Communicative Listener Feedback in Human–Agent Interaction: Artificial Speakers Need to Be Attentive and Adaptive

Socially Interactive Agents Track

Hendrik Buschmeier
Social Cognitive Systems Group
CITEC, Faculty of Technology, Bielefeld University
Bielefeld, Germany
hbuschme@techfak.uni-bielefeld.de

Stefan Kopp
Social Cognitive Systems Group
CITEC, Faculty of Technology, Bielefeld University
Bielefeld, Germany
skopp@techfak.uni-bielefeld.de

ABSTRACT

In human dialogue, listener feedback is a pervasive phenomenon that serves important functions in the coordination of the conversation, both in regulating its flow, as well as in creating and ensuring understanding between interlocutors. This make feedback an interesting mechanism for conversational human–agent interaction. In this paper we describe computational models for an 'attentive speaker' agent is able to (1) interpret the feedback behaviour of its human interlocutors by probabilistically attributing listening-related mental states to them; (2) incrementally adapt its ongoing language and behaviour generation to their needs; and (3) elicit feedback from them when needed. We present a semi-autonomous interaction study, in which we compare such an attentive speaker agent with agents that either do not adapt their behaviour to their listeners' needs, or employ highly explicit ways of ensuring understanding. The results show that human interlocutors interacting with the attentive speaker agent provided significantly more listener feedback, felt that the agent was attentive to, and adaptive to their feedback, attested the agent a desire to be understood, and rated it more helpful in resolving difficulties in their understanding.

KEYWORDS

artificial conversational agents; attentive speaking; listener feedback; adaptation; computational modelling; interaction study

ACM Reference Format:
Hendrik Buschmeier and Stefan Kopp. 2018. Communicative Listener Feedback in Human–Agent Interaction: Artificial Speakers Need to Be Attentive and Adaptive . In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 9 pages.

1 INTRODUCTION

A central concern in conversational interaction is to make oneself understood and to understand what one's interlocutor means with an utterances. This is not simply a task that can be solved by pure 'natural language processing'. Understanding goes well beyond extracting literal meaning from a sentence. It very often requires cooperative interaction between interlocutors, which makes language use in conversation a 'collaborative effort' [17]. When producing an utterance, speakers need to design it in a way that makes it likely understandable to the listeners. Conversely, listeners need to reason about and infer what speakers have likely meant with their utterance. These processes can be thought of as coordination tasks and are facilitated by various resources, concrete ones such as the situation in which a conversation takes place, mental ones such as the 'common ground' [17] between dialogue partners, or interactional ones such as meta-communication. Giving artificial conversational agents the ability to use theses types of resources is a key challenge for enabling them to engage in communicative interactions that go beyond today's simple question answering or information presentation scenarios.

When problems in understanding occur and come to light, interaction partners can try to mitigate, or even solve, them, e.g., by making a new approach to the utterance that takes the updated information into account. One prevalent meta-communicative mechanism that is used in such situations is 'listener feedback' [4], which serves multiple functions from low-level coordination of the interlocutors' behaviours (e.g., the amount of information that the speaker provides), to higher-level coordination of beliefs and attitudes (e.g., what are the beliefs that can be considered shared, i.e., in the common ground) [26].

The present work deals with such feedback-based coordination between (embodied) artificial conversational agents and their human interaction partners. In previous work we presented computational models for an artificial conversational agent — described elsewhere [12, 14–16] — which has the capability to interpret the conversational feedback of its human interaction partners and to adapt its language production processes according to the attributed needs (e.g, there seems to be an understanding problem, so let's try again, this time providing some more information). In this paper, we describe an interaction study to investigate (1) whether human interlocutors are actually willing to provide natural communicative listener feedback to artificial conversational agents, and (2) whether human interlocutors notice that they are interacting with an (attentive speaker) agent that invests a certain collaborative effort in the joint project of making itself understood.

The paper is organised as follows: section 2 provides some background on communicative listener feedback and describes related work on feedback in human–agent interaction. Section 3 briefly describes the computational models underlying the attentive speaker agent, followed, in section 4, by a description of the study on interactions between the attentive speaker agent and human interlocutors.

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

2 BACKGROUND

2.1 Listener Feedback

Communicative feedback is an inconspicuous phenomenon that does not take centre-stage but is secluded in the background. Feedback takes place in the 'back channel' [47, p. 568], on 'track 2' [17, p. 241]. One of its defining features is that it does not adhere to — nor interferes with — the systematics of turn-taking. It does not occupy a turn, but may be placed with relatively few restrictions in parallel to an ongoing turn. To be unobtrusive, feedback signals are generally (i) short (i.e., they consist of minimal verbal/vocal expressions), (ii) locally adapted to their prosodic context (i.e., the speaker's utterance) by being more similar in pitch to their immediate surrounding than regular utterances [23], or (iii) taking place in the visual modality, for example as head gestures or facial expressions [2, 3].

Despite being such an inconspicuous phenomenon, communicative feedback is very expressive. Short communicative feedback expressions, such as *yeah*, *okay*, *hm*, can be recombined and reduplicated in many ways [1, 44], as well as changed continuously through prosodic modification [e.g., 21, 43]. In this way, feedback spans a large space of forms that map on an equally rich meaning space. In addition to this, non-verbal feedback signals are, depending on the modality, specialised to express their own meaning spaces (e.g., facial expressions are well suited to express attitudinal feedback).

From a cognitive point of view, communicative feedback has been framed as a mechanism to express (with different levels of awareness) the listening-related mental states (being in contact, willingness and ability to perceive, understand, accept, and agree) that an interlocutor experiences when listening to the dialogue partner that currently holds the speaking role [4, 27].

2.2 Feedback in human–agent interaction

There is an ongoing research effort to model feedback computationally for artificial conversational agents. Considerable effort has been spent on the question when it would be appropriate for an agent to produce feedback signals. This has been approached from multiple perspectives, by identifying prosodic cues in speakers' speech [45], by identifying multimodal cues of speakers [32], or triggered through problems when processing speakers' utterances [27]. Research has also tackled the question which specific feedback signal should be generated, e.g., based on the listener agent's mental and emotional state [7, 27], or justified by the reaction that specific feedback signals will likely evoke in the speaker [35].

