Engagement in Collaborative Construction Tasks with Max

Nadine Leßmann, Stefan Kopp
Artificial Intelligence Group, Faculty of Technology
University of Bielefeld, D-33594 Bielefeld
{nlessman, skopp}@techfak.uni.bielefeld.de

Abstract

In Virtual Reality environments, real humans can meet virtual humans to collaborate on tasks. The agent Max is such a virtual human. In construction tasks, he is standing face-to-face to a human partner as a co-situated companion. To maintain a social conversation, Max has to perceive the user's engagement and to solicit the user's engagement, he needs to demonstrate engagement himself. This paper presents ongoing work on the development of a model that describes interacting levels of engagement. We present how the agent's perceptive, cognitive, and expressive capabilities interact on these levels ranging from observing the human partner, to taking her goals, mental state or emotions into account when making decisions on how to interact and intervene.

1. Introduction and Motivation

During the last years, the anthropomorphic agent Max [4] has been developed at Bielefeld University's A.I. Group in cooperation with the Collaborative Research Center SFB 360 (Situated Artificial Communicators). Our research aims to investigate aspects of situated communication and collaborative, mixed-initiative discourse in a changing environment. Max meets the human user in a CAVE-like VR environment where he is supposed to assist in building complex aggregates out of the Baufix toy-kit (see Fig. 1). The user, who is equipped with stereo glasses, data gloves, optical position trackers, and a microphone, can issue natural language commands along with coverbal gestures to trigger construction steps or to interact with Max. Or she can directly grasp and manipulate the 3D Baufix models to carry out assembly actions. At any time, the user can address Max and ask him how to go on with the assembly of an aggregate. Max is able to guide the user interactively through assembly procedures, combining capabilities of performing construction steps himself with capabilities of mixed-initiative dialogs in which he employs speech, gaze, facial expression, and gesture (see [5] for the generation of multi-modal behavior). While the overall discourse is guided by the user's wish to build a certain assembly, the interaction is open in that the roles of instructor and constructor may switch according to the competences of the interaction partners, i.e., both may assemble by themselves or instruct the other to perform an action.

Figure 1. Max encourages the user to engage in the interaction.

With his perceptive, cognitive, and expressive capabilities becoming more developed and powerful, Max has turned into what can be considered a social presence, i.e., an (artificial) interaction partner that recognizably possesses and is led by cognitive qualities such as beliefs, goals, desires, or emotions. At this point, it becomes more and more apparent that we must generally draw our attention towards the social aspects of the task-oriented collaboration both partners are pursuing. What are the crucial ingredients for a successful joint activity of two partners with human (or human-like) characteristics? We focus here on engagement, the degree to which the human and the artificial interaction partner take part and are actively involved in their collaborative task and in the ongoing conversation.

We define engagement in construction dialogs according to Sidner et al. [10]: *"Engagement is the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake.*" In our case, the joint interaction lies in the assembly realm. In this paper, we propose a model which to account for the processes and the connections engagement both brings about and relies on in cooperative assembly tasks. We present ongoing work on using this model as a basis to enable Max to create and maintain bonds with his human construction partner through engagement. In the next section, we first describe the model and then discuss requirements and challenges that arise from simulating engagement in collaborative construction tasks with an artificial agent. In the following sections, we address in a more detailed way the agent's abilities affording engagement behavior and how they connect to the different levels of our model. First, the agent's perception and recognition abilities needed to classify and understand the user's behavior are presented.

How all of these aspects affect the agent's decision and the behavior generation process, is then shown in Section (3.5). These discussions lead into interaction examples in Section (4), presenting first achievements of engagement behavior. The last Section (5) finishes our paper with a conclusion and our plans for future work.

