Functions of Speaking and Acting - An Interaction Model for Collaborative Construction Tasks

Stefan Kopp & Nadine Pfeiffer-Leßmann

Artificial Intelligence Group
Faculty of Technology, Bielefeld University
{skopp, nlessman}@techfak.uni-bielefeld.de

Abstract. This paper describes how a virtual agent assists a huam user in collaborative construction tasks by combining manipulative capabilities for assembly actions with conversational capabilities for mixed-initiative dialogue. We present an interaction model for representing the evolving information states of the participants in this interaction. It includes multiple dimensions along which an interaction move in general can be functional, independent of the concrete communicative or manipulative behaviors comprised. These functions are used in our model to interpret and plan the contributions that both interactants make.

1 Introduction

Virtual humanoid agents offer an exciting potential for interactive Virtual Reality (VR). Cohabiting a virtual environment with their human users, virtual agents may ultimately appear as equal partners that share the very situation with the user and can collaborate on or assist in any task to be carried out. To investigate such a scenario the embodied agent *Max* (Kopp, Jung, Leßmann, Wachsmuth, 2003) is visualized in human-size in a CAVE-like VR environment where he joins a human user in assembling complex aggregates out of virtual Baufix parts, a toy construction kit (see Fig. 1). The two interactants meet face-to-face over a table with a number of parts on it. The human user—who is equipped with stereo glasses, data gloves, optical position trackers, and a microphone—can

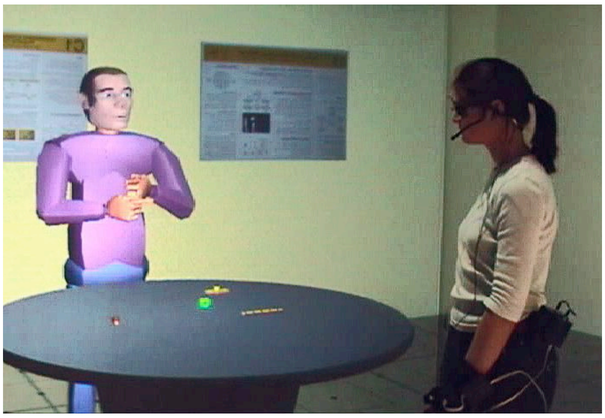

Figure 1: In his CAVE-like Virtual Reality environment, Max guides the human partner through interactive construction procedures.

issue natural language commands along with coverbal gestures or can directly grasp and manipulate the 3D Baufix models to carry out assembly actions. Further, the user can address *Max* in natural language and gesture. The agent is, likewise, able to initiate assembly actions or to engage in multimodal dialogue using prosodic speech, gesture, eye gaze, and emotional facial expressions.

In this setting, the two interactants can become collaboration partners in a situated interaction task as follows; see (Leßmann, Kopp & Wachsmuth 2006) for a detailed analysis. The user wants (or is) to construct a certain Baufix aggregate (e.g. a propeller) and is free to directly carry out assembly steps, either if they are known to her or if she wants to give certain assembly procedures a trial. At any time, she can ask *Max,* who shares the situation with her and attends her actions, for assistance. *Max*'s task is then to collaborate with the human user in jointly walking step-by-step through the construction procedure and to provide support whenever his partner does not know how to go on constructing the aggregate.

While the overall interaction is guided by the user's wish to build a certain assembly, once the human and Max have engaged in the collaborative construction activity the scenario is symmetric in that roles may flexibly switch between the interaction partners according to their competences. That is, either the human or Max may carry out an action or may instruct the other one to perform an action. This demands for the agent being able, both to collaborate by taking actions in the world and conversing about it, possibly in an intertwined manner. The scenario is hence characterized by a high degree of interactivity, with frequent switches of initiative and roles of the interactants. The participants hence need to reconcile their contributions to the interaction with the need for regulating it by performing multiple behaviors simultaneously, asynchronously, and in multiple modalities. That is, their multimodal contributions are multi-functional, either communicative or manipulative, and the effects of the latter can only be taken in from the situational context. We subsume each of those contributions that Max or the user can perform under the term *interaction move* (as an extension of the common term *dialogue move*).

