How Should Attentive Speaker Agents Adapt to Listener Feedback?

Hendrik Buschmeier, Stefan Kopp

Sociable Agents Group, CITEC and Faculty of Technology, Bielefeld University
PO-Box 100131, 33501 Bielefeld, Germany
{hbuschmeier,skopp}@uni-bielefeld.de

Introduction In dialogue, a speaker's communicative actions are not only shaped by her communicative intentions and her ability to express them linguistically, but importantly also by the actions and reactions of her interlocutor, which she – being cooperative – cannot ignore. Both dialogue partners contribute to dialogue success and collaborate on the interaction to make it as efficient as possible by responding to each other's needs.

One important mechanism for dialogue coordination is *communicative feedback* in the form of short verbal-vocal expressions (such as 'uh-huh', 'yeah', 'huh?'), head movements (nods, shakes, etc.), facial expressions (e.g., smiling, frowning, raising an eyebrow) and gaze. The use of feedback is prevalent in spoken interaction. Listeners often produce it concurrently to the speaker's communicative actions and convey that they are in contact with the speaker, whether they perceive and understand what the speaker says, whether they accept, adopt or agree with the speaker's utterance and also further attitudes towards it [1].

By providing feedback, a listener thus reveals parts of his mental state and indicates in a timely manner how the interaction is going or which attitude he has towards an utterance. On the basis of feedback, a speaker can then reason about the listener's mental state and use this information to adapt her subsequent utterances to the listener's needs. If, for example, the listener frowns right after the speaker mentions an object which the speaker thinks she unambiguously referred to, she can use this evidence of difficulties of understanding and clarify the reference by providing additional information.

Attentive speaker agents Currently, artificial conversational agents (such as dialogue systems or embodied virtual agents) lack capabilities to deal with user feedback. This is one aspect why interacting with them is often cumbersome. Users are forced to communicate meta-information on the state of the conversation explicitly and adhere to strict turn-taking behaviour while doing it. And if agents can react to this information at all, they do so no sooner than in their next utterance.

In previous work [2], we proposed that conversational agents should be *attentive speakers*, which we define as being able to (i) elicit feedback from users; (ii) detect and interpret concurrent user feedback; and (iii) respond to feedback by adapting their conversational actions to accommodate the user's needs. Such an attentive speaker agent can determine problems as soon as they become evident and is thus able to respond immediately by adapting the still unspoken part of its current utterance.

In the above mentioned work, we also presented a first approach towards conversational agents that can attend to and adapt to communicative listener feedback. The agent, which assists users in organising their weekly calendar, attributes a simple numerical model of *listener state* ($C, P, U, A, dU, dP \in [0, 1]$) to the user. These values are updated when feedback signals (head gestures, simple feedback expressions, gaze) are encountered. Based on this *attributed listener state*, the agent's incremental natural language generation component then changes parameters and constraints that shape the form of the unspoken increments of the utterance. When users show difficulties understanding what the agent means, for example, redundancy is introduced by making implicit communicative effects explicit.

Corpus analysis Here, we present first results from a dialogue study of human–human interactions in the calendar domain. We analyse the semantic and pragmatic properties of listeners' feedback signals as well as speakers' utterances in their vicinity.

Listeners' feedback signals are annotated on multiple dimensions. We classify them according to their basic communicative function and also look for signs of uncertainty, progressiveness and attitude that are often conveyed. We use this information to reason about a listener's mental state with respect to the utterance a feedback signal refers to.

Speakers' utterances, on the other hand, are analysed for their illocutionary force as well as for grounding status and information state of their content (e.g., is it new and possibly unexpected or already known to both interlocutors). The utterance parts succeeding listers' feedback signals are further analysed with respect to the parts preceding them. We do this in order to find out whether new information is introduced and what role it plays; whether old information is clarified or used redundantly; whether implicit content is made explicit; etc.

The insights gained from this corpus analysis will be used to inform the design of a Bayesian model of the listener that takes the speaker's utterances, contextual factors as well as the listener's feedback signals into account when reasoning about the attributed listeners state [3]. The analysis of feedback-succeeding utterance parts will help us identify and implement additional adaptation mechanisms and strategies for the natural language generation component of the attentive speaker agent.

References

[1] J. Allwood, J. Nivre, and E. Ahlsén, "On the semantics and pragmatics of linguistic feedback," *Journal of Semantics*, vol. 9, pp. 1–26, 1992.

[2] H. Buschmeier and S. Kopp, "Towards conversational agents that attend to and adapt to communicative user feedback," in *Proceedings of the 11th International Conference on Intelligent Virtual Agents*, Reykjavik, Iceland, 2011, pp. 169–182.

[3] ——, "Unveiling the Information State with a Bayesian model of the listener," in *SemDial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue*, Los Angeles, CA, 2011, pp. 178–179.