This is the published version of a paper presented at Young Researchers Roundtable on Spoken Dialogue Systems.

Citation for the original published paper:


N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-257885
1 Research interests

My research interests lie generally in the area of embodied multimodal interfaces, with a special focus on interaction capabilities in such systems. Particularly, I am interested in multimodal communication between humans and machines with a focus on visual attention and natural language during dialogue. In my PhD so far (Kontogiorgos, 2017), I start with computationally describing certain elements of interaction and conversational dynamics in humans, and model such behaviours in embodied conversational interfaces. I then evaluate how such systems are perceived by naive participants, when changing the verbal and non-verbal behaviour and embodiment of these systems. Such behaviours are annotated and modelled from human-human interactions, and influenced by previous work in psycholinguistic studies.

1.1 Human-human interaction

I am particularly interested in how humans establish, maintain and repair common ground, in a process defined in literature as grounding (Clark et al., 1991). The approach I have been adopting so far is to take incremental steps on the analysis of human-human conversation collected corpora (Kontogiorgos et al., 2018a), and inspired from a conversational analysis point of view (Schegloff, 2007), make use of multimodal data from speakers and listeners to infer the current state of the conversation.

Using a framework for multisensory input (Jonell et al., 2018), I have used low-level multimodal signals to extract high level features that represent the humans’ current state of uncertainty, or features that represent when humans establish common ground in reference resolution during task-oriented dialogues (Kontogiorgos et al., 2018b).

1.2 Human-robot interaction

The main idea is therefore to inspire studies that use models from analyses in human-human data, and then test in manipulation with social robots, as embodied spoken dialogue systems. Such systems need to infer the user state and intention using real-time input from multimodal features. Given the complexity of modelling open-world dialogues (Bohus and Horvitz, 2009), I focus my PhD work in instructional HCI systems, and therefore task-oriented dialogue, where such systems need to be able to deal with communication breakdowns and user uncertainty, and detect incrementally when humans require more information to adapt their repair strategies.

I am interested in applying the outcome of previous work in two types of embodied interfaces: smart-speakers that can only interact using speech, and anthropomorphic social robots that can portray non-verbal behaviour along with voice when they interact with humans. So far, I have experimented with manipulating these embodiments in interactions with humans and our results have shown that human interactive behaviour shifts when changing the embodiment and non-verbal behaviour of a conversational agent (Kontogiorgos et al., 2019a,b,c).

2 Spoken dialogue system (SDS) research

Where do you think the field of dialogue research will be in 5 to 10 years? Research in dialogue systems dates a few decades, but is now more timely than ever, as spoken dialogue systems have currently become ubiquitous in users’ homes. This in fact, helps dialogue researchers finally understand how people interact with these devices in large scale, and how do these people adapt to the systems’ capabilities. While information retrieval is the main functional characteristic of these devices, understanding language and the nuances of conversation in humans remains largely unsolved. I think, in the next 5-10 years we will encounter a lot of research in conversation dynamics and language understanding, and we may even see the first devices that are able to converse with humans further from factual question answering and information retrieval based interactions.

What kind of questions need to be investigated to get the field to that point? I think one should start by asking why do users even need a conversational interface that they can engage in dialogue with. We still do not know if users will actually spend time in conversation with such devices, or if factual question answering should suffice. I would also question how should such devices be embodied, as smart speakers have been successful, however more and more social robot companies get to close doors. Why are social robots preferred by users in research studies, but never make it to users’ homes?
What are the most important things for users of SDSs? Another aspect which is important for users is to bond with devices, therefore more research in long-term interaction is needed. Dialogue systems should remember what users like, what they have talked about in the past and what could they be interested in the future.

3 Suggested topics for discussion

- How should dialogue systems be embodied and why does it matter?
- Why do deep learning methods work so well in restricted environments, but are still not able to model conversation?
- What is the societal impact and consequences of machine generated text? Should dialogue systems be regulated to limit their capabilities?

References


Dimosthenis Kontogiorgos, Vanya Avramova, Simon Alexandersson, Patrik Jonell, Catharine Oertel, Jonas Beskow, Gabriel Skantze, and Joakim Gustafsson. 2018a. A multimodal corpus for mutual gaze and joint attention in multiparty situated interaction. In LREC.

Dimosthenis Kontogiorgos, Andre Pereira, Olle Andersson, Marco Koivisto, Elena Gonzalez Rabal, Ville Vartiainen, and Joakim Gustafson. 2019a. The effects of anthropomorphism and non-verbal social behaviour in virtual assistants. In International Conference on Intelligent Virtual Agents. ACM.


Biographical sketch

Dimosthenis Kontogiorgos is a PhD Student at KTH Royal Institute of Technology and works in the field of human-robot interaction and multimodal conversational interfaces. He is interested in interaction and computational modelling in the area of grounding through the use of conversational cues and visual attention. He applies his research in user studies with social robots in different forms of embodiment, and in interactions with humans.

He got his Master degree in computer science with focus in human-computer interaction at the University of Copenhagen, and Bachelor degree in Informatics from the Technological Educational Institute of Athens. Throughout this PhD, he currently works in a project funded by the Swedish Foundation for Strategic Research (SSF), where in collaboration with robotics researchers, the project interests are based on robots that can collaborate with humans.

Acknowledgements

I would like to thank my supervisor Joakim Gustafson and my co-supervisor Gabriel Skantze for the support and advice so far on my PhD.