Less work has approached the problem of modelling the processes within an artificial conversational agent in the role of the speaker. In an early, but already quite comprehensive, approach, Dohsaka and Shimazu [20] describe a dialogue model for incremental response generation to user utterances, including feedback, which, however, was only evaluated in simulation. Nakano and colleagues [34] present a model, informed by an analysis of a human–human interaction study, for estimating groundedness and choosing subsequent acts in a direction giving scenario based on multimodal user feedback. In a preliminary evaluation study some non-verbal but no verbal feedback was provided by users. Reidsma and colleagues [36] describe various efforts towards a conversational agent that is 'attentively speaking': it can perceive and classify feedback produced by the listener, elicit feedback from the listener, and flexibly adapt behaviours in response to feedback. Skantze and colleagues [39] focus on interpreting multimodal signals such as gaze, verbal expressions, but also timing and what these says about the listener's uncertainty, completion of the current activity, etc. Misu and colleagues [31] focusses on feedback elicitation cue generation and whether this evokes feedback responses from the agent's human interlocutors.

In previous work [13] we defined 'attentive speaker agents' to be conversational agents that should be able to

(1) interpret communicative listener feedback from their human interlocutors, taking the dialogue context into account, and
(2) adapt their ongoing and/or subsequent natural language utterances, paying heed to their interlocutors' needs — as inferred in (1).

The underlying property that makes such agents 'attentive', is that they need to be 'willing' to work towards being understood and to make extra efforts — if necessary — to achieve this. This desire plays a role in (1) and (2), and additionally requires attentive speaker agents to

(3) invite feedback from their interlocutors by providing opportunities or producing feedback elicitation cues when needed.

These three properties follow directly from psycholinguistic research on speaker-listener interaction in dialogue: (1) Communicative feedback serves as 'evidence of understanding' [17, 19] and information on further basic communicative functions [4, 27], see below. (2) Speakers monitor their interlocutors and immediately adapt to their feedback [5, 18, 28]. (3) speakers actively seek feedback from their interlocutors [6, 9]. All this shows that feedback provided by attentive listeners and responded to by attentive speakers goes well beyond simple backchannelling, which dominates research on feedback in artificial conversational agents.

3 MODELLING THE ATTENTIVE SPEAKER

The three properties of attentive speaking require dedicated computational models and an integration in an attentive speaker agent. Each of these models has been presented individually in previous work [12, 14–16]. Here, we briefly summarise them before turning to the study evaluation them in human–agent interaction.

3.1 Feedback interpretation as mental state attribution

The first requirement for attentive speaker agents is that they are able to interpret the feedback that their human interaction partners provide to them. In our model, we adopt Allwood, Kopp and colleagues' [4, 27] cognitive perspective on feedback, namely that feedback is an expression of the listener's listening-related mental state. Feedback can thus be framed as evidence of the listeners willingness and ability to understand (perceive, accept, and so forth; see section 2.1). Because of this, we model feedback interpretation as a mental state attribution process. The model assumes that a listener *L* whose current level of understanding is 'high' (for example), has a mental state that represents this phenomenal state in terms of a

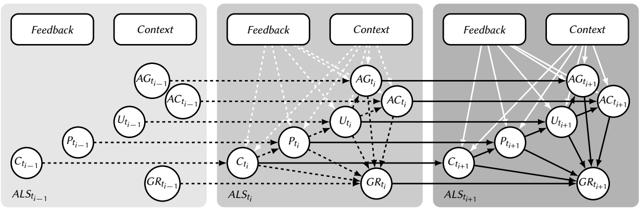

Figure 1: Illustration of a dynamic ALS two time-slice Bayesian network model unrolling over three steps in time, each corresponding to one dialogue segment. Dashed arrows are disregarded during inference in subsequent time-slices, i.e., variables from time slice $\mathrm{BN}_{\mathrm{ALS}_{t-1}}$ and evidence variables in time slice $\mathrm{BN}_{\mathrm{ALS}_t}$ have no influence on variables in time slice $\mathrm{BN}_{\mathrm{ALS}_{t+1}}$. The posterior distributions of attributed listener state variables (*C, P, U, AC, AG*) as well as the groundedness variable *GR* in time slice $\mathrm{BN}_{\mathrm{ALS}_t}$ are taken as prior distributions at time t_{i+1}.

belief, say $B_L(U = high)^1$. The mental state attribution process of the attentive speaker thus needs to be able to infer this belief of the listener and represent it as a belief itself, which is known as mentalising or theorising about the mind of the other [22]. As the mapping from mental state to feedback is not conventionalised, this attribution process involves uncertainty on the side of the attentive speaker. Hence the belief of a speaker S that the listener L believes that her level of understanding is *high*, becomes a matter of degree, which can be represented in terms of subjective probability (the confidence that the speaker has in this belief being true):

$$b_S(B_L(U = high)) = \mathrm{Pr}_{B_L(U)}(high)$$

Where $\mathrm{Pr}_{B_L(U)}$ is a probability distribution over the individual states that U may be in (in our model we use a ternary grading, namely *low*, *medium*, and *high*). It is important to note that this model of the listener's mental state implies that her belief to be in a specific level of understanding is certain and crisp. Uncertainty about the listener's belief is only modelled on the side of the speaker.

This allows for a representation of a single attributed mental state as a random variable (e.g., U). The full model for interpreting the listener's feedback thus needs to represent one random variable for each of the speaker's beliefs about the listener's various listening-related mental states (contact, perception, understanding, acceptance, and agreement). This belief state, the 'attributed listener state' (ALS) can be represented as a joint probability distribution $\mathrm{Pr}_{ALS}(C, P, U, AC, AG)$, but also can be represented, assuming conditional independence between some of these variables, in a more compact way using the representational and computational formalism of Bayesian networks [25]. This approach also allows us to model the hierarchical relationships between listening-related mental states that are described in the literature on the semantics and pragmatics of feedback [4, 10] and build on the general principle of 'upward completion' in language use [17, p. 147].

As listening-related mental states evolve over time and in parallel to the dialogue, the attribution process also needs to take temporal dynamics into account. We model this by using dynamic Bayesian networks [25, sec. 6.2.2]. Each time slice of the network corresponds to the specific increment in which the attentive speaker agent moves forward in the dialogue (in our case these are dialogue segments roughly the size of intonation units, which are produced by the incremental and adaptive natural language generation component of the speaker agent, see section 3.2).

