

Timo Baumann

University of Potsdam
Department for Linguistics
Karl-Liebknecht-Str. 24
D-14476 Potsdam
Germany

timo@ling.uni-potsdam.de
<http://www.ling.uni-potsdam.de/~timo>

1 Research Interests

My research is geared towards **interaction management** in spoken dialogue systems. Specifically, I am interested in the **timing** of dialogue and dialogue-related phenomena, and in taming dialogue systems to respond as quickly and in similar ways as humans. For a dialogue system to react to a user while she is still speaking, it is necessary for the system to run **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 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 traditionally the system could only be sluggish or fast enough, a proactive system's **timing** must try to temporally align to the user (or to deliberately break the alignment). I believe, that **prosody** plays a vital role in everyday conversation and that it is still too often ignored due to a prevalence of written language. I believe that a leap in spoken dialogue systems design and performance will result from considering prosodic information across the board and more generally from a **dense coupling** in the SDS' architecture.

1.1 Incremental Processing

In a modular system, an *incremental module* is one that generates (partial) output while input is still ongoing. I have thoroughly investigated incremental ASR which outputs word and sub-word hypotheses while recognition is still ongoing (Baumann et al., 2009a). We developed measures to be able to assess the incremental behaviour of ASR. These deal with how often hypotheses change (every change means that consuming modules have to re-process their input) and others describe timing properties of when words are first considered and first decided upon by the ASR. I showed influences between optimizing timing and change measures and derived a measure of certainty from the different timing measures. Having assessed incremental properties of ASR, I analysed methods to improve incremental behaviour by simple (and generic) filtering approaches (Baumann et al., 2009a).

Together with my colleagues, we applied the work on evaluation of incremental components to semantic interpretations (Atterer et al., 2009; Heintze et al., 2010), the evaluation of incremental reference resolution (Schlangen et al., 2009), and to n-best processing in both ASR and semantic interpretation (Baumann et al., 2009b). As part of our venture into incremental analysis, we built a toolkit to process and visualize incremental data (Malsburg et al., 2009).

1.2 Predictive Processing

In an SDS, some processing latencies are inevitable. Hence, for reactions to be *right on time*, they must be issued *before the fact*. In other words, for natural interaction, an SDS must anticipate future events (e. g. that a back-channel or speaker change will be required soon) and predict when exactly it should be placed. We have investigated end-of-turn (EoT) prediction (Baumann, 2008) and end-of-utterance (EoU) detection (Atterer et al., 2008) using rather crude acoustic-prosodic features. In (Baumann, 2008), I presented a system for turn-taking simulation, which stress-tests EoT prediction. Two artificial dialogue participants converse with each other; the resulting turn-taking behaviour is similar to human-human dialogue. While the need for EoT detection is obvious, EoU detection can be important for processing more complex user turns, and for system feedback at TRPs. We were also able to predict the remaining duration of a speaker's words that are *currently ongoing* (Baumann and Schlangen, 2011). This allows the system to precisely time its contributions to the user. Immediate system reactions result in a tight feedback loop between the user and the system and the system must be aware of the fact that its feedback will immediately influence the user's actions.

1.3 Future Work

Micro-temporal considerations of incrementality have so far only been applied to the input side of dialogue processing, with the NLG and TTS components remaining on a "large" scale (i. e. > 1 sec. chunks and latencies). I would like to focus on incremental and low-latency output generation and to combine it with input processing in order to improve interaction capabilities of SDSs.

2 Future of Spoken Dialogue Research

Currently deployed SDSs are mostly tailored towards information access and simple tasks. To some extent they can be seen rather as VUIs than as full dialogue.

I believe that in the future, dialogue systems will appear as **conversational assistants** in many areas, such as hospitals, for elderly people, in tutoring (not only for foreign language learning, but in all areas), and as more natural interfaces for general-purpose personal digital assistants. Computer games make for an especially interesting sandbox for advanced SDSs, as domains are controlled and consequences of errors are small, while demand for “new stuff” and enthusiastic users are plentiful.

While human-like behaviour is not needed or could even distract in simple task-oriented systems, human-like behaviour becomes more important for future applications, as they will be less recognized as tools but as real interlocutors. For better intuitivity, **interaction behaviour** (turn-taking, and -yielding, understanding and hinting below the content level) must be improved.