2. Collaboration and Engagement - Requirements

The realization of an embodied conversational agent as a both engaging and engaged collaborator poses a lot of challenges. We conceive of the engagement that is taking place between two interactants as arising out of a complex interplay between several interacting processes. Fig. 2 shows our process model of engagement which is meant to capture the relevant processes from the perspective of one agent in the interaction. At first, it distinguishes between the *perception* and the *generation* of engagement-relevant behaviors. To act successfully in a shared environment, the agent must be able to closely monitor both the environment and the user, and he must interpret all percepts to understand the situational context - the shared (virtual or physical) environment, in which complex reference to objects or events and joint manipulations of objects are carried out. As engagement is a bidirectional process, the agent does not only need to look out for engagement cues, but also needs to demonstrate engagement in the interaction himself to make the human user feel accepted and equally involved. The perception layer comprises all processes that are concerned with how the agent perceives the behaviors of the user which may indicate and influence the user's engagement in the collaboration. Likewise, the generation layer involves the actions the agent carries out in order to engage in the interaction himself and to demonstrate this to the user.

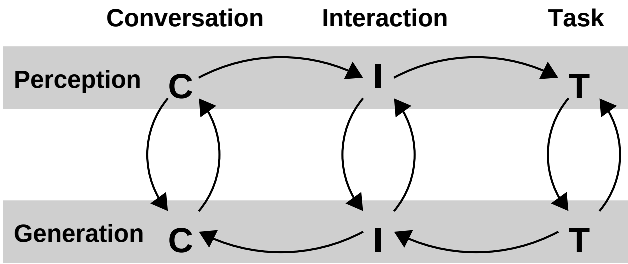


Figure 2. Interacting levels of engagement.

Orthogonal to this two-layer structure, the model distinguishes between three levels at which engagement behavior can take place. The first level accounts for *conversational engagement*; these processes take place on a lower level and consist mainly of reactive behaviors including gaze and glance behavior and the production of immediate feedback signals to insure the interlocutor that one is following the conversation. Then, on a higher level, *interactional engagement* can be classified. Turn-taking behaviors and social rules for beginning and disengaging from a conversation belong to the interactional engagement repertoire. These rules also incorporate the agent's conversational obligations and expectations. The third level accounts for *task-level engagement* which deals with the impact certain actions and interactions have with respect to the establishment and progress of a joined task.

At each of these levels, perception and generation are closely interwoven by feedback loops that ensure the agent's reaction to respective cues from the user. For example, when the user references an object, the agent should immediately show that he is listing and following the conversation, e.g. by glancing at the referred object. At the interaction level, the agent should immediately demonstrate his willingness to interact, e.g. by nodding when the user asks him for help. Finally, the agent must also engage on the task level, by adopting the user's goal, carrying out the required reasoning processes to determine a proper answer, and possibly even committing himself to long-term behavior, e.g., to guide the user through a complex procedure.

Indicated by the inner loops at each level, all of these processes involve perceiving the user's action or wants, and reacting with behavior that demonstrates the agent's willingness to collaborate and desire for a successful interaction. In addition, just like the three levels of conversation interaction, and task depend on and rest upon each other, do the processes trigger and influence each other across these levels. On the perception layer, the direction of inter-level interaction is thereby from conversation to interaction to task, whereas it is the other way around for generation processes.

Based on our model, we can derive several requirements for successful engagement. Sidner et al. [10] point out that "*engagement is supported by the use of conversation (that is spoken linguistic behavior), the ability to collaborate on a task (that is, collaborative behavior), and gestural behavior that conveys connection between participants.*" In light of out model, gestural behavior and the production of multi-modal utterances can be used to express engagement on the *conversational* and *interactional* level, whereas spoken linguistic behavior or multi-modal utterances and the performance of collaborative actions are forms of expression for *interactional* and *task-level* engagement. As collaboration and thus engagement evolves over time perceived information needs to be integrated into some form of a memory playing a connecting role between the perception and generation processes. Different information needs to be collected and interpreted to account for the different levels of engagement.

On the task level, the agent needs to maintain a form of behavior that guides him to accomplish the overall goal of the collaboration, and that affects his decision process on how and when to interact. To appear thereby committed to mutual support, requires the agent to keep records of open tasks and task achievements.