Figure 1 shows how the attributed listener state model unrolls over three steps in time. At each time step t, properties of the human interlocutor's feedback signals as well as information about the immediate dialogue context (e.g., how difficult is the utterance that the speaker produces) are provided as evidence to the dynamic Bayesian network, which is then used to infer the current belief state of the attentive speaker agent regarding the interlocutor's listening-related mental states.

3.2 Adaptive behaviour generation

Our second requirement for attentive speaker agents is that they can adapt their behaviour, especially their language production, in such a way that they take the listener's needs into account. The model for feedback interpretation described in the previous section infers and represents an up-to-date mental state of the listener at any point in the dialogue. Our model for adaptive behaviour generation makes use of this information for incrementally adapting the speaker agent's language generation and dialogue management decisions.

The natural language generation component we use is an incremental and adaptive variant of the SPUD microplanner [40, 41]. How it generates natural language incrementally is described in [12], here it suffices to say that we use it to incrementally produce descriptions of calendar operations (announcement of new appointment, changes to existing appointments due to conflicts, etc.) and that the generation mechanisms takes the most recent attributed listener state into account each time it generates an increment of a description. This allows for adaptation to take place while utterances are ongoing. If, for example, a degradation of the listener's understanding is indicated through feedback, the natural language generation component can take this information from the attributed listener state into account in its next increment, e.g, by inserting additional or redundant information, or by repeating the increment that may have led to the problem in understanding. The information in the attributed listener state is used to parametrise the central decision-making function in the generation component as in [29].

On the level of dialogue management, adaptation also happens in the decision mechanism, e.g., after the presentation of an information presentation unit such as a calender item, the attributed listener state is evaluated whether the listener's understanding is sufficient for current purposes or whether actions are necessary to increase the listener's level of understanding (e.g., by repeating the complete information presentation unit).

3.3 Elicitation cue generation

The final requirement for attentive speaker agents is that the agent should actively seek feedback from listeners if they do not provide feedback pro-actively. This is modelled via an assessment of the information needs of the agent and realised with a simple threshold model that evaluates the attributed listener state (e.g., has the mental

[1] B is the belief operator and $B_L(x)$ means L beliefs x

state of understanding currently attributed to the listener been static for a specific amount of time and is the listener's understanding not as high as it should ideally be?). If such an information need is observed, the agent produces multimodal feedback elicitation cues by inserting pauses, gazing at the listener, and may even provide a verbal elicitation cue such as *okay?*. These cues are generated and realised incrementally as well and can, in principle, be inserted at each boundary of the increments generated by the the adaptive incremental natural language generation component.

3.4 Integration in an attentive speaker agent

The three models described above have been integrated in an attentive embodied speaker agent that can present appointments and calendar operations to its human interlocutors (see [11, chp. 8] for details). The agent is currently operated in a semi-autonomous manner: signal properties of the interlocutor's feedback behaviour (function, polarity, modality, gaze) are entered into the system in real-time by a human Wizard-of-Oz that observes the system's human interlocutor.

This information is then used by the three computational models to autonomously interpret the listener's feedback given the dialogue context, adapt its incrementally generated ongoing verbal and nonverbal behaviour, and generate feedback elicitation cues if needed. All information in the system is processed incrementally, based on the principles of the incremental unit framework [38].

4 INTERACTING WITH THE ATTENTIVE SPEAKER

The artificial attentive speaker agent described above was used to investigate the research questions we pursue here: (1) are human interlocutors willing to provide natural communicative listener feedback to artificial conversational agents, and (2) do human interlocutors notice that they are interacting with an agent that makes a collaborative effort in the joint project of making itself understood?

We carried out an interaction study in which the attentive speaker agent described calender-related information to its human interlocutors. Participants had the task to understand this information as well as possible so that they would be able to recall it afterwards.

Experimental conditions. In order to be able to compare the effects of the attentive speaker agent, the interaction study consisted of three experimental conditions. The target condition

AS In the 'attentive speaking' condition participants interacted with the attentive speaker agent, which perceived participants' feedback (via the wizard, see section 3.4), probabilistically attributed listening-related mental states based on this information, adapted its speech production on the levels of natural language generation and dialogue management, and elicited feedback using verbal and nonverbal cues — if it had information needs. The dialogue strategy used in condition AS is illustrated in fig. 2 (left): immediately after a complete information presentation unit is incrementally generated and realised (U), the attributed listener state is evaluated (E). Based on this evaluation, the agent decided to either continue with the next utterance (if a participant's understanding is estimated to be sufficiently high; C), repeat the current utterance (if a participant's understanding is estimated to be insufficient; R), or, if the evaluation result did not allow for a clear decision, ask the participant whether it should either continue or repeat (A). Participants then had to answer this question for the interaction to proceed.

was compared with two control conditions that served as baselines:

EA In baseline condition 'explicit asking' participants interacted with a speaker agent that did not process or consider any feedback signals, but followed a fixed dialogue strategy according to which the agent explicitly asked, after the realisation of each complete information presentation unit, whether it should either continue with the next one, or repeat the current one (see fig. 2, centre). Participants then had to answer this question for the interaction to proceed.

NA In baseline condition 'no-adaptation' participants interacted with a speaker agent that did not process or consider any feedback signals and followed a fixed dialogue strategy according to which the agent immediately continued with the next information presentation unit after the current utterance was realised (see fig. 2, right).

Conditions EA and NA were created to test agents with different levels of adaptivity to the interlocutor. Furthermore they served as upper- and lower-bound baselines on participants' understanding of the information that the agents presented and on standard dialogue evaluation measures (such as PARADISE; [42]). As participants in condition EA were always asked by the agent how to continue, they could freely choose to have items repeated as often as they like. We expected dialogues with these agents to be long, but information transfer to be effective. Participants in condition NA on the other hand, were never asked whether they would like to hear a repetition of a unit. We expected dialogues with these agents to be short, but information transfer to be ineffective. The target condition AS should lie right in the middle of these two conditions, as the agent only repeated information if it inferred this to be necessary, or if it was uncertain and the participant wanted to hear a repetition. Interaction with the target agent should be shorter in duration than with the agent in condition EA, but more effective than with the agent in condition NA. We do not consider measures of effectiveness and costs in the context of this paper, though.

