A Cognitively Motivated Architecture for an Anthropomorphic Artificial Communicator

Nadine Leßmann (nlessman@techfak.uni-bielefeld.de)
Faculty of Technology, University of Bielefeld
33594 Bielefeld, Germany

Ipke Wachsmuth (ipke@techfak.uni-bielefeld.de)
Faculty of Technology, University of Bielefeld
33594 Bielefeld, Germany

Introduction and Challenge

This poster presents demonstrable work towards developing an architecture for an artificial conversational agent which aims to integrate communicative abilities with autonomous intelligent behavior. The context for our research is the Collaborative Research Center SFB 360 which is concerned with situated artificial communicators. Since situatedness involves embodiment, emphasis is laid on the agent's physical representation. The application scenario consists of a task-oriented, mixed-initiative discourse between an instructor and a constructor who collaboratively build aggregates, like a model aeroplane, from a wooden toykit ("Baufix"). Multimodal interaction including interpretation abilities as well as production competency plays a central role in the scenario.

One realization exhibiting such abilities is an artificial anthropomorphic agent embodied in large-scale virtual reality, "Max" (cf. Figure 1). Max is capable of interpreting multimodal input comprising speech and gesture and of producing multimodal output including gestures, facial expressions, and speech. The agent's movements are calculated on the basis of a kinematic skeleton as real-time computer animations (Kopp & Wachsmuth, 2000). Max has access to a knowledge base about possible connections of Baufix-parts and has additional knowledge on how to construct specific aggregates. The production of multimodal utterances and intelligent behavior on the one hand and the perception of the environment and interpretation of multimodal communicative acts on the other hand demand a complex interplay of sensory, cognitive and actoric abilities:

- the agent should be able to perceive his environment and especially the user, interpreting perceived input in a cognitively motivated way
- situated communication goes beyond a pure processing of instructions insofar as it requires dialog competences in understanding and generating context-dependent utterances
- to approach this aim a discourse memory and a partner model have to be integrated
- turn-taking is important in mixed-initiative dialog characterized by asynchrony, change of initiative, openness and unpredictability
- the agent's embodiment plays an important role in providing the possibility to use several channels to convey information

Figure 1: User and Max conducting multimodal dialog.

For the realization of an embodied conversational agent, different research areas are relevant, i.e., autonomous agent architectures, cognitive architectures, dialog systems, ECAs, and speech acts. The demand for both deliberation and reactivity suggests a hybrid approach to be used. Architectures that lead our work comprise on the one hand approaches such as (Gat, 1997), and (Blumberg & Galyean, 1995). On the other hand, approaches modeling cognitive processes such as Soar, ACT-R, and PECS have influenced our work, and we explicitly build on JAM (Huber, 1999). The third research area of interest is that of dialog systems and especially embodied conversational agents. Works of (Cassell et al, 2000), FXPAL and Speech-act theory are relevant for the development of our architecture.

An Architecture for a Conversational Agent

The conceived architecture is outlined in Figure 2 and at first displays the classical *perceive-reason-act* triad. In addition there is a direct *perceive-act* connection which marks a reactive component. Incoming sensor information can directly trigger reactive responses. The deliberative component is found in the *reason* section. A concurrent, parallel processing in the triad enables the simultaneous calculation of reactive responses and deliberative intentions. Both modules are supplied with permanent feedback information originating at different abstraction levels including the physis. The deliberative module represents communicative acts as intentional actions. It conducts dynamic, self-contained operating planners, including a turn-taking planner, and several memories, e.g. one for discourse. Both the reactive and the deliberative component use behaviors to perform their actions which are provided with priority values.

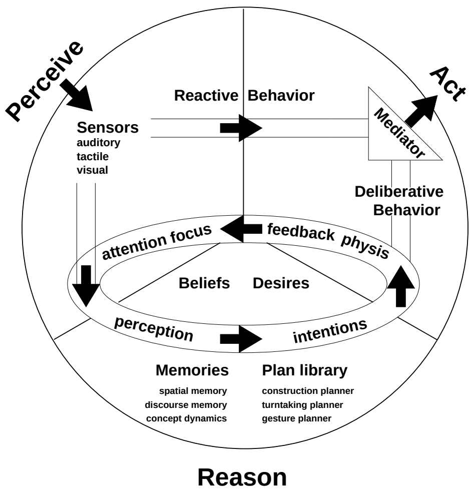

Figure 2: Architecture design.

The perceived information as well as the feedback information get interpreted by attention foci and the results are written in different memories which use specialized representations. All memories write the most relevant information in the working memory, the *beliefs*. This either happens when new important data is recognized or when the focus of attention is shifted. In our scenario, Max's perception is simulated through virtual sensors and real cameras to perceive the user. The core of the deliberative module consists of a BDI kernel being partly based on JAM. *Desires* can be asserted from internal processing but also through interactions from the outside. Persistent top-level goals are differentiated from instantiated subgoals. Possible courses of action are represented as plans with preconditions, context conditions, an effect and a utility function. The plan library contains simple plans to trigger specific behaviors and dynamic planners which construct complex plans, all expressing their competence by utility values. The plan with the highest utility value (the *intention*) is allowed to manipulate internal variables or to instantiate behaviors to act which compete for control together with the active behaviors of the reactive component. A mediator resolves conflicts and selects the most adequate behavior. In addition it allows all behaviors access to actuators which can act simultaneously without interfering with the chosen behavior. The behaviors differ in their complexity.

The reactive module of the architecture is firstly responsible for secondary behaviors to make the agent appear more lifelike and secondly it serves for fast responses providing system protection in form of reflexes. In the field of conversation reactive behavior can be found e.g. in turn-taking activities. Being responsible for intention building as well as serving as a basis for action selection and behavior activation, the priority and utility calculation play a central role. The priority value development of instantiated behaviors can be guided by time considerations and success.

The current status of our work includes two applications. First, Max can explain the complete construction of an aggregate. Using synthetic speech and gestures, he explains how to connect the individual Baufix parts. Secondly, there is an interactive construction scenario in which Max explains the construction steps and waits for the user to perform the actions. By way of a simple perception of the scene, Max is able to evaluate the success of the user's action steps and to provide feedback. He has also rudimentary turn-taking abilities and is equipped with reactive and secondary behaviors.

Acknowledgments

This research is partially supported by the DFG in the Collaborative Research Center SFB 360.

References

Blumberg B.M., & Galyean, T.A. (1995). Multi-level direction of autonomous creatures for real-time virtual environments. *Computer Graphics 29*, 47-54.

Cassell, J., Bickmore, T., Campbell, L., Vilhjalmsson, H., & Yan, H. (2000). Human conversation as a system framework: Designing embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied Conversational Agents* (pp 29-63). Cambridge (MA): The MIT Press.

Gat, E. (1997). On three-layer architectures. In D. Kortenkamp, R.P. Bannasso, & R. Murphy (Eds.), *Artificial Intelligence and Mobile Robots.* MIT/AAAI Press.

Huber, M.J (1999). JAM: A BDI-theoretic mobile agent architecture. *Proceedings of the Third International Agents'99* (pp. 236-243). Seattle, WA.

Kopp, S., & Wachsmuth, I. (2000). A knowledge-based approach for lifelike gesture animation. In W. Horn (Ed.), *Proceedings of the 14th ECAI 2000* (pp. 661-667). Amsterdam: IOS Press.