Creating familiarity through adaptive behavior generation in human-agent interaction

Ramin Yaghoubzadeh and Stefan Kopp

Sociable Agents Group, CITEC, Bielefeld University
P.O. Box 10 01 31, 33501 Bielefeld, Germany
{ryaghoub,skopp}@techfak.uni-bielefeld.de

Abstract. Embodied conversational agents should make use of an adaptive behavior generation mechanism which is able to gradually refine its repertoire to behaviors the individual user understands and accepts. We present a probabilistic model that takes into account possible *socio-communicative* effects of utterances while selecting the behavioral form.

Keywords: Behavior generation, Familiarity, Addressee design, Personalized communication, Adaptivity, Social intentions

1 Introduction and motivation

Suppose you were to ask a good friend, a colleague in your building, or an unknown distinguished older man to give you the time. How would you formulate your request? Possible ways are “could you please tell me the time, sir?”, “what time is it, please?”, “what’s the time?”. These options differ with regard to their social acceptability and amount of elicited face threat [4] – depending on the context, the addressee, and the personal common ground the two of you have. Human speakers take these contingencies into account (e.g., as audience design [3]) and expect others to do the same. However, it is not so clear which changes in the way things are preferably expressed are licensed at what time of an ongoing familiarization process between interactants. Many coordination effects can be found at different levels of verbal and nonverbal behavior, possibly serving cognitive, communicative, as well as social functions (see [7] for an overview). Pickering and Garrod [10] argued for automatic alignment in communication as the result of priming processes on all levels of linguistic processing, both intrapersonally and interpersonally. Deliberately taking the perspective of the interlocutor, based on rough assumptions about their needs and knowledge, can additionally change the used communicational repertoire early on, by estimating only a few bits of information about the other [3]. Continued and repeated interactions furthermore change the social distance and relationship between interlocutors, resulting in higher familiarity and leading to adaptations of the production repertoire [6].

We investigate if and how these phenomena are also expected from, and can be modeled for, embodied conversational agents. ECAs, or ‘virtual humans’, as artificial interlocutors of human users, demand planning processes to decide what to communicate and how to communicate it. With the SAIBA architecture [8], the community attempted to establish a model of behavior generation with a trichotomy between the Intent Planning ('function'), the Behavior Planning (choosing a 'form'), and the Behavior Realization layer, bridged by the Functional Markup Language (FML) and the Behavior Markup Language (BML), respectively. Bickmore [2] presented a model that allowed agents to actively manage the interpersonal distance by navigating, at the level of Intent Planning, through a taxonomy of dialogue topics to reach a point where relational constraints for intimate questions were met. We propose that such selection and gradual navigation is also effective and possible on the behavior ('form') level, accounting for adaptability in the intermediate layer between intent planning (where knowing what the other wants and needs is the basis for adaptation) and low-level alignment (where superficial features of behavior realization synchronize, as with *mimicry* [9, 7]). A study by Bergmann & Kopp [1] demonstrated that human users rate an agent's communicative behavior highest when it is produced with a production model learned from a single human speaker's behavior. This suggests that behavioral coherence, i.e, the extent to which people can familiarize with and predict behavior, is a key ingredient for effective and acceptable interactive agents.

In this paper, we propose a quantifiable model that shall allow an ECA to formulate its multimodal behaviour not only based on criteria of communicative effectivity and efficiency, but also in pursuit of additional social intentions. This model is meant to revise and refine expectations about the overall effects of utterance selection. It enables flexible and continuous adaptation in behaviour formation as well as crystallizing towards stable patterns that reflect familiarity between the agent and its user.

2 From utterances to socio-communicative acts

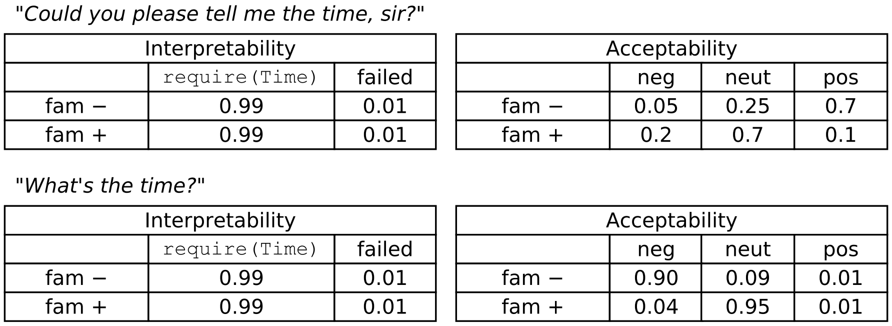
According to the Dynamic Interpretation Theory of dialogue [5], dialogue acts may carry meaning in more than one functional dimension: In addition to the presentation or request of task-specific information, other functions include management of contact, dialogue turns, timing, dialogue structure and topic, error signaling and correction, and feedback functions revealing the own state and the estimated state of the interlocutor.