Interactional engagement accounts for initiating a conversation, maintaining it, or even disengaging from it. To this end, the agent needs to maintain a model of the current conversational state including the last conversational functions, e.g. turn-taking functions, expressed by his interlocutor.

Concerning the conversational engagement, it is of particular importance to keep track of what the user is attending to and to link these percepts with verbal messages in order to establish joint, shared reference. As Nakano et al. [9] point out, to achieve engagement in a situated conversation, the agent "*should be able to be aware of the user's attention in the environment, and establish a communication channel by connecting the user's attention with the linguistic context of the conversation*". Likewise, Whittaker and colleagues [12, 13] claim that attention to a shared reference indicates that the interlocutor is engaged in a task as well as in a conversation. Knowing the communication partner's attention focus is thus a prerequisite to resolve references and interpret user actions in the appropriate context and to process engagement more generally.

3. Realization in Max

We now turn to the models and methods that are needed to meet the aforementioned requirements. In the following Section (3.1), we first focus on Max's perception and interpretation abilities enabling him to follow and manage the conversation. Concerning the measurement of the user's engagement in the current conversational process, Max needs to watch out for subtle engagement cues, revealing whether the user is paying attention and is following the conversational flow and whether the user wants to interact. In this context feedback and turn-taking signals play an important role which are processed by the the agent's perception module.

Max, who is obliged to assist the user, engages on the task level by adopting the goal, planning the requested assembly explanations, and then 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. As all of these aspects crucially depend on the interaction context, Max needs to build up two situational models:

- he needs to maintain a *task model* to find out whether and in what way a communicative move or environmental manipulation contributes to the overall goal;
- he needs to have a *user model* to relate the user's actions to her goals and intentions, which enables him to get an impression of the user's engagement at the task level.

The required models and their effects are described in the Sections (3.2) and (3.3). Other important cues that occur perpetually and in parallel with informational components are emotional aspects, which hint to the arousal and valence of engagement. They influence the way utterances are interpreted by the communication partner and provide some insight into her mood and attitude, which in turn are strongly related to the disposition and willingness to cooperate. These aspects are covered in Section (3.4). The way the agent's overall behavior and his decision process are influenced by all these aspects is then discussed in Section (3.5).

3.1. Perception

Max is supposed to engage in highly situated interactions and to recognize engagement cues ranging from task-level engagement to interactional and conversational engagement. He thus needs to pay close attention to the changing environment as well as to his human interaction partner.

Visual perception of the environment is emulated by virtual sensors attached to his eyes, simulating his point of view and calculating the virtual objects sighted in a view frustum. Information about the environment is needed to know which objects are present and which assemblies have been performed so far. In addition, the agent relies on information about the success or failure of construction actions emanating from the underlying assembly simulation system [3]. All this information helps Max to analyze and evaluate the human partner's actions and to keep his task model (see Section 3.2) up to date which thereby provides information needed to measure the user's engagement on the *task-level*.

Perception of the human partner in the real world is achieved by an infrared tracking system that detects the positions and orientations of the user's head and hands, as well as data gloves for gathering information about hand posture. The collected data are interpreted by detectors using the PrOSA framework [6]. Diverse detectors are realized using compute nodes which can be combined in hierarchically organized compute networks. The calculation of these real-world detectors runs in parallel with the perception of the virtual scene.

By focusing on subtle cues, these detectors concentrate, on the one hand, on calculating the engagement of the user on the *conversational level* and are responsible for discovering the user's focus of attention. These engagement cues are calculated via tracking the position and orientation of the user's head. The tracking results can then be used to estimate the user's gaze direction and the objects lying in her attention focus. Max is thus able to keep track of what the human partner is focusing at, e.g., when a verbal reference must be resolved.

On the other hand, detectors are used which are responsible for turn-taking signals. In the current implementation, detectors are available for signals of various turn-taking functions: *Wanting-Turn* (facing the agent and raising a hand), *Taking-Turn* (raising a hand and saying halt), and *Giving-Turn* (facing the agent, a metaphoric giving gesture, and spoken key words like "okay"). These detectors account for the gestural part of engagement on the *interactional level*.