Enabling *Max* to engage in such an interaction requires to embed him tightly in the situational context and, based on the perception of the user‘s interaction moves, to reason about what the proper next interaction move may be right now. The result of this deliberation process is passed on to modules that generate the required behaviors (Leßmann, Kopp & Wachsmuth 2006). This general layout resembles the SAIBA pipeline architecture (Kopp et al. 2006), with dedicated representations to interface (1) between the planning of an interaction move and its behavioral realization, and (2) between behavior planning and realization. In the remainde of this paper, we described the first representation as part of an information state-based interaction model that attempts to explicate all the functional aspects of an interaction move which must be taken into account in order to keep track of the interaction and adequately specify an interaction move independent of its behavioral realization.

2 Interaction Model

Adopting the information state approach to dialogue (Traum & Larsson, 2003), we frame an interaction model that defines the aspects along which the collaborative interaction evolves from the agent's point of view. This model stipulates what facets of interaction Max is to take into account, without making any provisions as to how these aspects can be fully represented or reasoned about.

2.1 Information state-based interaction model

The information state approach to dialogue modelling provides a framework for formalizing both, state-based/ frame-based as well as agent-based dialogue models in a unified manner (Traum & Larsson, 2003). It assumes that each dialogue particpant maintains information states (IS) that are employed in deciding on next actions and are updated in effect of dialogue acts performed by either interactant. A particular dialogue manager, then, consists in a formal representation of the contents of the ISs plus update processes that map from IS to IS given certain dialogue moves. Several systems have been based on this framework, concentrating on different aspects of dialogue, e.g., the GODIS system (Larsson et al. 2000), or in the WITAS project (Lemon et al. 2001). Traum & Rickel (2002) have proposed a model of multimodal dialogues in immersive virtual worlds that comprises layers for *contact*, *attention*, *conversation*, *obligations*, and *negotiation*. The *conversation* layer further defines separate dialogue episodes in terms of *participants*, *turn*, *initiative*, *grounding*, *topic*, and *rhetorical connections* to other episodes. Rules state how communicative signals can be recognized and selected to cause well-defined changes to a layer.

However, most existing models have focused either on dialogues where the agents are only planning—with the plan to be executed at a later time—or on dialogues where the agents are only executing some previously created plan. As Blaylock et al. (2003) point out, this does not allow for modeling dialogues where mixed-initiative collaboration and interleaved acting and planning take place, as in our setting. In addition, we posit that not only spoken communicative acts but also manipulative actions must be characterized as interaction moves. We therefore introduce a model that includes a layer accounting for the formation, maintenance, overlap, and rejection of the goals of interactants. Goals cover the rational behind any kind of intentional action and abstract away from their situational realization. The interaction model consists of the following layers:

- *Initiative*: Who has brought up the goal that the interactants are pursuing in the current discourse segment.
- *Turn*: Who has the turn. We distinguish between four states: *my-turn, others-turn* (or a unique name for a specific interaction partner, respectively), *gap, overlap*.
- *Goals*: The goals that have been pursued so far as well as the, possibly nested, goals that are still on the agent's agenda for the remainder of the interaction. Each goal may either have arisen from the agent's own desires, or was induced due to obligations following social norms or a power relation Max is committed to (Traum & Allen 1994).
- *Content*: The propositional content that has been or will be subject of the discourse, defined in a logic-based notation.

- *Grounding*: The discourse state of content facts, denoting whether the agent assumes a fact to be new to the conversant or part of the common ground (Clark & Brennan, 1991).
- *Discourse structure*: The organization of the discourse in segments that can relate to goals and refer to or group the entries in the content and goal layers. Each discourse segment has a purpose (DSP; Grosz & Sidner 1986) and they are related based on the relationships among their DSPs, all of which are part of one intentional structure.
- *Partner model*: What is known about the dialogue partner(s), also covering aspects of negotiation and obligations. It is the basis of retrospective analysis and thus plays a central role for the agent being situated in an evolving interaction.

