Towards a Cognitively Motivated Processing of Turn-Taking Signals for the Embodied Conversational Agent Max

Nadine Leßmann, Alfred Kranstedt, Ipke Wachsmuth
Artificial Intelligence Group, Faculty of Technology
University of Bielefeld, D-33594 Bielefeld
{nlessman, akranste, ipke}@techfak.uni-bielefeld.de

Abstract

Max is a human-size conversational agent that employs synthetic speech, gesture, gaze, and facial display to act in cooperative construction tasks taking place in immersive virtual reality. In the mixed-initiative dialogs involved in our research scenario, turn-taking abilities and dialog competences play a crucial role for Max to appear as a convincing multimodal communication partner. The way how they rely on Max's perception of the user and, in special, how turn-taking signals are handled in the agent's cognitive architecture is the focus of this paper.

1. Introduction

This work is embedded in the Collaborative Research Center SFB 360 which aims at realizing situated artificial communicators. Our research scenario is concerned with task-oriented discourse between an instructor and a constructor building aggregates, such as a model airplane, from parts of the *Baufix* toykit [16]. These dialogs take place in a CAVE-like virtual environment in a face-to-face manner. The scenario permits the exploration of the interplay between speech, gestures, dialog competences, knowledge, and planning as well as sensomotoric aspects in a restricted setting. The anthropomorphic conversational agent Max was developed in this context. On the one hand Max is able to interpret multimodal (speech and gesture) input by a human instructor, and on the other hand he has abilities to produce multimodal output involving synthetic speech, facial display, and gesture.

The topic of the dialog is restricted to the assembly realm and may be set or switched by the user at any time. Max is obliged to assist the user and hence must adopt the topic by, first, planning the requested assembly explanations and, second, demonstrating the construction procedure in a step-by-step manner, sometimes committed to initiate actions himself when the user refuses or hesitates to do so. At any stage, the discourse is influenced by the situational context, e.g., the mutual consent on individual parts employed so far, the state of the ongoing assembly, or the outcome of a user action.

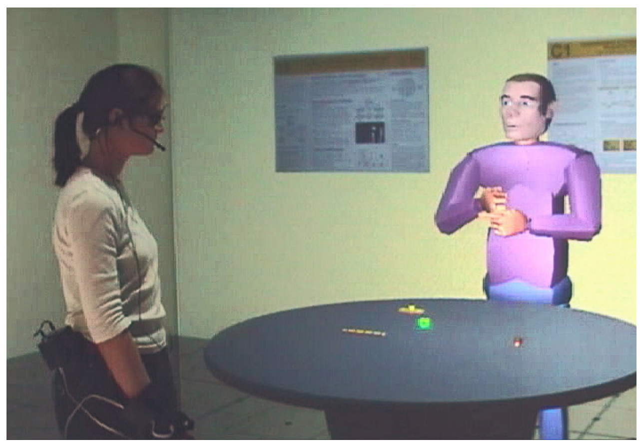


Figure 1. In a CAVE-like virtual environment a user meets the multimodal communication partner Max.

In this paper we present ongoing work on equipping Max with advanced dialog competences, in particular, turn-taking abilities. In Section 2 we discuss requirements and challenges arising in this context. How these are approached in related and in our own work is presented in Section 3. Section 4 in detail explains the interplay of different perceptual modules within a cognitively motivated agent architecture and how this leads to the agent's turn-taking capabilities. In Section 5 examples are described illustrating the agent's current communicative abilities. In the last section we give some ideas for future work.

2. Requirements

Realization of an embodied conversational agent as a pleasant and convincing communication partner has a lot of challenges that pertain to different aspects. The agent should be able to perceive the environment and especially the user, and interpret that what is perceived in a cognitively motivated way. An anthropomorphic appearance of the agent yields expectations by users which, e.g., concern the possible field of vision of the agent. Moreover, situated communication goes beyond a pure input-output processing of instructions in that it requires dialog competences that involve understanding and generating context-dependent utterances.

One advantage of an embodied agent is the possibility of using several channels for conveying information about the agent's inner state. For instance, Max can employ facial expressions for feed-back and gestural movements of his body simultaneously while explaining a construction step by speech. In addition multimodal production of utterances makes it easier to refer to an object, e.g., by using a deictic gesture. These abilities have to be integrated in the architecture and the agent must coordinate intelligent behavior with communication acts fulfilling the characteristics of mixed-initiative dialogs in which turn-taking plays an important part.