3 Suggestions for Discussion

- *Turn-taking vs. continuous interaction*: Engineers of applied dialogue systems think of “barge-ins” when they talk about flexibility in their system’s turn-taking scheme. While the *turn-by-turn paradigm* helps to arrange contributions to dialogue conceptually, I believe that it is becoming a handicap in dialogue research and development, as it barely reflects “real” dialogue, in which people constantly interact, give feedback about understanding, consent, etc. with much of this interaction happening on the sub-word level.
- *Building truly incremental SDSs*: Tightly coupled with the previous point, I believe that only truly incremental SDSs (which happen to be my very research objective) can overcome the problems of turn-by-turn processing. While mostly an engineering problem at its core, incremental SDSs allow for plenty of interesting research opportunities.
- *What to say vs. when to say it*: I believe that *micro-temporal* aspects of dialogue are, if not understudied, too often ignored in applied research. This is largely due to technical difficulties with the precise alignment of cut-ins, back-channels, mimics and gesture. At the same time, I believe that there is a huge difference between a dialogue system acknowledging (with a back-channel) at precisely the “right” moment, or as little as 150 ms off. This timing difficulty may account for the fact that good back-channeling is not yet available, even though I believe that it would have a tremendous influence on dialogue system acceptability and efficiency.

References

- Michaela Atterer, Timo Baumann, and David Schlangen. 2008. Towards Incremental End-of-Utterance Detection in Dialogue Systems. In *Procs. of Coling 2008*, pages 11–14, Manchester, UK.
- Michaela Atterer, Timo Baumann, and David Schlangen. 2009. No Sooner Said Than Done? Testing the Incrementality of Semantic Interpretations of Spontaneous Speech. In *Procs. of Interspeech 2009*, pages 1855–1858, Brighton, UK.
- Timo Baumann and David Schlangen. 2011. Predicting the Micro-Timing of User Input for an Incremental Spoken Dialogue System that Completes a User’s Ongoing Turn. In *Procs of SigDial 2011*, Portland, USA.
- Timo Baumann, Michaela Atterer, and David Schlangen. 2009a. Assessing and Improving the Performance of Speech Recognition for Incremental Systems. In *Procs. of NAACL-HLT 2009*, pages 380–388, Boulder, USA.
- Timo Baumann, Okko Buß, Michaela Atterer, and David Schlangen. 2009b. Evaluating the Potential Utility of ASR N-Best Lists for Incremental Spoken Dialogue Systems. In *Procs. of Interspeech 2009*, pages 1031–1034, Brighton, UK.
- Timo Baumann. 2008. Simulating Spoken Dialogue With a Focus on Realistic Turn-Taking. In *Procs. of the 13th ESSLLI Student Session*, pages 17–25, Hamburg, Germany.
- Silvan Heintze, Timo Baumann, and David Schlangen. 2010. Comparing Local and Sequential Models for Statistical Incremental Natural Language Understanding. In *Procs. of SigDial 2010*, Tokyo, Japan.
- Titus von der Malsburg, Timo Baumann, and David Schlangen. 2009. TELIDA: A Package for Manipulation and Visualisation of Timed Linguistic Data. In *Procs. of SigDial 2009*, pages 302–305, London, UK.
- David Schlangen, Timo Baumann, and Michaela Atterer. 2009. Incremental Reference Resolution: The Task, Metrics for Evaluation, and a Bayesian Filtering Model that is Sensitive to Disfluencies. In *Procs. of SigDial 2009*, pages 30–37, London, UK.

Biographical Sketch



Timo Baumann is a research assistant and PhD candidate at the University of Potsdam, Germany, working under the supervision of David Schlangen.

He previously studied computer science, phonetics and linguistics at Hamburg University and received his master’s degree in 2007 for work on prosody analysis carried out at IBM Research.

In his free time, Timo likes to go hiking or cycling and sings in a choir. He prefers organic food and is interested in renewable energies. Also, Timo enjoys to stay abroad and live on student grants (USA 1997, Switzerland 2003, Spain 2005, Sweden 2009).