The study followed a between-subject design. Each participant interacted with one of the embodied conversational agents in an information presentation task in a calender assistant domain. 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. Importantly, participants across all conditions were told that they cannot speak freely with the agent (for technical reasons; the agent did not respond to such utterances), but may provide multimodal listener feedback, which the agent might take into account in its own behaviour. Information was presented in six dialogue phases (each consisting of two to three information presentation units), which were followed by a recall phase in which calender items needed to be written down in a paper calendar template.

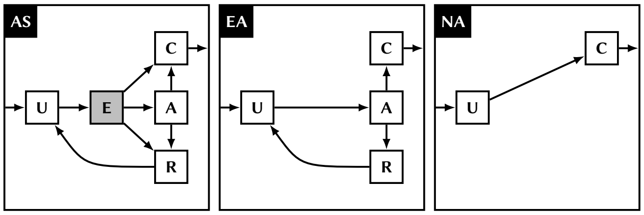

Figure 2: Dialogue phases in the study consisted of two to three information presentation units, the structure of which differed depending on the experimental condition (AS/EA/NA). Nodes represent the following actions: U – present information in an incremental utterance; E – evaluate current attributed 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.

Participants. Thirty-six participants, twelve per experimental condition, were recruited from Bielefeld's student population and compensated with 5 euro. Participants were between 18 and 40 years old ($M = 24.2, SD = 4.4$); 25 reported to be female, 11 to be male; 33 reported to be native speakers of German (the others reported a mean of 13.8 years of experience speaking German); 18 of the participants reported prior experience interacting with virtual agents or humanoid robots; four participants reported non-normal and non-corrected vision; one participant reported non-normal and non-corrected hearing. Participants were blindly assigned to experimental conditions. An analysis of the distribution of participant features across conditions yielded that an influence of gender and vision on the outcome of our study could not be ruled out completely.[2]

Supplementary Material. Research data and (statistical) analyses in Python/R are archived and available as a citable data publication at https://doi.org/10.6084/m9.figshare.6171260.

4.1 Participants' feedback behaviour

The first question we addressed is whether participants actually provided feedback to the conversational agents that they interacted with. We annotated participants' feedback behaviour based on the webcam-videos, which gave annotators the same information that the wizard had. Using the audio-tracks only, participants' utterances were segmented and those with feedback-characteristics that were not produced in response to questions of the agent were classified as verbal/vocal feedback and transcribed, using Praat [8], based on the conventions used in the ALICO-corpus [30]. Similarly, using the video-tracks only, participants' head gesture feedback was segmented and labelled using ELAN [46], head gestures produced while responding to questions of the agent were filtered out afterwards. Head gesture unit labels were limited to the head movement types nod, shake, jerk, tilt, turn, protrusion, and retraction [30]. Thirty-three interactions (of 36; audio-visual data was missing for three interactions) were segmented, transcribed, annotated, and analysed. Transcriptions and annotations were checked and corrected through systematic listening.

In total, 734 feedback signals were encountered. 127 (17.3%) of these were unimodal verbal 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 were produced in overlap. This showed that participants of the interaction study produced feedback signals in response to the communicative actions of the artificial conversational agents, a result that was found previously (e.g, in [36], see section 2.2).

Next, we looked at these feedback signals in a slightly more detailed way — looking at classes of feedback form and their distribution — and investigated whether the feedback behaviour of participants was similar to the feedback behaviour that humans produce in natural human–human dialogue.

Of the 436 verbal/vocal feedback signals produced, the most frequently used feedback expression was *okay* (41.5 %), followed by *mhm* (18.3 %), *ja* (14.2 %), *m* (6 %), *nein* (3.2 %), *hm* (1.8 %), *ja okay* (1.6 %), and *nee* (1.4 %). These were followed by a 'long tail' of expressions (each < 1 %). This distribution resembles the one we found in human–human conversations in the ALICO-corpus [30, tbl. 7], where the four most frequent feedback expressions are *ja*, *m*, *mhm*, and *okay*, too. Of the 598 head gestures produced as feedback, 81.6 % were labelled nod, 8.9 % tilt, 6.4 % shake, and 3.2 % jerk. The distribution of head gesture types is also similar to human–human conversation [30, tbl. 4]. Both distributions are 'nod-heavy' and in both cases the four most frequent units are nod, jerk, tilt, and shake. The 311 bimodal feedback signals participants produced were coherent across modalities. Almost all head gestures of the category nod occurred together with verbal feedback expressions of positive polarity (*okay*, *mhm*, *ja* and *m*). Head gestures of the type shake occurred together with verbal feedback expression of negative polarity (*nein* and variants such as *hm nein*). Head gesture types tilt and jerk rarely occurred in overlap with verbal/vocal feedback. We conclude that feedback is comparable in form and distribution to the feedback that humans use in human–human dialogue.

Participant across all three conditions were told in the instructions that they can provide multimodal communicative listener feedback, only in the target condition AS, however, did the agent attend to the feedback, adapt its own behaviour, and elicited feedback from its interlocutors. Because of this, we hypothesised that participants' feedback behaviour differed between experimental conditions, more specifically, we expected that participants provide more feedback in the target condition AS than in the two control condition EA and NA.

To test this hypothesis, we analysed participants' feedback frequency, which, in contrast to the absolute number of feedback signals, is a measure that is comparable across conditions. We defined feedback rate to be the number of feedback signals per presentation (which varies with the number of repetitions).

Across all conditions, participants produced between 0 and 2.4 feedback signals per presentation unit, with a mean feedback rate

[2] In both cases contingency table Bayes factor tests (with prior concentration set to $a = 1$) only find 'anecdotal' evidence in favour of the null-hypothesis of independence of participant feature from experimental condition ($\text{BF}_{01} = 2.557$ for gender; $\text{BF}_{01} = 2.661$ for vision). ¶ The interpretation of the strength of evidence for a hypothesis A in comparison to a hypothesis B (the Bayes factor BF_{AB}) follows Jeffreys [24, p. 432] and is stated in single quotes throughout this paper. ¶ All Bayesian statistics was computed using the 'BayesFactor' R package [33].

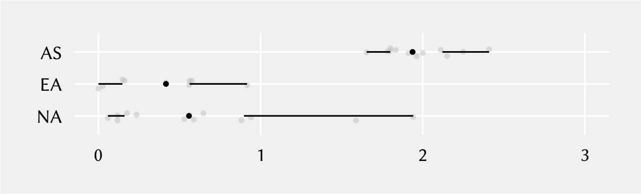

Figure 3: Distribution of participants' feedback rate (number of feedback signals per presentation; see text) by experimental condition. Data points are y-jittered in translucent light grey; black dots are medians, black lines are whiskers, mid gaps are quartiles.

