Extended abstract presented at *the 5th European Symposium on Multimodal Communication*, Bielefeld, Germany, 2017.

Conversational Agents Need to Be 'Attentive Speakers' to Receive Conversational Feedback from Human Interlocutors

Hendrik Buschmeier, Stefan Kopp
Social Cognitive Systems Group, CITEC and Faculty of Technology, Bielefeld University
{hbuschme,skopp}@techfak.uni-bielefeld.de

Conversational listener feedback – small verbal/vocal signals such as 'mhm', 'yeah', 'huh?', head gestures, and facial expressions (Allwood et al., 1992) – is an important mechanism in dialogue that enables interlocutors to efficiently coordinate meaning and understanding. This mechanism could also be important for the interactive coordination between humans and artificial conversational agents. In previous work (Buschmeier and Kopp, 2011, 2014), we defined that agents capable of dealing with listener feedback, so called 'attentive speaker agents', should be able (i) to interpret their interlocutors' listener feedback, (ii) to incrementally adapt their language production to the interlocutors' needs, and (iii) to elicit feedback when needed.

Previous research found evidence that humans are generally willing to provide listener feedback in response to feedback elicitation cues produced by artificial conversational agents (Misu et al., 2011; Reidsma et al., 2011). In the present work, we investigate the influence that a conversational agent's behaviour and capabilities have on its human interlocutors' willingness to provide communicative listener feedback. We present the results of a semi-autonomous interaction study between three embodied conversational agents with different feedback processing behaviours: the attentive speaker agent (AS), which pays attention to its interlocutors' feedback, attributes listening-related mental states to them, incrementally adapts its behaviour accordingly, and elicits feedback when needed; a control agent (EA), which does not pay attention to its interlocutors' feedback, but explicitly asks, following each presentation, whether it should continue or repeat; and a control agent (NA), which does not pay attention to its interlocutors' feedback and does not adapt its behaviour at all (see fig. 1).

The study followed a between-subject design. Each participant interacted with one of the embodied conversational agents in an information presentation task. Participants were told that they cannot speak freely with the agent, but may provide multimodal listener feedback,

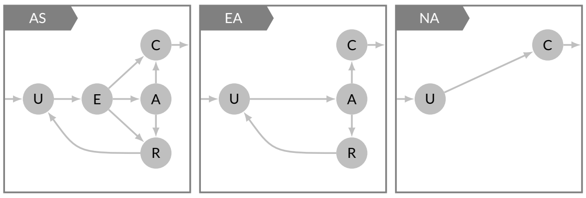

Figure 1: Dialogue phases consist of two to three information presentation units, the structure of which differs depending on the experimental condition (AS/EA/NA), as the flow charts schematically illustrate. Nodes represent the following actions: U – present information in an incremental utterance; E – evaluate current attributed listener state, decide what to do next, and describe this to the participant; C – continue with next unit; R – repeat this unit; A – ask interlocutor whether to repeat or continue.

which the agent might take into account in its own behaviour. The agent talked about calendar items and changes to the calendar (e.g., 'The events are: on Tuesday from 13 to 14 o'clock Lunch and directly afterwards from 14 to 16 o'clock Math 101.'), which participants had to understand well enough to be able to recall afterwards. Information was presented in six dialogue phases (each consisting of two to three presentation phases), which were followed by a recall phase in which calender items needed to be written down in a paper calendar template.

Participants' feedback behaviour (verbal/vocal utterances and head-gestures) was annotated following the conventions of the ALICO-corpus (Malisz et al., 2016). In total, participants produced 734 feedback signals, 127 (17.3%) of which were unimodal verbal/vocal signals, 296 (40.3%) were unimodal head gestures, and 311 (42.4%) were bimodal signals in which a verbal/vocal feedback expression and a head gesture unit are produced in overlap. Analysing these feedback signals, we found their form and distribution to be comparable to feedback in human dialogue (Malisz et al., 2016).

Having seen that participants are, in general, willing to provide feedback to the artificial conversational agents, we analysed whether the different agents' behaviour influenced participants' feedback rate. Participants in condition AS had a mean feedback rate – number of feedback signals per presentation – of $M = 1.97$ ($Mdn = 1.93$, $SD = 0.22$). Participants in condition NA followed with a mean feedback rate of $M = 0.65$ ($Mdn = 0.56$, $SD = 0.6$). Participants in condition EA only had a mean feedback rate of $M = 0.1$ ($Mdn = 0.41$, $SD = 0.31$). A Bayesian ANOVA yields the Bayes factor $BF_{10} = 1.042e7$, which is considered 'decisive' evidence for the alternative hypothesis that feedback frequencies differ between experimental conditions (against the null hypothesis that only contains the intercept). Post-hoc Bayes factor two-sample t-tests yield a value of $BF_{>0} = 3.934e8$ when comparing AS to EA and $BF_{>0} = 5.148e4$ when comparing AS to NA, both of which are considered 'decisive' evidence for the hypothesis that participants produce more feedback when interacting with the attentive speaker agent. In contrast to this, a Bayes factor t-test only finds anecdotal evidence for a difference in feedback rate between conditions EA and NA ($BF_{01} = 1.467$).