2.2 Interaction moves

Any action by an interactant may alter the agent's internal representation of the state of interaction. We focus here on intentional actions, either communicative or manipulative, which make up for most of the progress of interaction. They are selected and performed in order to achieve a desired state of the world. In mental attitudes terms, we think of them as being produced following specific kinds of nested intentions, which in turn are resultant of some goal-directed planning process. For example, when touching a bolt or saying "move the bolt closer to the bar", one might have the intention not just of manipulating the object but of mating it with a bar, building a propeller, and assisting the partner.

Modelling such behaviors requires having an account of their functional aspects, in addition to their mere overt actions. For manipulative acts, these aspects can easily be defined as the direct consequences (or post-conditions) of the manipulation in the world, which in turn need to be related to the current intention structure. For communicative acts, the functional aspects are not so easy to pinpoint despite of a long tradition of viewing speaking as acting. Austin (1962) pointed out multiple acts (or "forces") of making a single utterance: the locutionary act of uttering words, and the perlocutionary act of having some effects on the listeners, possibly even effecting some change in the world. To its perlocutionary end, an utterance always performs a certain kind of action, the illocutionary act. Adopting this notion, Searle (1969) coined the term *speech act* which is supposed to include an attitude and a propositional content. Other researchers have used terms like *communicative acts* (Allwood, 1976; Poggi & Pelachaud, 2000), *conversational moves* (Carletta et al., 1997), or *dialogue moves* (Cooper et al., 1999) for related ideas.

The notion of functional aspects of communicative actions is particularly beneficial to multimodal systems, for it allows abstracting away from a signal's overt form to core aspects that only got realized in certain ways (Cassell et al. 2000; Traum & Rickel 2002). In Cassell et al.'s model, the function a behavior fulfils is either propositional (meaning-bearing) or interactional (regulative), and several behaviours are frequently employed at once in order to pursue both facets of discourse in parallel. Poggi & Pelachaud (2000) define a *communicate act*, the minimal unit of communication, as a pair of a signal and a meaning. The meaning includes the propositional *content* conveyed, along with a *performative* that represents the action the act performs (e.g. request, inform, etc.).

Based on the interaction model layed out above, we define interaction moves as the basic units of interaction in terms of the following slots:

- *Action*: The illuocutionary act the move performs. The act can either be purely *manipulative* (connect, disconnect, take, or rotate) or *communicative*. In the latter case, a performative encodes the kind of action as described below.
- *Goal*: The perlocutionary force of the move, i.e., what the move is meant to accomplish. This can be either a desired state of the world (*achieve* something), or it can be the mere performance of an action (*perform* something).
- *Content*: Information conveyed by the move, needed to further specify the action and the goal. This can accommodate either propositional specifications (e.g. for language or symbolic gestures) but also an analog, quantitative representation of imagistic content (e.g. for iconic gestures).
- *Surface form*: The entirety of the move's overt verbal and nonverbal behaviors, employed to convey all of the aspects represented here.
- *Turn-taking*: The function of the move with respect to turn-taking, either *take-turn*, *want-turn*, *yield-turn*, *give-turn*, or *keep-turn*.
- *Discourse function*: The function of the move with respect to the segmental discourse structure, either *start-segment*, *contribute*, or *close-segment* (cf. Lochbaum, Grosz, Sidner, 1999).
- *Agent*: The agent that performs the move.
- *Addressee*: The addressee(s) of the move.

These slots together capture the informational aspects that are relevant about an action in our model of interaction. They are not independent from each other, nor are they self-sufficient. Instead, the slots are supposed to mark the specific signification of particular informational aspects. In general, they provide a frame structure that can be incrementally filled when generating an action through subsequent content planning and behavior planning. Some of the slots may thereby remain empty, e.g., for smallish moves like raising a hand which may solely fulfil a turn-taking function.

2.3 Semantics – Expectations – Obligations

The meaning of an interaction move is defined by what it realizes, i.e. by the effects it brings about in terms of the aspects captured by the interaction model. For a manipulative move, the effects are directly implied by the action and can be extracted via perception of the environment. Given powerful enough perceptive capabilities, the current scene thus serves as an external part of the representation of the interaction state for the agent.

