Feedback in Adaptive Interactive Storytelling

Timo Baumann

Department for Informatics
University of Hamburg, Germany
baumann@informatik.uni-hamburg.de

Abstract

Telling stories is different from reading out text: a speaker reponds to the listener's feedback and incorporates this into the ongoing talk. However, current computer systems are unable to do this and instead just non-attentively read out a text, disregarding all feedback (or the absence thereof). I propose and discuss an idea for a small research project and a plan for how an attentive listening storyteller can be built.

Index Terms: Incrementality, Feedback, Storytelling, Adaptation, Prosody

1. Introduction

Interactive storytelling [1] deals with adaptations to stories based on listener input to change the content of the story (making this area somewhat similar to computer games). So far, such adaptations happen at a relatively coarse granularity and user input is integrated only with some delay. Recent advances in incremental speech processing technology [2, 3] enable adaptations to happen at a much finer granularity. Thus, it is not only possible to react to a user's verbalized requests by changing the *content* of the story, but also by adapting the *delivery* of the story, based on reactions to concurrent feedback utterances (or the absence thereof).

I consider interactive storytelling (or more to the point: interactive story delivery) an ideal testbed for fine-granular, micro-temporal turn-taking behaviour, as the content and types of adaptations are fully controlled by the system, far more than for task-driven dialogue systems.

2. Research and Development Goals

I believe that micro-temporal interactive storytelling can help in several areas of feedback research.

- Develop techniques to elicit feedback in an interactive system, which involves building speech synthesis systems with *conversational* abilities. (Work on eliciting feedback utterances for spoken dialogue systems [4] nonetheless will also apply to interactive storytelling.)
- Advance ASR and VAD technology to reliably recognize feedback utterances in an incremental fashion to allow for timely adaptation; this includes recognizing para-linguistic phenomena such as inhalation, lip-smack, and the like.
- Develop techniques to incorporate feedback into the ongoing speech.

Once these basic technological questions have been addressed, such a system can be used to further study feedback behaviour in a controlled way by deliberately manipulating the feedback (and turn-taking) behaviour and analyzing the users' reactions, for example studying entrainment phenomena.

3. Plan and System Architecture

At first, a system that tells a short, cyclical story would be conceived, that reacts to feedback utterances (depending on whether they occur in expected places) change its emotional state (which, in turn, changes the system's speech delivery). The overall architecture of the system is depicted in Figure 1 and a short utterance plan [5] for the beginning of a story is depicted in Figure 2. As can be seen, depending on the system's affective state, it might (somewhat aggressively) require a feedback utterance before continuing. User feedback is analyzed for prosody, content, and expectancy, and results in changes to the system's affective state. The system will be implemented in

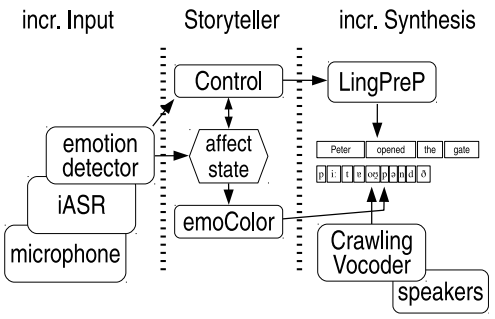

Figure 1: Overview of system modules and the hierarchic structure of incremental units describing an example utterance as it is being produced and adapted during delivery.

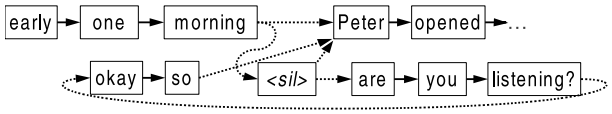

Figure 2: Possible changes to an ongoing utterance for a listening storyteller (which might be in a slightly annoyed mood).

INPROTK [6], an incremental dialog processing toolkit that allows to interconnect modules that exchange their information as *incremental units* [7].