Mixed-initiative dialogs are characterized by asynchrony, changes of initiative, openness, and unpredictability of discourse. Max needs to keep track of the dialog state w.r.t. *turn*, *initiative*, *topic*, and *obligation*. By initiative we consider the power to seize control of the dialog by presenting or confining a domain goal for the interlocutors to achieve. Thus, sudden switches of initiative may occur, e.g., when the user asks for explanation of a new aggregate at some stage in the discourse, but also when Max explains the user how to conduct an assembly action, possibly bringing up the same goal again. Besides switches of initiative, both Max and the user may take the turn or assign it to the interlocutor.

In our cooperative construction scenario, Max is supposed to act as an autonomous agent pursuing his own goals, but also to interact with the user. Therefore gaze and turn-taking gestures help the user to get indication of the agent's mental state. For instance, by paying attention to turn-taking signals the user may be able to tell whether the agent wants to say something or is just listening to him. Gaze further helps to recognize where the agent's attention focus is at that specific moment.

3. Related work

Early examples of embodied conversational agents that conduct multimodal dialog with a human user are Gandalf [19], who can answer questions about the solar system, or REA [3], who provides house descriptions in the real-estate domain. These systems focus on the processing of multimodal input and output, i.e., how information is intelligibly conveyed using synchronized verbal and nonverbal modalities.

The realization of synthetic agents engaging in natural dialog has drawn attention to questions on how to model social aspects of conversational behavior in dialog, in particular, turn-taking and feedback signals. Turn-taking, as a basic interactive mechanism for scheduling the speaker role in conversation, has been investigated since more than thirty years. Whereas conversation analysis emphasizes the context-free rule-base character of this mechanism [17], Duncan [6] and successors have done empirical investigations which document the role of interactive signals for the negotiation of the speaker role. Both these aspects are reflected in modern dialog theories which emphasize the interactive character of dialog (e.g., [7, 5]).

The Ymir architecture developed for Gandalf [18] played a fundamental role for the development of computational models for turn-taking mechanisms in human-machine communication. Motivated by the work of Goodwin [7], central aspects were the explicit detection of interactive functions concerning turn-taking (giving-turn, taking-turn, and wanting-turn) in the incoming signals from the dialog partner and their processing in an interaction loop. This approach was integrated in the FMTB architecture [4] demonstrated with REA. The allocation of the speaker role is explicitly represented by conversational states, and possible changes are modeled by a finite state machine.

In the aforementioned systems, communication takes place in rather static scenarios, with the agent fulfilling the role of a presenter and the user only observing presented scenes. In contrast – and comparable to our assembly assistance scenario – many educational applications allow a human student to perform actions that are subject of a training process, while being monitored by a tutoring agent. Such agents thus need to combine communicative behaviors with the ability to observe, and react to, environmental changes. This poses greater demands on more general perceptual and cognitive capabilities. In the STEVE system [15], this has led to a general framework for modeling cognitive processes of an intelligent agent, based on Soar [11]. In recent work by Traum and Rickel [21], the STEVE architecture was extended by a comprehensive dialog system that accounts for multimodal, multi-party, and multi-utterance conversations with open, unpredictable dialogs. Based on the dialog theory of Clark [5], different layers of dialog management are modelled, each including a specific information state. A set of dialog acts can change that state. On the turn-taking layer five different types of dialog acts – called turn-taking actions (Take-turn, Request-turn, Release-turn, Hold-turn, and Assign-turn) – are classified which are responsible for shifting the turn-holder state.

For the Max system, we adopted this classification (with different labels, see Section 4.3 on turn-taking). As in the FMTB architecture we distinguish between a behavior and its conversational function, which is theoretically founded in the communicative act theory [13]. Finally, similar to the STEVE architecture, we chose to build Max's deliberative processes on top of a general model of agent rational reasoning. However, rather than Soar we adopted the BDI architecture [14], for it provides provisions for modeling intentional actions in the form of plans, which help to perform complex tasks under certain conditions while being interruptible and able to recover from failure.

4. Max in Dialog

4.1. Perception

As Max is situated in a virtual environment but also needs to perceive the user in the real world, different aspects of perception have to be taken into account.