The effects of communicative moves can often not be defined in a clear-cut way as they depend on multiple aspects of the interaction move and the context in which it is performed. For example, even a simple move that fulfils the turn-taking function *want-turn* will result in a new state *my-turn* only when executed in the state *gap* (no one has the turn). For this reason, information state-based systems typically employ a large number of update rules to model the context-sensitive effects of particular moves. Furthermore, the effects of a communicative move depend on social attitudes like the expectations a sender connects to a move or the obligations it imposes on the recipient. Traum (1996) argues that *obligations* guide the agent's behavior, without the need for recognizing a goal behind an incoming move, and enable the use of shared plans at the discourse level. As *Max* is supposed to be cooperative, obligations are therefore modelled to directly lead to the instantiation of a *perform*-goal in response to an interaction move. If this move was a *query* or *request*, Max will thus be conducting the action asked for in a reactive manner. In case of a *proposal*, he is only obliged to address the request, and his further deliberation decides upon how to react. We hence explicitly encode the *Action* of each move by distnguishing between four types of performative (e.g. cf. Poggi & Pelachaud 2000):

1. *inform-performatives*: provide informational facts, characterized by the desire to change the addressee's beliefs
2. *query-performatives*: procure informational content to establish a new belief or to verify an existing one
3. *request-performatives*: request a manipulative action
4. *propose-performatives*: propose propositional content or an action

Derived from one of these general types, a performative can often be narrowed down through subsequent specification during interpretation or generation (Poggi & Pelachaud 2000). The final performative will be tailored to the content on which it operates (e.g., whether it asserts or requests a propositional fact, an action, or an opinion) as well as to contextual factors like the actual situation, the addressee, the type of encounter, or the degree of certainty. This can even happen retrospectively, when the performative of a previous move is fully disclosed at a later point in the conversation. We represent these refinements using a compositional notation, e.g. *inform.agree* or *propose.action*.

See (Leßmann, Kopp & Wachsmuth 2006) for a description of how the different *performatives* are endowed with semantics, obligations and expectations, and how they are used along with the *discourse function* of a move to state rules for identifying the holder of the *initiative*. To illustrate our model here, we analyse in Table 1 some interaction moves of an example dialogue. To represent propositional facts in the *goal* and *content* slots, we use a formal, logic-like notation in which predicates/relations are indicated by a capital letter, and unbound variables are prefixed with a $. For example, "(Inform.ref $s)" represents that a reference will be communicated, with the unbound variable $s being the referent. An unbound variable indicates underspecification inherent to the content of the move (in this example, there is no particular entity being referred to, yielding an indefinite article). Note further that the *content* and the *goal* slot together specify the informational aspects of the move, with some of these aspects marked as being the goal.

	MAX	USER	MAX
Interaction Move	"*Let us build a propeller.*"	"*Ok*."	"*First, insert a bolt in the middle of a bar.*"
Action	propose.action	inform.agree	request.order
Goal	(Achieve (Exists prop))	(Perform (Inform.agree))	(Achieve (Connected $s $b $p1 $p2))
Content	(Build prop we)	(Build prop we)	(Connect $s $b $p1 $p2) (Inst $s bolt) (Inst $b bar) (Center_hole $b $p2))
Surface form	<words>$_t$	<words>$_t$	<words>$_t$

Turn-taking	take | give	take | keep	| give
Discourse function	start-segment (DSP=prop)	contribute	start-segment (DSP=prop-s1)
Agent	User	Max	Max
Addressee	Max	User	User

USER	MAX	USER	MAX
"*Which bolt?*"	"*Any bolt*."	User puts bolt into the first hole of bar.	"*No, that was the wrong hole*."
query.ref	inform.ref	connect	inform.disagree
(Perform (Query.ref $s))	(Perform (Inform.ref $s))	(Achieve (Connected bolt-2 bar-1 port-1 port-3))	(Perform (Inform.disagree (Connect …)))
(Inst $s bolt)	(Inst $s bolt)	(Connect bolt-2 bar-1 port-1 port-3)	(Not (Center-hole bar-1 port-3))
<words>$_t$	<words>$_t$	<manipulation>	<words>$_t$
take | give	take | yield	take | yield	take | yield
contribute	contribute	contribute	contribute
User	Max	User	Max
Max	User		User

Table 1: An example dialogue analyzed into formal interaction moves.