Table 1: Bayes factor analyses of feedback rate (feedback signals per presentation). Bayes factor t-tests of conditions AS : EA and AS : NA analyse both alternative hypotheses against the null hypothesis, and against each other. The test EA : NA is two-sided and two sample.

Comparison	Bayes factor t-tests		
	$\mathrm{BF}_{>0}$	$\mathrm{BF}_{0<}$	$\mathrm{BF}_{><}$
AS : EA	3.934e8	225.395	8.868e10
AS : NA	5.148e4	100	5.016e6
EA : NA	$\mathrm{BF}_{01} = 1.467$		

of $M = 1.1$ ($SD = 0.8$). Analysing participants' feedback behaviour by experimental condition we found differences in feedback rate. Participants in condition AS had a mean feedback rate of $M = 1.97$ ($Mdn = 1.93, SD = 0.22, \min = 1.65, \max = 2.4$). Participants in condition NA followed with a mean feedback rate of $M = 0.65$ ($Mdn = 0.56, SD = 0.6, \min = 0.06, \max = 1.94$). Participants in condition EA only had a mean feedback rate of $M = 0.1$ ($Mdn = 0.41, SD = 0.31, \min = 0, \max = 0.92$). Figure 3 shows the distribution of feedback rate by condition.

To confirm our hypothesis that participants provided more feedback in the target condition AS, we analysed feedback rates in a Bayesian framework[3]. A Bayesian ANOVA yielded the Bayes factor $\mathrm{BF}_{10} = 1.042e7$, 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. We further analysed this omnibus result with post-hoc tests. Firstly, we analysed our hypothesis that participants in target condition AS produced more feedback per presentation than participants in condition EA, i.e, AS > EA. We did this using a BayesFactor two sample t-test[4]. For the one-sided alternative hypothesis of a positive effect, i.e., that participants in target condition AS produced more feedback signals per presentation than participants in control condition EA, this yielded the Bayes factor $\mathrm{BF}_{>0} = 3.934e8$, which is considered 'decisive' evidence against the null hypothesis that there were no differences. The complementary alternative hypothesis of a negative effect yielded the Bayes factor $\mathrm{BF}_{<0} = 0.004$, which is considered 'decisive' evidence in favour of the null hypothesis. Directly comparing the two alternative hypotheses yielded the Bayes factor $\mathrm{BF}_{><} = 8.868e10$, which is 'decisive' evidence in favour of our hypothesis that participants in condition AS produced more feedback per presentation than participants in condition EA. Secondly, we analysed the hypothesis that participants in the target condition AS produced more feedback per presentation than participants in control condition NA, i.e, AS > NA, using the same approach. The results parallel the ones above, the details are shown in table 1. Finally, we analysed the experimental condition EA against condition NA. Here our hypothesis was that there should be no difference in feedback rate since the agents in both conditions ignored the participants' feedback signals. The analysis yielded the Bayes factor $\mathrm{BF}_{01} = 1.467$, which can be considered 'anecdotal' evidence for the null hypothesis that there is no difference in feedback rate. Such weak evidence, however, suggests that we do not have enough data to make a definite statement.

Both analyses show that participants who interacted with the attentive speaker agent (in the target condition AS) clearly produced more feedback signals per presentation than participants that interacted with the agents that were ignorant of participants' feedback (in the control conditions EA and NA). No difference in feedback rate between these latter two conditions were found. We can conclude from this that the attentive speaker agent's capabilities and behaviour had a decisive effect on the rate of communicative listener feedback being provided.

Analysing participants' feedback, we further found that their feedback rates did not vary much over time. In condition AS, the standard deviation of the mean feedback rate across dialogue phases and participants was $SD = 0.23$ and it was 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, probably 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.

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 rates in conditions EA and NA. The difference in feedback rate between condition AS and the control conditions thus cannot simply be due to the feedback elicitation cues that the attentive speaker agent produced.

[3] Carrying out the analysis using classic null-hypothesis significant testing yielded similar results.

[4] Using the default, 'medium'-scaled prior distribution ($r = \sqrt{2}/2$), for each one-sided alternative hypothesis (positive/negative effect) against the null hypothesis (no effect), and then against each other.

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

4.2 Subjective perception of the agents

We now turn to subjective factors that reflect the participants' perception of the agent. We measured these factors with a questionnaire that immediately followed the interaction study and asked participants to report their subjective experience of the interaction. In total, twenty items were presented in random order and had to be rated on seven-point Likert scales[5]. We looked at participants' responses to a selection of six of the twenty items that were of particular interest, namely those that focussed on the agents' perceived feedback processing capabilities, whether the agents were perceived to be attentive and adaptive to participants' feedback and needs, and whether the agents were perceived to be 'helpful' [37, p. 58] in resolving difficulties in understanding.

In contrast to usual analyses of questionnaires of Likert scale items, we compared ratings of items individually, i.e., not grouped into higher level factors, since the questionnaire was not developed with the intent of items to be grouped. This made it difficult to do an inferential analysis because type I errors due to multiple testing became very likely. We therefore carried out a descriptive analysis in which, for each item, we compared the median rating of the experimental conditions. We substantiated our arguments by making estimations of how likely, given the data, the observed ordering of a rating of an item is in relation to its alternatives. We did this — similar to the the Bayesian analysis in the previous section — by first computing the Bayes factor t-test[6] for each one-sided alternative hypothesis (positive/negative effect) against the null hypothesis (no effect), and then against each other. The relevant Bayes factor values, where evidence is at least 'substantial', are shown in fig. 4.

In general we expected that participants in the target condition AS would provide higher ratings than participants in both control conditions EA and NA. We also expected that participants who interacted with the agent that explicitly asked whether it should repeat or continue (EA) would be rated higher than the agent that did not adapt at all (NA) — at least for some of the questionnaire items.