First of all, Max needs to perceive his virtual environment. To this end he is equipped with virtual view sensors which simulate his point of view and calculate sighted objects in the virtual scene. The sensors register not only which objects are in the agent's field of view, but also their position, color, and type. The implementation of the virtual view sensors is done by attaching a view frustum at Max's eyes and calculating which objects are laying on the inside. Max is also provided with a simple visual short-term memory operating on the virtual sensors' data. Keeping protocols of the objects perceived, it is able to trigger events and reactions whenever there is a change, e.g., when an object disappears or a new object comes into sight. Another virtual input device is the scene simulator, which manages the physical properties of the objects and informs Max of their current connections, etc.. Information from the scene simulator enables Max to understand user actions more easily than if he would only rely on his view sensors.

The second kind of perception is that of the real world, i.e. the user, which is enabled by marker-based camera tracking, data gloves, and speech recognition. To collect data about the user's position and gaze, the glasses that the user wears have markers tracked by infrared cameras. By this the user's head position and head orientation are detected. Further, hand postures, positions, and movements are tracked by data gloves. A speech recognizer operates on vocabulary appropriate for the *Baufix* construction scenario. The collected multimodal data is interpreted by detectors using the PrOSA framework [12] realized with the AVANGO toolkit [20]. The diverse detectors are realized using compute nodes which can be combined in hierarchically organized compute networks [2]. For example, one detector responds to the user holding her hand up with fingers stretched (see Figure 3). The calculation of these "real-world" detectors runs in parallel with the virtual perception.

Information from the perceptual modules is forwarded to both the reactive and the deliberative modules of the architectural framework which is outlined in the next section.

4.2. Architecture

Max's overall behavior is controlled by a cognitively motivated architecture outlined in Figure 2. On the one hand it shows the classical perceive-reason-act triad with deliberation processes taking place in the reason section. On the other hand, reflexes and immediate responses are handled by a reactive component which has a direct connection between perceive and act. Processing in the triad runs concurrently such that reactive responses and deliberate actions are calculated simultaneously. Both the reactive and the deliberative module operate by the instantiation of behaviors which compete to control the agent, with both modules having the ability to overrule each other.

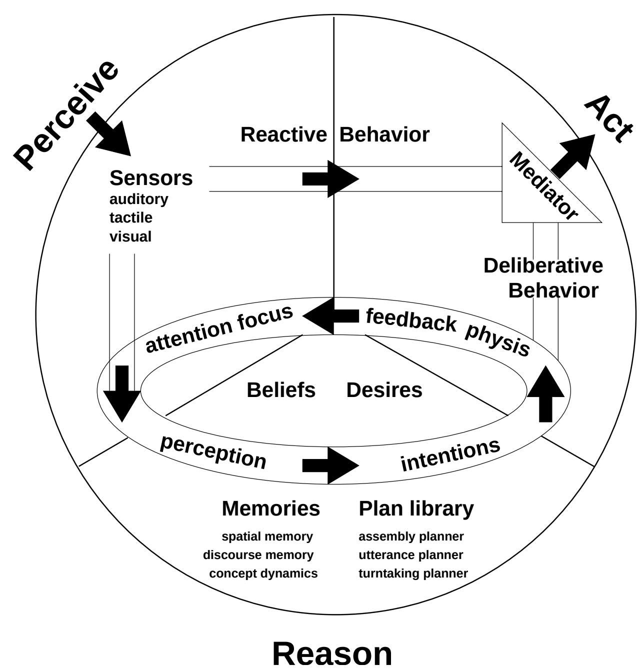


Figure 2. Overview of Max's architectural framework.

Behaviors of the reactive module, for instance, use sensor information expressing the user's head position to control gaze-following behavior. Information from the user's hand gestures, on the other hand, is used by the deliberative component to decide whether the user is trying to take the turn. Therefore it should be possible for Max to look up in reaction to noticing a significant movement of the user but then return to his interrupted behavior if he decides that it has not been important enough to change the intention he is pursuing.

At the core of the deliberative module is a BDI-kernel partly based on JAM [8] which operates on beliefs, desires, and a plan library. *Beliefs* are a part of the agent's working memory, *desires* represent the agent's goals emerging from internal processing as well as from interactions with the user and the environment, and the *plan library* contains possible courses of action which can be used to formulate intentions.

Possible courses of action are represented as plans with preconditions, context conditions, effect, and a utility function. Plans can either directly trigger specific action behaviors, or they may also invoke dynamic, self-contained planners which construct context-dependent plans that can be hierarchically expanded on demand by instantiating lower-level plans. The plan with the highest utility value is actually pursued by the BDI-interpreter.