3 Modeling situated interaction management

In this section we give a brief overview of how the functional aspects explicated in the move representation inform interaction management. Interaction moves are the main data structure for interfacing in Max's architecture between modules for perception, deliberation, and behavior planning. On the input side, interaction moves are used to specify and structure the incoming information, possibly relating the information to external objects or previous interaction moves; on the output side, they serve as a container which gets filled during the generation process of an agent's response. How the agent behaves in the interaction is determined by a Belief-Desire-Intention control architecture (Bratman, 1987; Rao & Georgeff, 1991), for which we use JAM/UM-PRS (Huber 1999, Lee et at. 1994). It draws upon specific plans for analysing input moves as well as generating responses in a context-dependent manner.

Plans can either directly trigger specific behaviors to act, but may also invoke dynamic, self-contained planners that construct context-dependent actions or, again, plans. A judicious use of plans allows the agent to reduce the complexity of controlling dynamic behavior and to constrain itself to work towards goals. Plans are therefore kept as general as possible and are refined to situational context not until necessary. That is, we too regard plans as mental attitudes, as opposed to recipes that just consist of the knowledge about which actions might help achieve a goal (cf. Pollack 1992, Pollack 1990). When Max selects a plan and binds or constrains its arguments, it becomes an intention on an intention stack, encompassing information about the context and the goals responsible for it to come into existence. In result, the plan is characterized by the agent's attitudes towards the realization of his goal.

3.1 Dealing with incoming interaction moves

A variety of plans are used for handling incoming interaction moves. Turn-taking functions are processed taking into account the mental state of the agent, the goal he pursues, and the dominance relationship between the interlocutors. A turn-taking model (Leßmann et al. 2004) is used that consists of two steps: First, a rule-based, context-free evaluation of the possible turn-taking reactions takes into account the current conversational state and the action of the user utterance. These rules are incessantly applied and integrated using data-driven conclude plans to ensure cooperative dialogue behavior. The second step is the context-dependent decision upon different response plans, possibly leading to new intentions.

The propositional content of an incoming interaction move, if present, is processed by plans for determining the context it belongs to (e.g. to resolve anaphora). To this end, it is checked whether the move relates to the current goals, or to an interaction move performed before. This is done by calculating a correlation value between the content facts carried by the interaction move and the goal context. In addition, the agent needs also to be able to resolve references to external objects in the virtual world, which is achieved using a constraint satisfaction-based algorithm (Pfeiffer & Latoschik 2004).

If the agent succeeds in searching a candidate context, it adds an *obligation*-goal to handle the interaction move as a sub-goal to the goal to which the move contributes to. Otherwise, a new *obligation*-goal is added as a top-level goal. By associating the interaction move with one of his goals, the agent is able to deal with multiple threads at the same time and keep apart their individual contexts. In result, incoming information is structured according to its content, but also depending on the context—an important aspect of being situated in the ongoing interaction. Finally, different plans are thus used to handle an incoming move depending on the *action* it performs. For example, these plans may include verifying a proposition, answering a question, or constraining a parameter involved in a plan in order to adapt it to the events occurring during the plan execution e.g. the usage of specific objects or proposals made by the interlocutor.

3.2 Output planning

If the agent has the goal to achieve a certain state of the world he will reason about which courses of action may lead to this goal. In result, he will formulate and aggregate situated plans that lead to the desired state of affairs. Each behavior that any of these plans incorporates may be either a communicative or a manipulative move, both represented as interaction move in terms of the same features defined above. Both kinds of behaviors hence stand in a meaningful relationship to each other and can be carried out in a nested way when suitable. The decision is based upon the role that the agent currently fulfills. If Max is to act as the instructor—his initial role upon being asked by the human—he will verbalize the steps.