Additionally, two functional domains underrepresented in the current taxonomies are worth considering: Firstly, emotional functions aimed both at signaling the own emotional state consciously, and at eliciting a change in the emotional state of the other, like phrasing an utterance in such a way as to give a comforting undertone; secondly, functions manipulating the interpersonal relationship, such as actively selecting a commanding stance to assert dominance.

We propose to move towards what we call "socio-communicative acts", by assigning to each producible utterance an independent probability distribution for a subset of relevant independent dimensions from the taxonomy. Each distribution is meant to represent the uncertainty about the effect of the utterance on the internal state of the human interlocutor in the respective dimension. These distributions can be hand-crafted or learnt from a corpus.

Using socio-communicative acts, the situation from the introductory example can be modeled as in Fig. 1, presenting two utterances for asking someone for the time. The utterances are associated with two distributions *Interpretability* and *Acceptability*, corresponding to the listener's expected interpretation of the utterance (content) and her social assessment of the way this content was put, respectively. Both distributions are dependent on the variable *Familiarity* (two-valued, for the sake of simplicity: *fam +* and *fam –*), which represents the influence of familiarity between receiver and producer (defined as social closeness and personal common ground) on the expected appraisals.

In the example, no effects of familiarity on the expected interpretation are present (both are unambiguous, with a 1% chance for failure). However, the utterances differ with respect to the expected acceptability of the recipient towards the selection of the utterance: while the longer sentence is received less favorably with known persons, it is a hardly objectionable way to ask a stranger. The concise request however is expected to have a negative effect on strangers, while acquaintances are neutral towards it.



"Could you please tell me the time, sir?"

Interpretability		
	`require(Time)`	failed
fam −	0.99	0.01
fam +	0.99	0.01

Acceptability			
	neg	neut	pos
fam −	0.05	0.25	0.7
fam +	0.2	0.7	0.1

"What's the time?"

Interpretability		
	`require(Time)`	failed
fam −	0.99	0.01
fam +	0.99	0.01

Acceptability			
	neg	neut	pos
fam −	0.90	0.09	0.01
fam +	0.04	0.95	0.01

Fig. 1. Two *socio-communicative construction* candidates

This representation of the degrees of belief about the socio-communicative effects of a construction allows for selecting utterances U from a repertoire by favoring utterances that yield a high probability for being understood correctly $P(\mathit{Interpretability} = \texttt{require(Time)} \mid U)$, while at the same time satisfying conditions from the social motivational system, namely not affronting its interlocutor $P(\mathit{Acceptability} \neq \text{neg} \mid U)$, and energy conditions, namely aiming for brevity when such is an option. Note that the 'best solution' towards the social motivation constraint in this instance is dependent on the estimated familiarity with the user. As long as no evidence is present that the system is talking to a known user, it must make a prior assumption, for example that they are to be treated as unfamiliar. Fig. 2 shows a Bayesian network capturing the assumed causations in the generation process. As can be seen, we propose Social Motivation and Intended Semantics to be the main factors influencing utterance selection, allowing both the determination of preferable behavioral forms – using the Interpretability and Acceptability distributions of production candidates – as well as the adaptation of this process according to observations and evidence gathered on their effectiveness during interactions.

Adaptation of the repertoire When an observation is made to support or challenge an assumption made during utterance production (Fig. 2, 'Indicators'), the socio-communicative act is updated with this new evidence, changing its probability distributions. For the above example, utterances which the user fails to understand are weakened in the hypothesis that the utterance fails to convey the desired semantics, while utterances for which interlocutors indicate that they liked or disliked the way they were talked to change the *Acceptability* distribution.

Familiarity through knowledge Over repeated interaction, the agents learns more about the effects of its actions, which helps it phrasing in a way that is understood with less confusion and received well by a specific dialogue partner. It is a core assumption that the maximization of these criteria can only work for single (or small groups of) users, the adaptions to whom taking place in repeated interactions. The degree to which the distributions are subject to change – or have 'settled' – can give a notion about the familiarity present in the dyad.

