Embodied Agents for Multi-party Dialogue in Immersive Virtual Worlds

David Traum
University of Southern California
Institute for Creative Technologies
13274 Fiji Way
Marina del Rey, CA 90292
traum@ict.usc.edu

Jeff Rickel
University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
rickel@isi.edu

ABSTRACT

We present a model of dialogue for embodied virtual agents that can communicate with multiple (human and virtual) agents in a multi-modal setting, including face-to-face spoken and nonverbal, as well as radio interaction, spanning multiple conversations in support of an extended complex task. The model builds on previous work in embodied agents and multi-layer dialogue models, and is being deployed in a peacekeeping mission rehearsal exercise setting.

1. INTRODUCTION

Immersive virtual worlds offer exciting potential for rich interactive experiences. Human users can cohabit three-dimensional graphical environments with virtual humans for entertainment, education, and training. They can have adventures in fantasy worlds. They can learn about history or other cultures by experiencing life in distant places and times. They can practice tasks, make mistakes, and gain experience without the consequences of real-world failure. In all these applications, virtual humans can play a wide variety of roles, including mentors and guides, teammates, companions, adversaries, and the local populace.

Perhaps the greatest challenge in creating virtual humans for interactive experiences is supporting face-to-face communication among people and virtual humans. On one hand, virtual worlds are an ideal application for current spoken language technology: they provide a microworld where conversation can legitimately be restricted to the events and objects within its confines. On the other hand, they raise issues that have received relatively little attention in computational linguistics. First, face-to-face communication in virtual worlds requires attention to all the nonverbal signals (e.g., gaze, gestures, and facial displays) that accompany human speech. Second, conversations that are situated in a 3D world raise a host of issues, including the attentional focus of the conversants, whether and to what degree they can see and hear one another, and the relative locations of conversants and the objects they are discussing. Finally, since there will typically be multiple real and virtual people, virtual worlds require support for multi-party conversations, including the ability to reason about the active participants in a conversation as well as who else might be listening. While there has been some early work in the area of embodied conversational agents [10, 19], and some of this work has addressed human-agent dialogues situated in 3D virtual worlds [31], there is currently no general model of such dialogues.

In this paper, we discuss progress towards a model of multi-party dialogue in immersive virtual worlds. Our model builds on prior work on embodied conversational agents, the social psychology literature on the nonverbal signals that accompany human speech, and models of collaborative dialogue from computational linguistics. The model is organized as a set of dialogue management layers, each including an information state and a set of dialogue acts that change that state. The layers include traditional ones, such as turn-taking and grounding, as well as several novel layers addressing the issues of multi-party dialogue in immersive worlds. Physical and verbal actions will often contribute to multiple layers.

2. EXAMPLE SCENARIO

The test-bed for our embodied agents is the Mission Rehearsal Exercise project at The University of Southern California's Institute for Creative Technologies. The setting for this project is a virtual reality theatre, including a visual scene projected onto an 8 foot tall screen that wraps around the viewer in a 150 degree arc (12 foot radius). Immersive audio software provides multiple tracks of spatialized sounds, played through ten speakers located around the user and two subwoofers.

Within this setting, a virtual environment has been constructed, representing a small village in Bosnia, complete with buildings, vehicles, and virtual characters. Within this environment has been placed an Army peacekeeping scenario: a U.S. Army lieutenant finds himself in the passenger seat of a simulated Army vehicle speeding towards a Bosnian village to help a platoon in trouble. Suddenly, he rounds a corner to find that one of his platoon's vehicles has crashed

Figure 1: An interactive peacekeeping scenario featuring (left to right in foreground) a sergeant, a mother, and a medic.

into a civilian vehicle, injuring a local boy (Figure 1). The boy's mother and an Army medic are hunched over him, and a sergeant approaches the lieutenant to brief him on the situation. Urgent radio calls from the other platoon, as well as occasional explosions and weapons fire from that direction, suggest that the lieutenant send his troops to help them. Emotional pleas from the boy's mother, as well as a grim assessment by the medic that the boy needs a medevac immediately, suggest that the lieutenant instead use his troops to secure a landing zone for the medevac helicopter.