When Max has decided to employ a communicative act, he has to decide what meaning should be conveyed (content planning) and how to convey it in natural language, gesture, etc (behavior planning). In general, the behavior produced must be tailored to the agent's current role in the collaboration, expressing his beliefs and goals. Our approach to natural language generation starts with a communicative goal to be achieved and relies on various sources of knowledge, including task and conversational competencies, a model of the addressee, and a discourse history (McTear 2002; Reiter & Dale 2000). Crucially, this process is an integral part of the agent's plan-based deliberations and is carried out naturally by dedicated plans when Max's mental attitudes decline him towards making a verbal utterance.

The communicative goal derives directly from the currently active intention on the agent's plan stack, e.g. to request the interaction partner to connect two objects: *request.order "user" "connect" "SCHRAUBE_23" $obj3*. The goal of the generation process, then, is to determine all information about this communicative move needed to render it as a multimodal utterance. Starting from a message (*goal*) as above, the general performative (*action*) is first derived directly from the type of the intended act (in our example *request.order*). Content selection then works to determine the information that specifies the parameters of the communicative act to concrete values and, if possible, refines the *performative*. Discourse planning determines the move's *discourse function* and the *discourse segment* (DSP) it is contributing to, both being derivable from the actual structure and ordering of the plans on the intention stack (cf. Lochbaum, Grosz, Sidner, 1999).

The final stages of output generation, not presented here, are multimodal behavior planning and realization; see (Leßmann, Kopp & Wachsmuth 2006) for details on the methods used in the present scenario.

4 Summary

We have presented an approach to formally model the moves that interactants take in a mixed-initiative, collaborative construction scenario. This scenario provides both the virtual agent and the user with rich possibilities for interaction, comprising praxic actios and conversing about it as equal contributions. We have defined a formal model that explicates and characterizes dimensions along which these situated interactions evolve, and which an embodied agent must actively manage in order to be able participate as a collaborative expert. Follwing an information state-based approach, we have further laid down in detail the notion of interaction moves that encapsulate the different functions a contribution can fulfill with respect to the dimenions of this conceptual framework. As these functions, among others not considered here (e.g. emotional display or epistemic qualification), bear significant influence on the behavioral realization of an interaction move, FML should be able to accommodate most or all of them.

References

Allwood, J. (1976). Linguistic Communication in Action and Co-Operation: A Study in Pragmatics. In *Gothenburg Monographs in Linguistics 2*, University of Gothenburg, Dept. of Linguistics.

Austin, J. L. (1962) *How to Do Things with Words*. In *Havard University Press*, Cambridge, MA.

Blaylock, N., Allen, J., Ferguson, G. (2003). Managing communicative intentions with collaborative problem solving. In Jan van Kuppevelt and Ronnie W. Smith, editors, *Current and New Directions in Discourse and Dialogue*, volume 22 of *Kluwer Series on Text, Speech and Language Technology*, pages 63-84. Kluwer, Dordrecht

Bratman, M. E., (1987). *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA

Carletta. J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., Anderson, A. (1997) The reliability of a dialogue structure coding scheme. Computational Linguistics 23:13-31.

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.

Clark, H. H. & Brennan, S. A. (1991). Grounding in communication. In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.). Perspectives on socially shared cognition. Washington: APA Books.

Cooper, R., Larsson, S., Matheson, Poesio, M., Traum, D. (1999) Coding Instructional Dialogue for Information States. Trindi Project Deliverable D1.1

Grosz, B. J., Sidner, C. L., (1986). Attention, Intentions, and the Structure of Discourse. In *Computational Linguistics,* Volume 12, Number 3, pp. 175-204: The MIT Press

Huber, M.J (1999). JAM: A BDI-theoretic mobile agent architecture. *Proceedings Third Int. Conference on Autonomous Agents*, pp. 236-243.

Kopp, S., Jung, B., Leßmann, N., Wachsmuth. I. (2003). Max – A Multimodal Assistant in Virtual Reality Construction. *KI-Künstliche Intelligenz* 4/03, pp 11-17

Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thorisson, K., Vilhjalmsson, H. (2006) Towards a common framework for multimodal generation in ECAs: The behavior markup language. Gratch, J. et al. (eds.): Intelligent Virtual Agents 2006, LNAI 4133, pp 205-217, Springer.