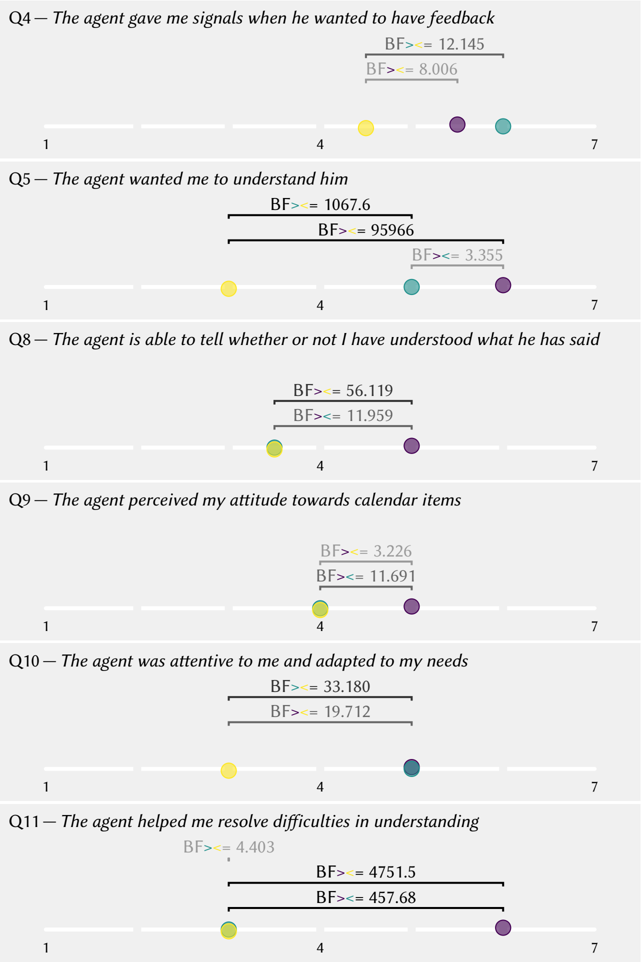

Figure 4: Median ratings and Bayes factor-based comparison of selected questionnaire items across experimental conditions (● AS; ● EA; ● NA). Brackets over two median dots show the Bayes factor t-test value comparing both one-sided alternative hypotheses (positive/negative effect) against each other. Colour-coded angle brackets indicate ordering of conditions (given, e.g., BF$_{><}$, a value > 0 is evidence in favour of the ordering AS $>$ EA, a value < 0 would be evidence in favour of the inverse ordering AS $<$ EA). Intensity encodes strength of evidence as follows: 'substantial' — 'strong' — 'very strong' — 'decisive'. Brackets for evidence considered merely 'anecdotal' are omitted.

We started by looking at the item *The agent gave me signals when he wanted to have feedback* (Q4). The analysis found 'substantial', respectively 'strong', evidence that participants in condition AS and EA were more convinced that the agent provided signals in order to elicit feedback than participants in condition NA, who where rather uncertain about this. The fact that target condition AS and control condition EA were rated similarly (there was not enough evidence to draw conclusions from the slight difference in median ratings) suggests that participants not only took the AS-agent's

[5] Labels for the response anchors were: 1 — strongly disagree, 2 — disagree, 3 — somewhat disagree, 4 — neither agree nor disagree, 5 — somewhat agree, 6 — agree, and 7 — strongly agree.

[6] We used the default, 'medium'-scaled prior distribution ($r = \sqrt{2}/2$). ¶ We merely used the tests as a tool for weighing the evidence for specific orderings of experimental conditions.

elicitation cues as requests for feedback, but also interpreted the EA-agent's explicit request (whether to continue or repeat) after the presentation of each unit as a feedback request.

Interestingly, this similarity between conditions AS and EA disappeared in the item *The agent wanted me to understand him* (Q5). Although the analysis still yielded 'decisive' evidence that participants in both conditions AS and EA attributed to the agent a desire to be understood that is higher than in condition NA, there was 'substantial' evidence that participants in condition AS had an even bigger sense of this than participants in condition EA. Although both were perceived as being similarly interested in receiving feedback (Q4), participants that interacted with the AS-agent had a stronger impression that the agent had a desire to be understood.

Next we looked at the agents' perceived abilities to interpret participants' feedback behaviour. The analysis of item *The agent is able to tell whether or not I have understood what he has said* (Q8) yielded 'strong', respectively 'very strong' evidence that participants in the attentive speaking condition AS were convinced to a higher degree that the agent was able to tell whether they understood (or not) than participants in the two control conditions EA and NA (which did not differ). Similar, but a little less pronounced, results were obtained for the item *The agent perceived my attitude towards calendar items* (Q9). There was 'strong', respectively 'substantial', evidence that participants in the attentive speaking condition AS rated the agent's ability to perceive their attitude higher than participants in the control conditions EA and NA. Again, the Bayes factor analysis found no evidence for a difference between the two control conditions.

Turning to the agents' abilities to adapt to participants' needs, results had a similar relationship as in Q4 and Q5. For the questionnaire item *The agent was attentive to me and adapted to my needs* (Q10) there was 'strong', respectively 'very strong', evidence that participants in conditions AS and EA felt more strongly that the agent they interacted with was attentive and adaptive to their needs than participants in condition NA. Again, as in Q4, there was not enough evidence for a difference between the rating of the two conditions. A plausible explanation may, again, be that even the EA-agent adapted (by continuing or repeating as requested).

In contrast to this, however, clear differences in participants' ratings could be observed for the item *The agent helped me resolve difficulties in understanding* (Q11). Here, there was 'decisive' evidence that participants in the attentive speaking condition AS felt to a larger degree that the agent helped them resolve difficulties in understanding as compared to the two control conditions. In addition, there was substantial evidence that the agent in control condition EA was rated higher than the agent condition NA.

In conclusion, it can be said that, unsurprisingly, the attentive speaker agent clearly received more favourable ratings than the agent in control condition NA, which neither attended to its human interlocutors' feedback, nor adapted to their needs at all. The comparison of the attentive speaker agent to the agent in the control condition EA, in which the agent explicitly asked its human interlocutors how to proceed (continue or repeat), is more complex. Interlocutors of the EA-agent noticed that it wanted to get 'feedback' from them (Q4) and 'adapted' accordingly (Q10) — here the participants' subjective perception of the agent did not differ from the attentive speaking agent. The similar, but more specific, items Q5 and Q11, however were evaluated differently by participants. There was 'substantial' evidence that participants more strongly attested the attentive speaker agent a desire to be understood (Q5) and there was even 'decisive' evidence that the participants more strongly felt that the attentive speaker agent helped them resolve difficulties in understanding (Q11). In accordance with these findings, participants that interacted with the attentive speaker agent clearly noticed that this agent had the ability to interpret their feedback (Q8 and Q9), which was felt to a lesser degree by participants that interacted with the EA-agent.