Analysing participants' feedback, we further found that their feedback rates do not vary much over time. In condition AS, the standard deviation of the mean feedback rate across dialogue phases and participants is $SD = 0.23$ and it is even smaller in conditions EA ($SD = 0.08$) and NA ($SD = 0.11$). That is, participants in the control conditions stopped providing feedback early on. This could be the case because they felt that providing feedback does not have an effect. An alternative explanation for this might simply be that the attentive speaker agent actively elicited feedback from its interlocutors, which the agents in the control conditions did not do.

To investigate this issue, we analysed how 'effective' (being responded to by participant feedback within a 4 seconds interval) the feedback elicitation cues of the attentive speaker agent were. Across interactions in condition AS the attentive speaker agent had a mean elicitation cue rate (defined analogously to feedback rate) of $M = 1.8$ ($Mdn = 1.86$, $SD = 0.26$), that is, on average 1.8 feedback elicitation cues were produced for each presentation. On average, 61% of the cues were effective. Overall, however, only 54% of participants' feedback signals were preceded by an elicitation cue of the agent, which, in turn, means that 46% of participants' feedback signals were produced 'pro-actively'. Participants in condition AS pro-actively produced 0.91 feedback signals per presentation, which is 9.1, respectively 1.4, times as high as the mean feedback rate of participants in conditions EA and NA. The difference in feedback rate between condition AS and the control conditions thus cannot simply be reduced to the factor that the attentive speaker agent produces feedback elicitation cues.

In summary we can say that (i) in conversation with attentive speaker agents, human interaction partners provide communicative listener feedback that is similar in surface form to feedback that occurs in human–human interaction; (ii) the behaviour of the agent is decisive for its human interaction partner's feedback behaviour; (iii) participants interacting with agents that do not respond to communicative listener feedback seem to become, on some level, aware that providing feedback has no effect – and stop doing it; and (iv) feedback elicitation cues are effective, but the rate of pro-actively produced feedback still exceeds the feedback rate in both control conditions. This suggests that participants that interacted with the attentive speaker agent noticed that their feedback behaviour has an effect on the agent and the interaction.

We conclude that feedback might be a viable coordination mechanism in conversational human–agent interaction if the artificial conversational agent shows an interest in its interlocutors' feedback and responds to it through appropriate adaptation of its behaviour.

Acknowledgements This research was supported by the German Research Foundation (DFG) in the Cluster of Excellence 'Cognitive Interaction Technology' (EXC 277) and by the German Federal Ministry of Education and Research (BMBF) in the project 'KOMPASS' (FKZ 16SV7271K).

Allwood, J., Nivre, J., & Ahlsén, E. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9, 1–26. doi:10.1093/jos/9.1.1

Buschmeier, H. & Kopp, S. (2011). Towards conversational agents that attend to and adapt to communicative user feedback. In *Proceedings of IVA*, pp. 169–182, Reykjavík, Iceland. doi:10.1007/978-3-642-23974-8_19

Buschmeier, H. & Kopp, S. (2014). A dynamic minimal model of the listener for feedback-based dialogue coordination. In *Proceedings SemDial*, pp. 17–25, Edinburgh, UK.

Malisz, Z., Włodarczak, M., Buschmeier, H., Skubisz, J., Kopp, S., & Wagner, P. (2016). The ALICO corpus: Analysing the active listener. *Language Resources and Evaluation*, 50, 411–442. doi:10.1007/s10579-016-9355-6

Misu, T., Mizukami, E., Shiga, Y., Kawamoto, S., Kawai, H., & Nakamura, S. (2011). Toward construction of spoken dialogue system that evokes users' spontaneous backchannels. In *Proceedings of SIGdial*, pp. 259–265, Portland, OR, USA.

Reidsma, D., de Kok, I., Neiberg, D., Pammi, S., van Straalen, B., Truong, K., & van Welbergen, H. (2011). Continuous interaction with a virtual human. *Journal on Multimodal User Interfaces*, 4, 97–118. doi:10.1007/s12193-011-0060-x