Larsson, S., Ljunglöf, P., Cooper, R., Engdahl, E., Ericsson, S., (2000). GoDiS - An Accommodating Dialogue System. In *Proceedings of ANLP/NAACL-2000 Workshop on Conversational Systems*, pp 7-10.

Lee, J., Huber, M.J., Kenny, P.G., Durfee, E. H. (1994). UM-PRS: An Implementation of the Procedural Reasoning System for Multirobot Applications. *Conference on Intelligent Robotics in Field, Factory, Service, and Space (CIRFFSS),* Houston, Texas, pp 842-849.

Lemon, O., Bracy, A., Gruenstein, A., Peters, S. (2001) Information States in a Multi-modal Dialogue System for Human-Robot Conversation. In Proceedings *Bi-Dialog, 5th Workshop on Formal Semantics and Pragmatics of Dialogue*, pp 57 - 67

Leßmann, N., Kranstedt, A., Wachsmuth, I. (2004). Towards a Cognitively Motivated Processing of Turn-taking Signals for the Embodied Conversational Agent Max. *AAMAS 2004 Workshop Proceedings: "Embodied Conversational Agents: Balanced Perception and Action"*, pp. 57-64.

Leßmann, N., Kopp, S., Wachsmuth, I. (2006) Situated Interaction with a Virtual Human - Perception, Action, and Cognition. Rickheit, G., Wachsmuth, I. (eds.): Situated Communication, pp. 287-323, Mouton de Gruyter.

Lochbaum, K., Grosz, B. J., Sidner, C., (1999) Discourse Structure and Intention Recognition. In *A Handbook of Natural Language Processing: Techniques and Applications for the Procressing of Language as Text*. R. Dale, H. Moisl, and H. Sommers (Eds.)

McTear, M. (2002). Spoken Dialogue Technology: Enabling the Conversational User Interface. In *ACM Computer Surveys* 34(1), pp. 90-169,

Pfeiffer, T. & Latoschik, M., E. (2004). Resolving Object References in Multimodal Dialogues for Immersive Virtual Environments. In Y. Ikei et al. (eds.): *Proceedings of the IEEE Virtual Reality 2004*. Illinois: Chicago.

Poggi, I. & Pelachaud, C. (2000). *Performative Facial Expressions in Animated Faces*. . In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied Conversational Agents*, pp. 29-63. Cambridge: The MIT Press.

Pollack, M. E., (1990). Plans as Complex Mental Attitudes, P.R. Cohen, J. Morgan, and M. E. Pollack, eds., *Intention in Communication*, MIT Press, 1990

Pollack, M. E. (1992). The Use of Plans. In *Artificial Intelligence*, 57(1), pp.43-68

Rao, A. & Georgeff, M. (1991). Modeling rational behavior within a BDI-architecture. In *Proceedings Int. Conference on Principles of Knowledge Representation and Planning,* pp. 473-484.

Reiter, E. & Dale. R. (2000). *Building Natural Language Generation Systems*, Cambridge University Press

Searle, J.R. (1969). *Speech Acts: An essay in the philosophy of language*. In Cambridge University Press

Traum, D. R. & Allen, F. A. (1994). Discourse Obligations in Dialogue Processing. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics* (ACL-94), pp. 1-8

Traum, D. R. (1996). Conversational Agency: The TRAINS-93 Dialogue Manager. In *Proc. Twente Workshop on Language Technology 11: Dialogue Management in Natural Language Systems*, pp. 1-11

Traum, D. & Rickel, J. (2002). Embodied Agents for Multi-party Dialogue in Immersive Virtual World. In *Proceedings of AAMAS 2002*, pp. 766-773, ACM Press.

Traum, D. & Larsson, S. (2003). The Information State Approach to Dialogue Management. In R. Smith, J. Kuppevelt (Eds.) *Current and New Directions in Dialogue*, Kluwer.