Active social behavior To create familiarity and social connectedness, following an intention for a beneficial modification of the dyad, the system can decide not to select the least objectionable utterance in an unfamiliar dyad, but to take the initiative in shaping the dyad by choosing a slightly more 'audacious' style of expression, hoping to not affront the user but making them more prone to reciprocate and thus accept subsequent informal utterances. With a similarly socially-annotated repertoire used for utterance interpretation (combining likely meanings of the user's utterances with an estimation of their informality), one can thus produce a familiarity-enhancing feedback loop in the ECA.

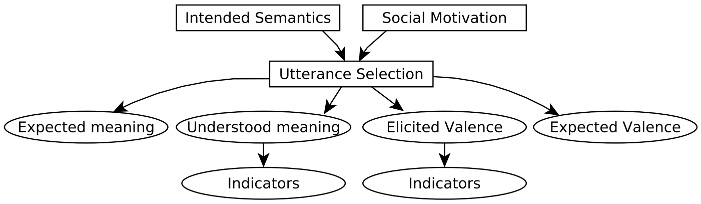


Fig. 2. Causation in the generation process

While the anticipated social effects from the socio-communicative constructions can be harnessed in a ‘fire-and-forget’ fashion, using a purely forward planning process as in SAIBA, their real strengths will be realized when they are embedded in an architecture that considers actually evoked effects (learnt from observation) to revise wrong and strengthen correct assumptions about constructions. The following section will introduce such a closed-loop account of behavior production and revision.

3 An adaptive generation model

To enable socio-communicative behavior generation in the sense described above, the SAIBA pipeline needs to be extended by an adaptive behavioral repertoire as well as a rich situation model (Fig. 3) that allow for considering potential effects of actions and revising beliefs about generable behaviors.

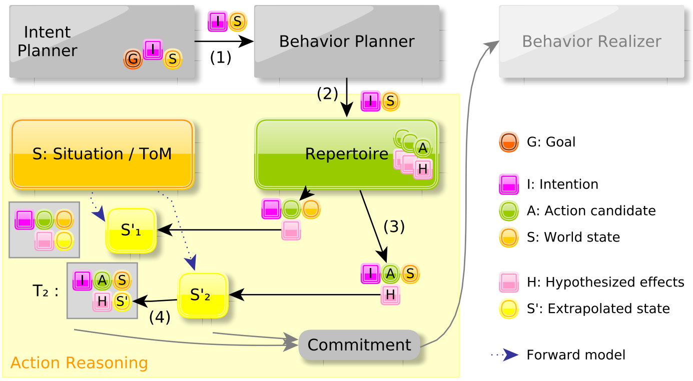


Fig. 3. Behavior selection: forward models explore uncertain action results.

Propositions and presumptions regarding the ongoing conversation are assumed to be stored in the ‘situation model state’ S known to the agent (in Fig. 3, ‘S’). When a state is formed as the agent’s goal G, differences between the current and the desired states are selected as the *intention* I. Both I and S are made available to the Behavior Planner (1). Each variable in I is considered a *desired effect*. The behavioral repertoire is searched (2) for all possible actions A likely to satisfy the desired effects when being performed. Note that the repertoire can be a lexicon of socio-communicative constructions, but could also be a generative production system that outputs new behaviors on the spot. This search produces a set of action candidates $A_1 \ldots A_n$ that are likely to satisfy most of the intended effects and to carry a number of known probable side-effects (3). The whole set of *hypothesized effects* is denoted $H_{A_i} = \{E_j\}, H_{A_i} \supset I$, with probabilities $0 < P(E_j|A_i, S) \leq 1$. The system produces a set of hypothetical extrapolated worlds $\mathcal{S}$ incorporating these effects (4), yielding tuples $T(S, I, A_i, H_i, S'_i)$. The action reasoning module rates the tuples according to a utility function over their extrapolated world states, and has to commit to one of them. The associated behavior is then realized, and the world state is updated with the *overlay* S' of the executed tuple, resulting in a new conjectured situation model state, conditional on the success of the excuted action.

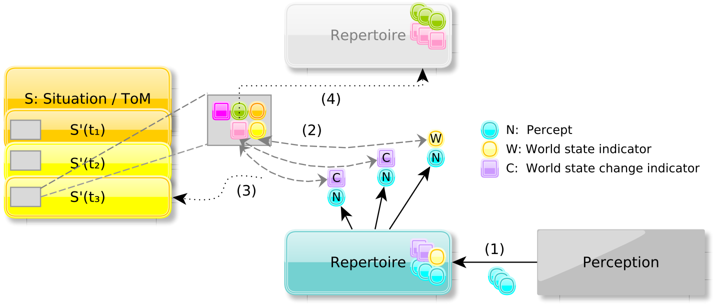


Fig. 4. Revision: Perceptual evidence (1) can lead to repertoire modification (4).