The plan library contains plans of two different types, *goal-directed* plans and *data-driven* plans. Goal-directed plans contain a goal specification and are either of an *achieve* or a *perform* type. In contrast to *perform*-plans, *achieve*-plans are only executed as long as their specified goal is not achieved which is continuously checked. Data-driven *conclude*-plans can be used to model the agent's reaction in specific situations. These plans contain a data-relation which is watched over by the plan. If the relation changes or is inserted in the beliefs the precondition is checked and if there is a match the plan tries to become the current intention. Both the *achieve*-plans as well as the *conclude*-plans play an important role for realizing the agent's turn-taking abilities (see Section 5).

4.3. Turn-Taking

The control of Max's turn-taking behavior is integrated in the cognitive loop of the BDI module as well as in his reactive module. On the one hand the agent is to react when he receives turn-taking signals of the user, and align his plans. On the other hand, he must be able to use turn-taking mechanisms in an active manner, e.g., if he has the intention to communicate an utterance, he first must try to get the turn.

To handle such situations a turn-taking model is proposed which adopts concepts of the FMTB architecture developed by Cassell [4] and the dialog management layers of Traum and Rickel (see Section 3), and we are currently working on integrating it in our architecture. The turn-taking mechanisms for Max operate on an information state called *conversational state* and receive input from a set of detectors filtering *conversational functions* from the data produced by the user.

Conversational state	**Detected conversational function**	**Alternative reactions**
MyTurn	WantingTurn	GivingTurn YieldingTurn HoldingTurn
	TakingTurn	GivingTurn HoldingTurn
OthersTurn	GivingTurn	TakingTurn
	YieldingTurn	TakingTurn
	HoldingTurn	no reaction WantingTurn
Gap	WantingTurn	GivingTurn WantingTurn
	TakingTurn	no reaction
	GivingTurn	TakingTurn
	YieldingTurn	TakingTurn GivingTurn
Overlap	GivingTurn	no reaction
	YieldingTurn	no reaction GivingTurn
	HoldingTurn	GivingTurn

Table 1. Possible turn-taking reactions depending on the conversational state and the detected conversational function. Only those detected functions which bear a meaning with respect to the conversational state are listed, *no reaction* expressing that a turn-taking reaction is not required.

The *conversational state* should be distinguished from the subjective conversational role with the exclusive values *speaker* or *listener*. In contrast, the *conversational state* may not only represent the holder of the turn, but also gaps and overlaps, which are called failure states in most of the classical models for turn-taking. We prefer to name them *transitional states*. Despite the efforts to avoid such states they are a normal and frequently observed phenomenon in natural conversation where the turn often has to be negotiated. Dealing with them is an important component of turn-taking. As a consequence (in a conversation with one other participant) the conversational state can take four different values: *MyTurn*, *OthersTurn*, *Gap*, and *Overlap*. Analog to [21] we distinguish five *conversational functions*: *WantingTurn*, *TakingTurn*, *HoldingTurn*, *GivingTurn*, and *YieldingTurn*.

In the current implementation only a few detectors are available. We are able to detect signals with the functions *WantingTurn* (facing the agent and raising a hand), *TakingTurn* (raising a hand and saying halt), and *GivingTurn* (facing the agent, a metaphoric giving gesture, and spoken key words like ok). The number of detectors will be increased as works proceeds.

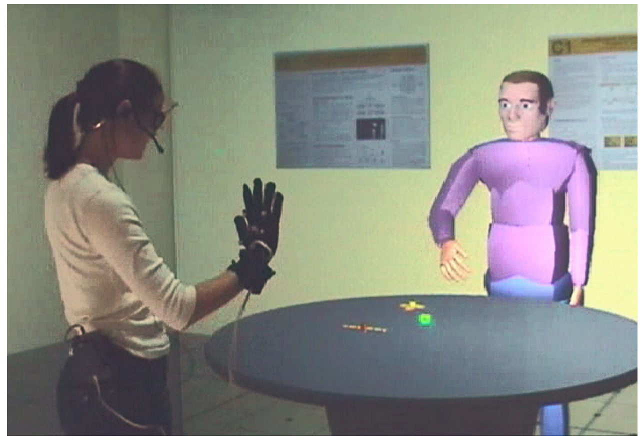
Dialog competences which operate on the described structures are realized on different levels of the architecture. The turn-taking mechanism itself consists of two steps. The *first step* is a rule-based, context-free evaluation of the possible turn-taking actions or reactions taking into account the current *conversational state* and the detected *conversational functions* revealed in the utterances of the conversational partner (see Table 1). These rules consist of a combination of the fundamental rules suggested by [17] together with simple rules for handling the transitional states *gap* and *overlap*. Altogether, these rules aim to ensure cooperative dialog behavior.