5 CONCLUSION

In this paper we framed communicative listener feedback as an important meta-communicative coordination mechanism for reaching joint understanding in conversational interaction. We argued that artificial conversational agents should have the capability to use such a mechanism, too, because it would allow them to approach potential or upcoming problems in understanding (and other listening-related communicative functions) before they become more serious and require costly repair actions. Adapting one's own language production to the needs of the interlocutor, as indicated through communicative feedback, is one way of contributing to this effort.

The specific contribution in this paper is a comprehensive evaluation study of such an attentive speaker agent, which rests on models for eliciting, interpreting, and adapting quickly to feedback. In this interaction study we compared three differently attentive speaker agents. We first analysed whether human interlocutors were willing to provide feedback to artificial agents in general and found that they only did so if the agent was actually attentive to their feedback and responds to it by adapting its behaviour. Following this we investigated whether participants subjectively perceived the agents to be different and whether they observed the attentiveness and adaptivity of the attentive speaker agent and were aware of the collaborative effort that it made to the interaction. Here participants noticed that both the attentive speaker agent as well as one control agent were attentive and adaptive, but only the attentive speaker agent was perceived as having a desire to make itself understood as well as being helpful in resolving difficulties in participants' understanding.

In conclusion, we can say that in order to receive listener feedback from their human interlocutors and to be perceived as an attentive speaker agent, artificial conversational agents need to actually be attentive and adaptive to their interlocutors' feedback and needs.

ACKNOWLEDGMENTS

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).

REFERENCES

[1] Jens Allwood. 1988. Om det svenska systemet för språklig återkoppling [On the Swedish system of linguistic feedback]. In *Svenskans Beskrivning 16*, Per Linell et al. (Eds.). Vol. 1. Linköping University, Tema Kommunikation, Linköping, Sweden, 89–106.

[2] Jens Allwood and Loredana Cerrato. 2003. A study of gestural feedback expressions. In *Proceedings of the 1st Nordic Symposium on Multimodal Communication*. Copenhagen, Denmark, 7–22.

[3] Jens Allwood, Stefan Kopp, Karl Grammer, Elisabeh Ahlsén, Elisabeth Oberzaucher, and Markus Koppensteiner. 2007. The analysis of embodied communicative feedback in multimodal corpora: A prerequisite for behaviour simulation. *Language Resources and Evaluation* 41 (2007), 255–272. https://doi.org/10.1007/s10579-007-9056-2

[4] Jens Allwood, Joakim Nivre, and Elisabeth Ahlsén. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics* 9 (1992), 1–26. https://doi.org/10.1093/jos/9.1.1

[5] Janet B. Bavelas, Linda Coates, and Trudy Johnson. 2000. Listeners as co-narrators. *Journal of Personality and Social Psychology* 79 (2000), 941–952. https://doi.org/10.1037/0022-3514.79.6.941

[6] Janet B. Bavelas, Linda Coates, and Trudy Johnson. 2002. Listener responses as a collaborative process: The role of gaze. *Journal of Communication* 52 (2002), 566–580. https://doi.org/10.1111/j.1460-2466.2002.tb02562.x

[7] Elisabetta Bevacqua, Sathish Pammi, Sylwia Julia Hyniewska, Marc Schröder, and Catherine Pelachaud. 2010. Multimodal backchannels for embodied conversational agents. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents*. Philadelphia, PA, USA, 194–201. https://doi.org/10.1007/978-3-642-15892-6_21

[8] Paul Boersma and David Weenink. 2016. Praat, a system for doing phonetics by computer. (2016). http://www.praat.org/

[9] Susan E. Brennan. 1990. *Seeking and Providing Evidence for Mutual Understanding*. Ph.D. Dissertation. Stanford University, Stanford, CA, USA.

[10] Harry Bunt. 2011. Multifunctionality in dialogue. *Computer Speech and Language* 25 (2011), 222–245. https://doi.org/10.1016/j.csl.2010.04.006

[11] Hendrik Buschmeier. 2018. *Attentive Speaking. From Listener Feedback to Interactive Adaptation.* Ph.D. Dissertation. Faculty of Technology, Bielefeld University, Bielefeld, Germany.

[12] Hendrik Buschmeier, Timo Baumann, Benjamin Dosch, Stefan Kopp, and David Schlangen. 2012. Combining incremental language generation and incremental speech synthesis for adaptive information presentation. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Seoul, South Korea, 295–303.

[13] Hendrik Buschmeier and Stefan Kopp. 2011. Towards conversational agents that attend to and adapt to communicative user feedback. In *Proceedings of the 11th International Conference on Intelligent Virtual Agents*. Reykjavík, Iceland, 169–182. https://doi.org/10.1007/978-3-642-23974-8_19

[14] Hendrik Buschmeier and Stefan Kopp. 2012. Using a Bayesian model of the listener to unveil the dialogue information state. In *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue (SemDial)*. Paris, France, 12–20.

[15] Hendrik Buschmeier and Stefan Kopp. 2014. A dynamic minimal model of the listener for feedback-based dialogue coordination. In *Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue (SemDial)*. Edinburgh, UK, 17–25.

[16] Hendrik Buschmeier and Stefan Kopp. 2014. When to elicit feedback in dialogue: Towards a model based on the information needs of speakers. In *Proceedings of the 14th International Conference on Intelligent Virtual Agents (IVA)*. Boston, MA, USA, 71–80. https://doi.org/10.1007/978-3-319-09767-1_10

[17] Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge, UK. https://doi.org/10.1017/CBO9780511620539

[18] Herbert H. Clark and Meredyth A. Krych. 2004. Speaking while monitoring addressees for understanding. *Journal of Memory and Language* 50 (2004), 62–81. https://doi.org/10.1016/j.jml.2003.08.004

[19] Herbert H. Clark and Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive Science* 13 (1989), 259–294. https://doi.org/10.1207/s15516709cog1302_7

[20] Kohji Dohsaka and Akira Shimazu. 1997. A system architecture for spoken utterance production in collaborative dialogue. In *Working Notes of the IJCAI-97 Workshop on Collaboration, Cooperation and Conflict in Dialogue Systems*. Nagoya, Japan.