The perception of spoken linguistic behavior is achieved by feeding the verbal input from the user into the agent's framework through a speech sensor that encapsulates a speech recognizer, which operates on a vocabulary appropriate for the Baufix construction scenario. The interpreted utterances can be used to additionally measure the *interactional* engagement of the user as rules are applied to recognize whether the user wants to initiate a conversation, maintain a conversation, transition the topic, or disengage from an interaction. Each dialog move of both Max and the user is memorized in a dialog history to allow for a temporal analysis of task-level and interactional engagement. Likewise, the perceptions about the user's behavior and her conversational engagement are fed into the user model to affect the agent's forthcoming interaction behavior.

3.2. Task Model

To collaborate with the user in a joint task and to be able to measure her engagement in the task, the agent needs to construct a *task model* which guides his supportive behavior in a goal-directed way. First of all, to create such a task model, the task must be known to the agent. If it not explicitly given, the agent must be able to perform some sort of intention recognition to find out the task. In our case, he needs to understand what the user wants to build which becomes clear either from explicit verbal statements of the user or implicitly from the sequence of assembly actions taken. Max disposes of long-term assembly knowledge that allows him to "conceptualize" the aggregates being constructed and to search for possible assembly goals along these lines. Thereupon, he constructs a task model by composing an assembly plan for the goal aggregate. This plan is suited to the situational context, e.g. the state of assembly and the concrete parts built in, and it is a shared plan in that it provides information about potential possibilities for collaboration.

Then, if Max and his human partner are working along a joint plan, the agent tries to relate every user action to the shared plan assuming a close engagement as a default. To keep the construction task as open as possible and to allow for flexible cooperation, Max uses an under-specified representation of individual construction steps. These steps are specified in terms of constraints, which result from the agent's construction knowledge or from the previous discourse and negotiation processes. By this means, the agent tries to analyze whether the user is engaged in the task by contributing to it in an active manner or is working on something else or even being counterproductive.

By way of refining single constraints, Max is able to adapt his plans to the ongoing joint task to be able to engage himself in a cooperative way. If he recognizes a conflict arising out of an assembly step performed by the user, he provides (partly emotional) feedback and informs his construction partner about the reason for the failure.

3.3. User model

The user model relies on both, perceptions of conversational and interactional engagement cues, and the deliberative analysis of user actions, thereby accounting also for engagement on the task level.

Concerning the interactional and conversational engagement the detected conversational functions and feedback signals the user has used are time-stamped and memorized in a stack-like manner to allow for retrospective analysis. Turn-taking gestures can be interpreted as indicative of engagement because the choice to take the turn presents an active way to participate in conversation. We want to use this information to calculate an engagement value which integrates the amount of engagement the user has shown over time and can be used to trigger interactional behaviors.

To deal with the user's focus of attention, a list of focused objects is acquired over time and is kept in a stack, carrying time-stamps. Furthermore, it should be possible to set an attention focus for certain time intervals to recognize whether the user has focused on a specific object during that time. This information can be used in the agent's decision process to recognize whether or when the conversational partner is paying attention or is unobservant. Additionally the information can be used for reference resolution.

With respect to the *task-level*, Max remembers the shared goals and plans that have been brought up and that were agreed upon. He also uses a simple form of intention recognition, trying to integrate the user's utterances and actions with context. If, for example, the shared goal consists of the construction of a propeller and the human partner puts a bolt into the middle hole of a bar, Max recognizes this as an attempt to build a rotor blade and thereby resolves the user's intention behind the construction step.