The *second step* of the turn-taking mechanism consists of a decision process between different courses of action and is integrated in the deliberative processes of the agent, leading to the instantiation of plans as intentions. For instance the agent is able to deliberately decide to try to take the turn if the situation seems suitable. Likewise, when the user tries to interrupt the agent, the processes generating a reaction take place in the deliberation process, but in this case data-driven plans are used. The agent may ignore the behavior of the user or can decide to respond to it. But in either case a significant signal should cause a simple reaction such as looking towards the source of the signal.

Generating the turn-taking actions of our agent Max, we go the reversed direction compared to the perception process, thus from conversational functions to concrete conversational behaviors. Once the agent has decided to show a turn-wanting behavior, this is accomplished by instantiating reactive behaviors automatically adapting to the situation, for example, gazing at the user even if he moves.

5. Experiences and Examples

In this section we present Max's conversational behavior resulting from our current model. When Max has the intention to communicate an utterance he first tries to get the turn. This is modelled by the desire to achieve that the conversational state has the value *MyTurn*. To deal with this, there exists an *achieve*-plan in the plan library. As an achieve-plan is only performed when the goal statement is not met the plan only produces actions when Max is currently not the speaker. In this case the plan is instantiated inheriting the communicational goal's utility. The applicable plans make Max perform a turn-wanting behavior which is increased until he gets the turn, or gives up. Max assumes that he has gotten the turn when the user gazes in his direction while finishing speaking.



Max: Insert this screw into the middle hole of a three-hole-bar.
User: [interrupts] Halt!
Max: [focuses on the user] Yes please?
User: [takes a close look at the scene, then returns the turn] Okay.

Figure 3. Max is interrupted by the user who performs a gesture to get the turn.

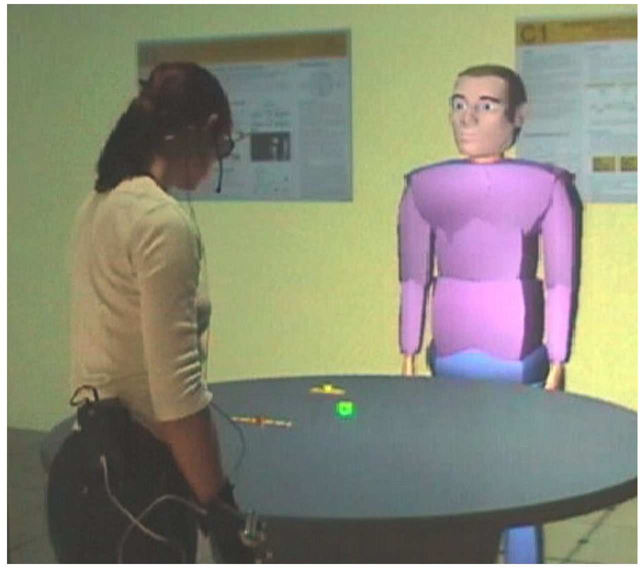
The loss of the turn is modelled by using data-driven *conclude*-plans. The precondition of these plans consists of the fact that Max believes the conversational state to be *MyTurn*. In the example illustrated in Fig. 3 the user claims the turn for herself while Max believes to be the speaker; she interrupts Max by telling him to halt and by performing a hand-up gesture. In this case Max will interrupt his current actions, shift his attention focus to, face the user and wait for the user to return the turn. After the user's turn, Max resumes his suspended actions if the context conditions of the interrupted plan are still valid.

In Figure 4 we present an example for the negotiation of the turn, serving as a motivation for the treatment of the transitional state *gap* as a legal conversational state. In the context of the construction of a propeller, Max proposes an assembly step to the user and yields the turn. Politely, he leaves it open who should go on with the construction. But he does expect a reaction of the user by showing either some confirmation or disconfirmation concerning the content of his proposal, or by the user performing the proposed action. As the user at first refuses to show any reaction, Max tries to encourage her to take the turn by performing some additional giving-turn behavior. But the user does not want the turn, she returns the turn to Max by performing a giving-turn gesture herself. Max accepts the turn and performs his proposed action.