A preliminary demonstration simulation has been implemented [36], using Hollywood storytelling, combined with technical expertise from USC. Figure 2 shows a small excerpt from the simulation script.

In this interaction, a number of issues arise for embodied agents, some going beyond capabilities of current implemented systems. First, at a broad level, we can see that the agents concern themselves with multiple agents and multiple conversations during this interaction. The main scene in Figure 2 concerned the Lt and Sgt, but the medic was also brought in, and the mother was an important overhearer. Other platoon members and townspeople may also be potential overhearers. There is also a separate conversation between the Platoon Sgt and the squad leaders, starting at the end of the excerpt given here. Also, in other parts of the scenario, the Lt engages in radio conversations with his home base, another platoon, and sometimes a medevac helicopter. Some of these conversations have fixed beginning and ending points (especially the radio conversations), while others are more episodic, trailing away as the local purpose of the interaction is established and resolved, and attention of the parties shifts to other matters. In all cases, agents must reason about who they are talking to, who is listening, and whether they are being addressed or not.

There is also an issue, in the immersive virtual world, of coordination of speech with other communicative modalities. In many cases, gestures and other nonverbal cues are important in carrying some of the communicative function. Some examples here are the way the Sergeant walks up to the Lt to initiate conversation, the way that the Sergeant glances at the medic to signal that he should take the turn and respond to the Lt's question, and the way the medic glances at the mother while formulating a less direct answer about the boy's health — focusing on the consequence of his condition rather than directly stating what might be upsetting to the mother.

3. PRIOR WORK

Our work builds on prior work in the areas of embodied conversational agents [10] and animated pedagogical agents [19]. Several systems have carefully modeled the interplay between speech and nonverbal behavior [9, 11, 8, 28], but these systems have focused exclusively on dyadic conversation, and they did not allow users and agents to cohabit a virtual world. The Gandalf system [11] allowed an agent and human to cohabit a real physical space, and to use gaze and gesture to reference an object (i.e., a wall-mounted display screen) in that space, but the agent's presence was limited to a head and hand on a 2D computer monitor. Similarly, the Rea agent [8] can transport herself to and into virtual houses and apartments, and the user can point to some objects within those virtual environments, but the user is not immersed in the environment, and Rea's movement and references within those environments is very limited. The Cosmo agent [24] includes a sophisticated speech and gesture generation module that chooses appropriate deictic references and gestures to objects in its virtual world based on both spatial considerations and the dialogue context, but the agent and its environment are rendered in 2D and the user does not cohabit the virtual world with Cosmo.

In contrast, Steve [31, 33, 32] cohabits 3D virtual worlds with people and other Steve agents, so it has addressed both multi-party and immersive aspects of dialogue in virtual worlds. Steve agents use path planning algorithms to

actor	speech	nonverbal
SGT		Walk up to LT
LT	Sergeant, what happened here?	
SGT	They just shot out from the side street sir.	Gesturing towards the civilian vehicle
	The driver couldn't see'em coming.	
LT	How many people are hurt?	
SGT	The boy and one of our drivers.	Gesturing toward the boy
LT	Are the injuries serious?	
SGT		Makes eye contact with medic and nods
MEDIC	Driver's got a cracked rib but the kid's –	Glancing at the mother
	Sir, we gotta get a Medevac in here ASAP.	
LT	We'll get it.	
LT	Platoon Sergeant, secure the area.	
SGT	Yes Sir!	
SGT	(Shouting) Squad leaders!	Raises arm
	Listen up!	Looks around at squad leaders

Figure 2: Multi-modal, multi-character interaction excerpt (many nonverbal behaviors omitted).

move around in virtual worlds, they are sensitive to where human users are looking, they can use gaze and deictic gestures to reference arbitrary objects in those worlds, and they can use gaze to regulate turn-taking in multi-party (team) dialogues. However, while Steve includes a dialogue model built on ideas from computational linguistics [33], it falls far short of the models in state-of-the-art spoken dialogue systems. Moreover, the model focuses primarily on the context of dyadic conversations between a Steve agent and his human student; there is very little dialogue context maintained for the multi-party dialogues between a Steve agent and his human and agent teammates.