Following Schröder [8], the affective state can be modelled along the two dimensions *activation* and *evaluation* which are also the basis for the rules in the speech synthesis component *emoSpeak* [9] which could be re-used. As a first shot, and without referring to the wealth of literature on affect recognition (e. g. [10]), we would use some simple rules-of-thumb: Every listener feedback increases *activation* while the magnitude is determined by the type of feedback (or words in the utterance) and by the user's pitch-range. Pitch-range and pitch-variability are also the influencing factors for the *evaluation* parameter. A gravitational pull for the evaluation dimension towards 0 and for activation towards $-\infty$ will lead the system to forget about positive or negative emotions and to eventually become inactive if not given any feedback in the long term.

Extensions of this first system would use improved techniques to elicit feedback utterances, include real story interactivity (i. e., allowing the user some freedom to change parts of the story), and support for deviations from ideal behaviour as outlined above.

4. Conclusions

Adaptive interactive storytelling is a promising 'micro-domain' [11] to understand and study feedback and micro-turn-taking behaviours, that should (following the idea behind micro-domains) later be applicable in full-blown conversational dialogue systems. Interactive storytelling has all limitations necessary for a micro-domain while at the same time offering a high flexibility to study feedback utterances and micro-temporal behaviour in dialogue. Interactive storytelling is especially interesting as it is a purely system-driven domain (and with the system doing the majority of the talking), which ensures a higher proportion of feedback utterances from the user than in typical task-driven dialogue systems.

The author is grateful to fruitful discussions on the topics presented and hopes to get, of course, feedback on the ideas presented.

5. References

[1] A. Glassner, *Interactive storytelling: Techniques for 21st century fiction*. AK Peters, 2004.

[2] T. Baumann and D. Schlangen, "Predicting the Micro-Timing of User Input for an Incremental Spoken Dialogue System that Completes a User's Ongoing Turn," in *Proceedings of SigDial 2011*, Portland, USA, 2011.

[3] H. Buschmeier, T. Baumann, B. Dorsch, S. Kopp, and D. Schlangen, "Combining incremental language generation and incremental speech synthesis for adaptive information presentation," in *Procs. of SigDial*, Seoul, Korea, 2012, pp. 295–303.

[4] T. Misu, E. Mizukami, Y. Shiga, S. Kawamoto, H. Kawai, and S. Nakamura, "Toward construction of spoken dialogue system that evokes users' spontaneous backchannels," in *Proceedings of the SIGDIAL 2011 Conference*, Association for Computational Linguistics. Portland, Oregon: Association for Computational Linguistics, June 2011, p. 259–265. [Online]. Available: http://www.aclweb.org/anthology/W/W11/W11-2028

[5] G. Skantze and A. Hjalmarsson, "Towards incremental speech generation in dialogue systems," in *Proceedings of SIGdial*, Tokyo, Japan, September 2010.

[6] T. Baumann and D. Schlangen, "The INPROTK 2012 release," in *Proceedings of SDCTD*, Montréal, Canada, 2012.

[7] D. Schlangen and G. Skantze, "A General, Abstract Model of Incremental Dialogue Processing," in *Proceedings of the EACL*, Athens, Greece, 2009, pp. 710–718.

[8] M. Schröder, "Dimensional emotion representation as a basis for speech synthesis with non-extreme emotions," in *Proc. Workshop on Affective Dialogue Systems*, 2004, pp. 209–220.

[9] M. Schröder, R. Cowie, E. Douglas-Cowie, M. Westerdijko, and S. Gielen, "Acoustic correlates of emotion dimensions in view of speech synthesis," in *Proceedings of Interspeech*, 2001, pp. 87–90.

[10] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.

[11] J. Edlund, J. Gustafson, M. Heldner, and A. Hjalmarsson, "Towards human-like spoken dialogue systems," *Speech Communication*, vol. 50, pp. 630–645, 2008.