The plans and mechanisms leading to the described behavior are the following. Max uses a *propose-performative* to express his ideas about the next construction step in order to achieve a specific assembly. The proposal ends with a *yielding-turn* gesture and Max changes his conversational role to that of a listener. Therefore both Max and the user are left in a situation in which neither of them is committed to the role of a speaker; the transitional conversational state *gap* occurs. As Max is expecting at least some reaction to his proposal, he tries to achieve the conversational state *OthersTurn*. For this purpose an *achieve*-plan, with the context condition of the conversational role being the listener, is used. The instantiated plan makes Max perform increasing turn-giving behavior. But the user still does not take the turn and instead performs herself a *giving-turn* gesture. As the conversational state is that of *gap* and Max realizes the conversational function *giving-turn* expressed in the user's gesture, he has to take the turn to be cooperative (see Table 1) and changes his role to speaker. This automatically leads to abandonment of the *achieve*-plan (*conversational state* being *OthersTurn*) as the context condition of the plan is no longer met. Because the user has shown some reaction and has not rejected his proposal and as Max has the role of the speaker, he tells the user that he will perform the action.

The examples illustrates not only the possibilities but also reveals limitations of our current technology. These limitations result especially from the input devices used. Detector nets which build on the aforementioned PrOSA framework allow detecting nearly every posture and a wide range of movements, but they are constrained by the incoming data. We can only track significant postures and well-formed movements. The gaze direction is computed using only the head orientation and also the speech recognizer is limited in its ability to detect verbal signals.

As for production the limitations are less strict. Currently Max is so far able to utter simple keywords as turn-taking signals and face the user if he wants to give the turn. When Max attempts to get the turn, he gazes at the user and raises his hand to signal that he wants the turn. The repertoire of Max's turn-taking behaviors can readily be extended. The utterance generators build on a database of utterances formulated in MURML, an XML-based representation language [10]. In this language it is possible to specify any coverbal hand and gaze gesture and a wide range of mimic postures in conjunction with co-uttered speech.



Max: Now we should turn the bars crosswise.
[expects the user to take the turn or to perform the assembly.]
User: [does not react; a conversational gap occurs]
Max: [after a while, performs a giving-turn gesture]
User: [refuses the turn by performing a giving-turn gesture herself]
Max: [takes the turn] Okay, then I will perform the action.

Figure 4. A scene in which the conversational state *gap* occurs is presented as well as the way Max handles this.

For the temporal synchronization between the different modalities an incremental model of speech and gesture production is used. It is based on the idea that continuous speech and gesture are co-produced in successive "chunks", whereby each chunk of speech-gesture production is a pair of an intonation phrase and a co-expressive gesture phrase. Within a chunk the synchrony between the affiliated word or sub-phrase and the gesture stroke is mainly accomplished by the gesture adapting to the structure and timing of running speech. In producing a single chunk, the intonation phrase can therefore be synthesized in advance, setting up timing constraints for co-verbal gestural or facial behaviors (for detail cf. [9]).

6. Conclusion and Future Work

To summarize, we have proposed ongoing work on equipping the conversational agent Max with more advanced turn-taking abilities. The classical turn-taking mechanism has been extended with respect to transitional conversational states like gaps and overlaps. In addition, the decision processes relating to the negotiation of the turn have been integrated into the deliberative processing loop of the agent. These approaches can be seen as first steps towards handling natural turn-taking phenomena in a cognitively motivated way. We are able to cover a number of simple cases and plan to extend the recognition as well as the production abilities.

As a next step Max will be equipped with additional peripheral view sensors which enable him to recognize gestures in a wider field of view. In the peripheral view area movements are more significant than static postures, but also harder to detect. There exist already some tools in the PrOSA network [12] that we can build on to detect specific movements. The detector nets used allow to combine the results of single detectors. So another next step will be to accumulate turn-taking signals registered in a short temporal sequence. This makes it possible to provide the detected conversational function with a parameter representing its urgency. The calculated urgency of all signals carrying the same function will be an important information source for the deliberative decision process on turn-taking.