Work in computational linguistics has focused on a complementary set of issues: it has largely ignored issues of embodiment and immersion in virtual worlds, but has produced relatively sophisticated models of spoken dialogue that include a variety of hooks for multiple modalities. We follow the framework of the Trindi project [23], using *dialogue moves* (in this case, corresponding to actions) as abstract input and output descriptions for the dialogue modelling component. This serves particularly well for considering multimodal communication, since it allows maximum flexibility of description, including moves that could be ambiguously realized using either speech or another modality, moves that require realization using a combination of multiple modalities, or moves that specify a modality. We also view the dialogue moves (and the affiliated information states) as segmented into a number of layers, each concerning a distinct aspect of information state, and using different classes of dialogue acts. Moreover, there is no one to one correspondence between dialogue acts and atomic communication realizations: a single utterance (or gestural communicative action) will generally correspond to multiple (parts of) dialogue acts, and it may take several communications (sometimes split into multiple modalities) to realize some dialogue acts.

As a starting point, we use the dialogue layers developed in the TRAINS and EDIS dialogue systems [39, 29, 26]. These included layers for turn-taking, grounding, core speech acts, and argumentation acts (later termed forward and backward-looking acts [16]). While not fully implemented within natural language dialogue systems, there has also been some other work on other layers that become important for dealing with the multi-character, multi-conversation domain. This includes work by Novick on meta-locutionary acts, including an attention level [27], work by Allwood and Clark on basic communicative functions [2, 13], work by Bunt on interaction management functions [7], and work on multi-level grounding in an extended multi-modal task interaction [15].

4. CURRENT WORK: MULTI-MODAL DIALOGUE MODEL

Our dialogue agents are designed to run within the Mission Rehearsal Exercise environment [36, 30]. This environment includes a message-passing event simulator, immersive sound, and graphics including static scene elements and special effects, rendered by Multigen/Paradigm's Vega.

Our agent model is based on Steve, as described in the previous section. Within the Mission Rehearsal Exercise scenario, Steve (and other agents) are given dynamically animated bodies from Boston Dynamics' PeopleShop [6]; the primitive motions were created using motion capture, and the Steve agents sequence these motions dynamically in response to the situation by sending commands to the PeopleShop run-time software. The medic and sergeant include expressive faces created by Haptek (www.haptek.com) that support synchronization of lip movements to speech. Steve's dialogue model and representation of the interactional state is being augmented with the new dialogue model presented here.

Our dialogue model currently consists of the layers shown in Figure 3. Each of these is modeled from the perspective of an agent involved in the interaction.

We will first briefly describe each of these, and then give details of the layers and how the associated acts may be realized using the palette of multi-modal communicative abilities. The *contact* layer concerns whether and how other individuals can be accessible for communication. Modalities include visual, voice (shout, normal, whisper), and radio. The

- contact
- attention
- conversation
  - participants
  - turn
  - initiative
  - grounding
  - topic
  - rhetorical
- social commitments (obligations)
- negotiation

Figure 3: Multi-party, Multi-conversation Dialogue Layers

attention layer concerns the object or process that agents attend to. Contact is a prerequisite for attention. The *Conversation* layer models the separate dialogue episodes that go on throughout the interaction. A conversation is a reified process entity, consisting of a number of sub-fields. Each of these fields may be different for different conversations happening at the same time. The *participants* may be active speakers, addressees, or overhearers [14]. The *turn* indicates the participant with the right to communicate (using the primary channel). The *initiative* indicates the participant who is managing the content expressed. The *grounding* component of a conversation tracks how information is added to the common ground of the participants. The conversation structure also includes a *topic* that governs relevance, and *rhetorical* connections between individual content units. Once material is grounded, even as it still relates to the topic and rhetorical structure of an ongoing conversation, it is also added to the social fabric linking agents, and not part of any conversation. This includes *social commitments* — both obligations to act or restrictions on action, as well as commitments to factual information. There is also a *negotiation* layer, modelling how agents come to agree on these commitments. We now turn to the layers in more detail.