3.4. Emotions

Concerning the generation side of engagement, the display of emotions helps to make the agent appear more lifelike and engaged. Emotions reflect the status and progress of the agent's inner processes in the sense that they represent records of success and failure in goal attainment and thereby reveal the agent's involvement and engagement in a task. Max comprises of a dynamic emotional system [1] which simulates his emotional state and its continuous evolvement over time. It modulates his facial expression, speech, secondary behaviors, as well as his cognitive functions. On the one hand, the current discrete emotion category, modulated by a continuous intensity value, is used to trigger emotional expressions of the face as well as to influence deliberative reasoning in the cognitive architecture. On the other hand, involuntary facets of Max's observable behavior e.g., his simulated breathing rate, eye blink frequency, as well as pitch level and rate of speech are modulated by the continuous emotional values like valence and intensity of arousal.

Having a symmetric system in which the human partner's emotional state is perceived and analyzed by Max as well would make the bonds between the interactants much closer. While work on emotion recognition is underway, right now we have only started to scan the user's utterances for verbal expressions that may count as emotional cues. These results will then be provided to and used in Max's decision processes.

3.5. Behavior Generation

The agent needs to adapt his behavior to the actual as well as the desired state of his own engagement, to that of his task partner, and to the situational context. In general, the engagement cues Max is supposed to provide in support of the joint task are inspired by the rules proposed in [10]. To create the prerequisites for conversation, the agent has to account for the following *interactional engagement* rules which are responsible for initiating and ending a conversation in socially accepted ways:

- to gaze at the human partner at first sight to engage her to interact, possibly followed by a conversational greeting
- to notice but not disengage when the human either fails to take the turn to speak and to possibly encourage her once more
- to ascertain, via dialog, the human's desire to disengage when the human is not watching Max or the scene for a long period of time and fails to take the turn
- to disengage quickly when the conversational partner desires to end the conversation
- to end the conversation with a normal conversational closing when there are no more overall goals left to discuss.

Once the interlocutors have started a conversation, the following rules need to be applied in order to evoke some form of *conversation engagement*.

- to briefly look away, when beginning to speak
- once the human has responded to the initiation of engagement, to look at the human partner when she takes the turn to speak in the conversation,
- to mainly look at the conversational partner or at the objects being referred to during the speaking turn

To know when to apply these rules, the agent relies on the perceptions and models described in the previous sections. The implementation of the different engagement cues poses several demands on the agent's underlying architecture. The architecture needs to support parallel, concurrent processing and generation of behavior [8] at different time scales.

Concerning the generation of *interactional* engagement cues, Max is able to produce synchronized multi-modal utterances [5]. He also uses turn-taking behaviors to inform the human partner about his intention to actively contribute. Upon deciding to show a turn-wanting behavior, reactive behaviors are instantiated that automatically adapt to the environment, for example, gazing at the user and tracking her if she moves. On the other hand, when Max detects a *Taking-Turn* or *Wanting-Turn* signal, the deliberative plan in charge of interpreting it, must take into account the mental state of the agent, the goal he pursues, and the dominance relationship between the interlocutors accounting for the agent's own current engagement. For example, depending on the performative Max is using, it is more or less likely that he will allow an immediate interruption even before he has analyzed the content of the interlocutors interaction move. To handle such situations, a turn-taking model is used [7] that combines concepts of the FMTB architecture [2] with dialog management layers as proposed in [11]. It consists of two steps: First, a rule-based, context-free evaluation of the possible turn-taking reactions taking into account the current conversational state and the performative of the user's utterance. These rules are incessantly applied and integrated using data-driven plans, and they aim to ensure cooperative dialog behavior. The second step is the context-dependent decision upon different response plans, possibly leading to new intentions.

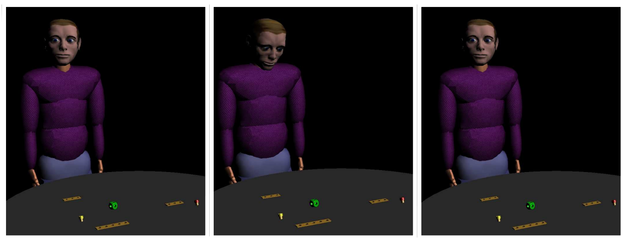
Additionally and as a form of *conversational engagement*, Max uses gaze and glance gestures to lead the user's attention to certain objects. In important cases, he also uses pointing gestures to make sure his construction partner is able to resolve a reference correctly. These glance gestures are produced when Max is the speaker and is referring himself to a certain object, but also when the user is talking and Max has been able to resolve the user's reference. In this case, the glance behavior can then be seen as a form of providing immediate feedback.