In future work we will also address mechanisms initiating and finishing a conversational situation. These include the explicit representation of pre- or post-conversational situations in the *conversational state* characterized by a *conversational role* of one or more participants not engaged in dialog, and behaviors realizing *conversational functions* like *turn towards* and *turn away*.

We are also planning a closer integration of emotional aspects in the architecture to model relations between communicative behavior and emotional states. One idea is that of Max being more or less reluctant to give away the turn depending on concepts like dominance which could be realized on the basis of our emotion model for Max [1].

Acknowledgements

This research is partially supported by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center SFB 360.

References

[1] C. Becker, S. Kopp, and I. Wachsmuth. Simulating the emotion dynamics of a multimodel conversational agent. In *Proceedings of the workshop Affective Dialogue Systems, Kloster Irsee Germany*. Springer, 2004.

[2] P. Biermann and I. Wachsmuth. An Implemented Approach for a Visual Programming Environment in VR. In *Proceedings Fifth Virtual Reality Interaction Conference (VRIC 2003), Laval, France*, pages 229–234, 2003.

[3] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjalmsson, and H. Yan. Human conversation as a system framework: Designing embodied conversational agents. In *Embodied Conversational Agents*, pages 29–63. MIT Press, 2000.

[4] J. Cassell et al. An Architecture for Embodied Conversational Characters. In *Proceedings of the first Workshop on Embodied Conversational Characters, October 12-15 1998, Tahoe City, California*, 1998.

[5] H. Clark. *Using Language*. Cambridge University Press, Cambridge, UK, 1996.

[6] S. Duncan Jr. On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, 3(2):161–180, 1974.

[7] C. Goodwin. *Conversational Organization: Interaction between Speakers and Hearers*. Language, thought, and culture: Advances in the study of cognition. Academic Press, New York, 1981.

[8] H. Huber. JAM: A BDI-theoretic mobile agent architecture. In *Proceedings of the Third International Agents, Seattle, WA*, pages 236–243, 1999.

[9] S. Kopp and I. Wachsmuth. Synthesizing multimodal utterances for conversational agents. *Comp. Anim. Virtual Worlds*, 15:39–52, 2004.

[10] A. Kranstedt, S. Kopp, and I. Wachsmuth. MURML: A Multimodal Utterance Representation Markup Language for Conversational Agents. In *Proceedings of the Workshop Embodied Conversational Agents – let's specify and evaluate them, AAMAS 2002, Bologna, Italy*, 2002.

[11] J. Laird, A. Newell, and P. Rosenbloom. Soar: An architecture for general intelligence. *Artificial Intelligence*, 33(1):1–64, 1987.

[12] M. Latoschik. A Gesture Processing Framework for Multimodal Interactions in Virtual Reality. In *AFRIGRAPH 2001, 1st International Conference on Computer Graphics and Visualization in Africa*, pages 95–100, 2001.

[13] I. Poggi and C. Pelachaud. Performative Facial Expressions in Animated Faces. In *Embodied Conversational Agents*. MIT Press, 2000.

[14] A. Rao and M. Georgeff. Modeling rational behavior within a BDI-architecture. In *Proceedings Int. Conference on Principles of Knowledge Representation and Planning*, pages 473–484, 1991.

[15] J. Rickel and W. Johnson. Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. *Applied Artificial Intelligence*, 13:343–383, 1999.

[16] G. Rickheit and I. Wachsmuth. Collaborative Research Centre "Situated Artificial Communicators" at the University of Bielefeld, Germany. *Collaborative Research Centre "Situated Artificial Communicators" at the University of Bielefeld, Germany*, 10(3-4):65–170, 1996.

[17] H. Sacks, E. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50:696–735, 1974.

[18] K. Thorisson. *Communicative Humanoids: A Computational Model of Psychosocial Skills*. PhD thesis, MIT Media Laboratory, 1996.

[19] K. Thorisson. Gandalf: An Embodied Humanoid Capable of Real-Time Multi-Modal Dialog with People. In *Proceedings First Int. Conference On Autonomous Agents*, pages 536–537, 1997.

[20] H. Tramberend. Avocado: A distributed virtual reality framework. In *Proceedings of IEEE Virtual Reality 99*, pages 14–21, 1999.

[21] D. Traum and J. Rickel. Embodied Agents for Multi-party Dialogue in Immersive Virtual Worlds. In *Proceedings First Internat. Joint Conference on Autonomous Agents and Multi-agent systems AAMAS02*, pages 766–773, 2002.