For the adaptation mechanism, any information from the perceptive system is analyzed for matches in the agent's understanding repertoire (Fig. 4 (1)). The system produces likely hypotheses about the indicative power of the received utterance for world state changes. Each overlay is checked (2) – if evidence contradicts a prior assumption, the associated A can be considered to have failed to convey that effect. The generating repertoire is informed of this, and is expected to adapt, lowering $P(E_i|A_i, S)$ for the future (4). If an observation confirms a hypothesized state from the overlay, the associated executed action can be assumed to have successfully caused the effect, and the conditional probability is raised. In case of a failure, the overlay must also be revised, so as not to negatively impact future intent planning due to an erroneous situation model state (3). When unplanned effects are observed, they could be tentatively attributed to the most recent utterances as new possible effects. Familiarity is seen in this sense as the reduction of uncertainty (or entropy) through evidence gathered in repeated interactions. Note that this model does not maintain the assumption of strict modularity of the intent and behavior planning systems.

4 Summary

In this paper we have proposed a model for ECA behavior generation that is capable of increased adaptation to individual users, with the goal to enable an agent to communicate with users both with higher communicative efficiency and social acceptability. The underlying assumption is that agents should be able to build up familiarity and increase social connectedness with a user by means of a personalized intent–behavior mapping, as found in human dyads. Additionally, the model proposed here can be employed for identification of different users by their mode of expression in reply to the agent's actions. We are currently implementing the proposed model for an agent serving as a personal calendar assistant. A study on human–human communication and familiarization in the same scenario has also been carried out; the resulting corpus is currently being annotated, with attention to standard vs. colloquial realizations of utterances, one important criterion for the implementation of the socio-communicative constructions for the calendar assistant. These data and the model will serve as a basis for enabling this agent to actively create familiarity with its user for the benefit of both effectiveness and social acceptability of human-agent interaction.

Acknowledgements – This research is supported by the Deutsche Forschungsgemeinschaft (DFG) in the Center of Excellence in 'Cognitive Interaction Technology' (CITEC).

References

1. Bergmann, K., Kopp, S., Eyssel, F.: Individualized gesturing outperforms average gesturing: evaluating gesture production in virtual humans. In: Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA 2010), LNCS (LNAI), 6356, pp. 104–117. Springer, Heidelberg (2010)
2. Bickmore, T.: Relational agents: effecting change through human-computer relationships. PhD Thesis, MIT, Cambridge, MA, USA (2003)
3. Brennan, S.E., Hanna, J.E.: Partner-specific adaptation in dialog. In: Topics in Cognitive Science, 1, pp. 274–291 (2009)
4. Brown, P., Levinson, S.C.: Politeness: Some universals in language usage. Cambridge University Press (1987)
5. Bunt, H.: The DIT++ taxonomy for functional dialogue markup. In: Heylen, D., Pelachaud, C., Catizone, R., Traum, D. (eds.): Towards a standard markup language for embodied dialogue acts (AAMAS 2009 Workshop), pp. 13–23. Budapest, Hungary (2009)
6. Cassell, J., and Gill, A.J., Tepper, P.A.: Coordination in conversation and rapport. In: Proceedings of the Workshop on Embodied Language Processing (EmbodiedNLP 2007), pp. 41–50. Prague, Czech Republic (2007)
7. Kopp, S.: Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. In: Speech Communication, 52, pp. 587–597 (2010)
8. Kopp, S., Krenn, B., Marsella, S., Marshall, A.N., Pelachaud, C., Pirker, H., Thórisson, K.R., Vilhjálmsson, H.H.: Towards a common framework for multimodal generation: The behavior markup language. In: Proceedings of the 6th International Conference on Intelligent Virtual Agents (IVA 2006), LNCS (LNAI), 4133, pp. 205–217. Springer, Heidelberg (2006)
9. Lakin, J.L., Chartrand, T.L.: Using nonconscious behavioral mimicry to create affiliation and rapport. In: Psychological Science, 14, pp. 334–339 (2003)
10. Pickering, M.J., Garrod, S.: Toward a mechanistic psychology of dialogue. In: Behavioral and Brain Sciences, 27, pp. 169–190 (2004)