[21] Konrad Ehlich. 1986. *Interjektionen*. Max Niemeyer Verlag, Tübingen, Germany. https://doi.org/10.1515/9783111357133

[22] Chris D. Frith and Uta Frith. 2006. The neural basis of mentalizing. *Neuron* 50 (2006), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

[23] Mattias Heldner, Jens Edlund, and Julia Hirschberg. 2010. Pitch similarity in the vicinity of backchannels. In *Proceedings of INTERSPEECH 2010*. Makuhari, Japan, 3054–3057.

[24] Harold Jeffreys. 1961. *Theory of Probability* (3rd ed.). Clarendon Press, Oxford, UK.

[25] Daphne Koller and Nir Friedman. 2009. *Probabilistic Graphical Models. Principles and Techniques.* The MIT Press, Cambridge, MA, USA.

[26] Stefan Kopp. 2010. Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. *Speech Communication* 52 (2010), 587–597. https://doi.org/10.1016/j.specom.2010.02.007

[27] Stefan Kopp, Jens Allwood, Karl Grammar, Elisabeth Ahlsén, and Thorsten Stocksmeier. 2008. Modeling embodied feedback with virtual humans. In *Modeling Communication with Robots and Virtual Humans*, Ipke Wachsmuth and Günther Knoblich (Eds.). Springer-Verlag, Berlin, Germany, 18–37. https://doi.org/10.1007/978-3-540-79037-2_2

[28] Robert M. Krauss and Sidney Weinheimer. 1966. Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology* 4 (1966), 343–346. https://doi.org/10.1037/h0023705

[29] François Mairesse and Marylin A. Walker. 2010. Towards personality-based user adaptation: Psychologically informed stylistic language generation. *User Modeling and User-Adapted Interaction* 20 (2010), 227–278. https://doi.org/10.1007/s11257-010-9076-2

[30] Zofia Malisz, Marcin Włodarczak, Hendrik Buschmeier, Joanna Skubisz, Stefan Kopp, and Petra Wagner. 2016. The ALICO corpus: Analysing the active listener. *Language Resources and Evaluation* 50 (2016), 411–442. https://doi.org/10.1007/s10579-016-9355-6

[31] Teruhisa Misu, Etsuo Mizukami, Yoshinori Shiga, Shinichi Kawamoto, Hisashi Kawai, and Satoshi Nakamura. 2011. Toward construction of spoken dialogue system that evokes users' spontaneous backchannels. In *Proceedings of the 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Portland, OR, USA, 259–265.

[32] Louis-Philippe Morency, Iwan de Kok, and Jonathan Gratch. 2010. A probabilistic multimodal approach for predicting listener backchannels. *Autonomous Agents and Multiagent Systems* 20 (2010), 70–84. https://doi.org/10.1007/s10458-009-9092-y

[33] Richard D. Morey and Jeffrey N. Rouder. 2015. BayesFactor: Computation of Bayes factors for common designs. (2015). https://CRAN.R-project.org/package=BayesFactor

[34] Yukiko I. Nakano, Gabe Reinstein, Tom Stocky, and Justine Cassell. 2003. Towards a model of face-to-face grounding. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*. Sapporo, Japan, 553–561. https://doi.org/10.3115/1075096.1075166

[35] Catharine Oertel, José Lopes, Yu Yu, Kenneth A. Funes Mora, Joakim Gustafson, Alan W. Black, and Jean-Marc Odobez. 2016. Towards building an attentive artificial listener: On the perception of attentiveness in audio-visual feedback tokens. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. Tokyo, Japan, 21–28. https://doi.org/10.1145/2993148.2993188

[36] Dennis Reidsma, Iwan de Kok, Daniel Neiberg, Sathish Pammi, Bart van Straalen, Khiet Truong, and Herwin van Welbergen. 2011. Continuous interaction with a virtual human. *Journal on Multimodal User Interfaces* 4 (2011), 97–118. https://doi.org/10.1007/s12193-011-0060-x

[37] Zsófia Ruttkay, Claire Dormann, and Han Noot. 2004. Embodied conversational agents on a common ground. A framework for design and evaluation. In *From Brows to Trust. Evaluating Embodied Conversational Agents*, Zsófia M. Ruttkay and Catherine Pelachaud (Eds.). Kluwer Academic Publishers, Dordrecht, The Netherlands, 27–66. https://doi.org/10.1007/1-4020-2730-3_2

[38] David Schlangen and Gabriel Skantze. 2011. A general, abstract model of incremental dialogue processing. *Dialogue and Discourse* 2 (2011), 83–111. https://doi.org/10.5087/dad.2011.105

[39] Gabriel Skantze, Anna Hjalmarsson, and Catharine Oertel. 2014. Turn-taking, feedback and joint attention in situated human–robot interaction. *Speech Communication* 65 (2014), 50–66. https://doi.org/10.1016/j.specom.2014.05.005

[40] Matthew Stone. 2002. Lexicalized grammar 101. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*. Philadelphia, PA, USA, 77–84.

[41] Matthew Stone, Christine Doran, Bonnie Webber, Tonia Bleam, and Martha Palmer. 2003. Microplanning with communicative intentions: The SPUD system. *Computational Intelligence* 19 (2003), 311–381. https://doi.org/10.1046/j.0824-7935.2003.00221.x

[42] Marylin A. Walker, Diane J. Litman, Candace A. Kamm, and Alicia Abella. 1998. Evaluating spoken dialogue agents with PARADISE: Two case studies. *Computer Speech and Language* 12 (1998), 317–347. https://doi.org/10.1006/csla.1998.0110

[43] Nigel Ward. 2004. Pragmatic functions of prosodic features in non-lexical utterances. In *Proceedings of the International Conference on Speech Prosody*. Nara, Japan, 325–328.

[44] Nigel Ward. 2006. Non-lexical conversational sounds in American English. *Pragmatics & Cognition* 14 (2006), 129–182. https://doi.org/10.1075/pc.14.1.08war

[45] Nigel Ward and Wataru Tsukahara. 2000. Prosodic features which cue backchannel responses in English and Japanese. *Journal of Pragmatics* 38 (2000), 1177–1207. https://doi.org/10.1016/S0378-2166(99)00109-5

[46] Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. In *Proceedings the 5th International Conference on Language Resources and Evaluation*. Genoa, Italy, 1556–1559.

[47] Victor H. Yngve. 1970. On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, Mary Ann Campbell et al. (Eds.). Chicago Linguistic Society, Chicago, IL, USA, 567–577.