We consider reaction time to be an important measurement of engagement. Max has to notice how long the user needs until she responds to his efforts, and he may encourage her if she does not react at all. To this end, the interaction model underlying Max's deliberative behavior takes explicit account of the agent's expectations about how the user is obliged to react. If Max asks a question, he expects an answer or reaction in some way; if he requests the user to perform a certain action, he expects that the user either performs the action, or otherwise informs him that he is not willing or able to do so. These conversational obligations ensure a natural coherent interaction, because Max will insist on a reaction if he does not get one. If that does not help, he may even abort the conversation. The time he leaves the human partner to react should thereby depend on the contextual factors like the mood of the agent as well as the importance of the conversational goal. Additionally, if Max notices that the user shows engagement cues while thinking about an answer, he should be more patient than when he does not notice any such cues.

In addition to the bonds between interactants that evolve through face-to-face engagement, personal bonds can be created by aligning the vocabulary and the utterances with the conversational partner's preferences. Therefore Max adapts in his utterances to the synonyms the communication partner uses.

4. Examples of Engagement

In this section we want to present some examples showing how Max, on the one hand, tries to signal engagement, and on the other hand, how he perceives and evaluates the user's engagement.



User: We could insert this bolt [*gazes at right bolt*] into a bar.
Max: [*gazes at right bolt, then focuses on user*] Okay.

Figure 3. Reference resolution

The first two examples show the perception and generating of *conversational engagement* cues. The first example (Fig. 3) presents how reference resolution needs to account for the user's gaze behavior and how Max shows his engagement and inner thoughts by focusing on the referenced object the user proposes to use. By this means, the user can get an impression whether Max has resolved the reference correctly. In addition, Max appears engaged during the time he thinks about the proposal because he shows interest in the referenced object.

User: Insert the blue bolt...
Max: [*gazes at blue bolts*]
User: ... into the middle hole of a bar!
Max: [*nods*]
User: And then turn the bars crosswise.

Figure 4. Immediate feedback during referenced resolution

The next example (Fig. 4) accounts for the production of engagement cues while listening, and thereby revealing information about inner ongoing processes and showing the interlocutor that one is still engaged in the conversation. The first glance gesture produced by Max indicates, that he has resolved the reference. The second feedback signal, the nodding gesture, goes even a step further and expresses that he has been able to resolve the complete instruction. But as the user is still keeping the turn, he keeps on listing and waits to perform the assembly until the user has finished her turn.

The third example (Fig. 5) shows how Max handles *interactional engagement*. It deals with a situation in which Max notices the user's disengagement. To recognize that the user is unobservant, Max's user model is consulted which includes a history of the user's conversational and interactional engagement cues. In the beginning, Max tries to encourage the user to re-engage. But then after a while, he disengages himself following the interactional engagement rules of social norms.

Max: Insert this bolt into a bar.
User: [*gazes in opposite direction (neither at Max, nor at the objects)*]
Max: [*strong turn-taking gesture, repeats*] Insert this bolt into a bar.
User: [*no reaction*]
Max: [*turn-giving gesture, waits*] Okay then we leave it like that. Bye!

Figure 5. Disengaging from a conversation

The last example (Fig. 6) accounts for *task-level engagement*. It is shown, how Max agrees to engage in the task of building a propeller. He builds up a task model, and explains the first assembly step which needs to be performed. As the user proposes an object, he relates the utterance to the situational context and interprets it as engaging in the task. He agrees and engages himself in the task by performing the construction step.

User: Let us build a propeller.
Max: Okay, let's do it together. We need to insert a bolt in the middle of a three-hole-bar.
User: The yellow one?
Max: Okay [inserts the yellow bolt into a bar].

