supervision of Ginevra Castellano. In her Bachelor and Master studies at the University of Hamburg, Germany, she focused both on robotics and embodied agents as well as dialogue systems. In her PhD in social robotics, she is now working on bringing the two fields of interest together.