

10.3 Incremental Dialogue Processing: Architecture and Lower-level Components

Timo Baumann (baumann@informatik.uni-hamburg.de)
Supervisor: Wolfgang Menzel

My research is geared towards interaction management in spoken dialogue systems. Specifically, I am interested in the **timing** of dialogue and dialogue-related phenomena, especially on the sub-turn level and its relation with other modalities such as gaze and gesture. We believe that a dialogue system should be able to both show reactions while a user is speaking, and to integrate reactions into its own behaviour while it itself is speaking.

For a dialogue system to react to a user while she is still speaking, it is necessary for the system to work **incrementally**, that is, to process the user's utterance while it is ongoing, and to come up with partial conclusions about what the user is saying, what the system should answer and how certain this is. Going one step further, I am also interested in **proactively** generating output, that is, to **predict** a short distance into the future in order to overcome delays or to –gasp– cut short the user. While conventional systems could only be 'sluggish' or 'fast enough', a proactive system's **timing** may try to temporally align to the user (or to deliberately break the alignment). Likewise, an incremental system's output must be produced incrementally, so as to allow the integration of user reactions into ongoing utterances.

In my work so far I have analyzed properties of incremental speech recognition, short-term prediction of dialogue flow, and recently incremental speech synthesis. Combined, these 'building blocks' allow for truly incremental dialogue systems that allow for much improved naturalness of interaction, the prompt integration of information from other modalities (such as gaze, prosody, or gesture), and as a combined result, better dialogue systems.