Figure 6. Engagement in construction tasks

---

5. Conclusion and Future Work

We have presented our work towards the development of an engaging and engaged collaboration partner. To address the issue of different forms of engagement, we have presented an analytical model to act as a foundation for our discussions. The model provides a basic structure to which engagement phenomena can be allocated, covering different skills of engagement. The implementation of these skills shows first promising behavior. Still, a lot of work remains to be done to make the agent appear convincing and equal, but first steps have been taken to successfully cope with aspects of engagement. Future extensions will cover the detection of more gestural cues and the consideration and investigation of appropriate time frames in which the engaged conversational partner has to react to account for the conversational obligations imposed by communicative acts. For example, an engaged and cooperative user will not hesitate to answer a question Max has asked.

Finally, and very importantly, we are planning evaluation studies to find out how well Max is perceived and accepted as a conversational partner and a first evaluation study on the social interactions with Max in a public museum is underway. In this context, our proposed engagement model can be used as an analytical basis to provide lines along which the measurement and classification of engagement behavior can be performed. Certainly, the results from such studies will be used to gradually adjust and enhance our models, ultimately yielding a better understanding of the underlying mechanisms of engagement in collaborative tasks.

Acknowledgments

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 multimodal conversational agent. In *Proceedings of Affective Dialogue Systems: Tutorial and research workshop (ADS 2004)*, LNCS 3068, pages 154–165, Berlin Heidelberg, 2004. Springer.

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

[3] B. Jung and I. Wachsmuth. Integration of geometric and conceptual reasoning for interacting with virtual environments. In *Proc. AAAI'98 Spring Symposium on Multimodal Reasoning*, pages 22–27, 1998.

[4] S. Kopp, B. Jung, N. Lessmann, and I. Wachsmuth:. Max - A Multimodal Assistant in Virtual Reality Construction. *KI-Kü nstliche Intelligenz*, 4/03:11–17, 2003.

[5] S. Kopp and I. Wachsmuth. Synthesizing multimodal utterances for conversational agents. *Computer Animation and Virtual Worlds*, 15(1):39–52, 2004.

[6] M. Latoschik. A Gesture Processing Framework for Multimodal Interaction in Virtual Reality. In *Afrigraph 2001, 1st*

International Conference on Computer Graphics, Virtual Reality and Visualization in Africa, 5 - 7 November 2001, 2001.

[7] N. Leßmann, A. Kranstedt, and I. Wachsmuth. Towards a Cognitively Motivated Processing of Turn-Taking Signals for the Embodied Conversational Agent Max. In *AAMAS 2004 Workshop Proceedings: "Embodied Conversational Agents: Balanced Perception and Action"*, 2004.

[8] N. Leßmann and I. Wachsmuth. A Cognitively Motivated Architecture for an Anthropomorphic Artificial Communicator. In *Proceedings of the Fifth International Conference on Cognitive Modeling (ICCM-5 2003), Bamberg, April 2003*, pages 277–278, 2003.

[9] Y. I. Nakano and T. Nishida. Awareness of perceived world and conversational engagement by conversational agents. In *Conversational Informatics for Supporting Social Intelligence & Interaction – Situational and environmental information enforcing involvement in conversation*. AISB 2005 Symposium, April 2005.

[10] C. L. Sidner, C. Lee, and C. Kidd. Engagement during dialogues with robots. In *Conversational Informatics for Supporting Social Intelligence & Interaction – Situational and environmental information enforcing involvement in conversation*. AISB 2005 Symposium, April 2005.

[11] 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 Multiagent systems AAMAS02*, pages 766–773, 2002.

[12] S. Whittaker. Theories and methods in mediated communication. In A.Graesser, M. Gernsbacher, S., and Goldman, editors, *The Handbook of Discourse Processes*. Erlbaum, NJ, 2003.

[13] S. Whittaker and B. O'Conaill. Evaluating videoconferencing. In *Proceedings of CHI'93 Human Factors in Computing Systems*. ACM Press, 1993.