The contact layer is modelled as a vector for all participants that the agent may interact with, each element indicating whether the participant is in contact in the media specified above. There are also dimensions for whether someone is in contact to *send* or *receive* communications by this modality. The actions influencing this layer are **make-contact**, which could be established by turning on a radio or walking over to within eye contact or earshot, and **break-contact**, which could be established by walking out of hearing, turning out of view (or moving behind something), or turning off the radio. Contact is not generally realized verbally, although one might indicate a desire for contact, e.g., by shouting for someone to come over. An example of a make-contact action is shown at the beginning of our example in Figure 2, where the Sergeant walks over to the lieutenant, to initiate contact (for the purpose of starting a conversation).

The attention layer is modelled by a similar vector to that of contact, though also including an entry for the agent itself, and attention is a one-way phenomenon, rather than having (potentially) distinct send and receive dimensions. The actions affecting this layer are divided into those that an agent performs concerning its own attention, and those related to the attention of other agents. **Give-attention** involves paying attention to some process, person, or object, as well as signalling this attention. This can be accomplished both verbally (e.g., saying "yes") or nonverbally (gazing at the object of attention). **Withdraw-attention** removes the current object from the attention entry of the agent. It can be implicit in giving attention to something else, or performed explicitly, by looking away in some cases (other than when engaged in conversation and serving some other purpose, such as planning a turn or indicating turn-taking). **Request-attention** signals to an agent that its attention is desired – it will also require a give-attention action by the other agent to change the attentional state. Request-attention can be signalled by a call on the radio, a shout, or using an agent's name, but also by gestures, such as raising an arm or waving. A **release-attention** act indicates that attention is no longer required. It occurs by default when a process or action that is the object of attention ends. It can also be explicit, in the form of a dismissal, or gesture indicating lack of attention (looking away). Attention of the released agent may still persist, however, until withdrawn, or given to something else. **Direct-attention** signals that attention should be given to another object or event, rather than the signaller. This can often be accomplished with a deictic gesture, or with an utterance such as "look up!"

Conversation is often a purpose for which attention is desired. In this case, attention will be assumed (unless explicitly withdrawn) for the duration of the conversation. There are also explicit indicators of conversational openings and closings [34, 22]. Conversations are often opened with verbal greetings, but nonverbal actions can be very important as well. Kendon found a variety of nonverbal actions involved in the initiation of conversation [22]. The interaction typically starts with a "sighting" before the orientation and approach. Individuals who do not know each other well and have no special reason to greet each other will "catch the eye" by gazing longer than normal. Next, a "distance salutation" involves definite eye contact. This is followed by an approach, which typically involves gaze avoidance. Finally, a "close salutation" involves resumed eye contact. The BodyChat system [12] was the first to model these acts in animated agents. Conversational openings and closings are very formalized in the military radio modality, e.g., saying "out" to close a conversation. Actions such as **open**, **continue** and **close** a conversation can be performed either explicitly or implicitly. Also, there are actions for maintaining and changing the presence and status of participants. An example from the sample dialogue is the way the medic is brought into the conversation from being an overhearer to an active participant.

Turn-taking actions model shifts in the turn holder. Most can be realized verbally, nonverbally, or through a combination of the two. **Take-turn** is an attempt to take the turn by starting to speak. **Request-turn** (e.g., signalled by various speech preparation signals such as opening the mouth or raising the hands into gesture space, or by avoiding a speaker's gaze at phrase boundaries) is an attempt to request the turn without forcibly taking it [3]. **Release-turn** (e.g., signalled by an intonational boundary tone, removal of the hands from gesture space, or a sustained gaze at the listener at the end of an utterance) is an attempt to offer the turn to the listener [3, 17]. **Hold-turn** (e.g., signalled verbally by a filled pause, or nonverbally by gaze aversion at phrase boundaries or hands in gesture space) is an attempt to keep the turn at a point where the listener might otherwise take it [3, 17]. These four turn-taking acts have been modeled in embodied conversational agents since the earliest systems [9]. In multi-party dialogue, there is one more act: **assign-turn** (e.g., signalled verbally by a vocative expression or nonverbally by a speaker's gaze at the end of an utterance) can be used to explicitly select the next speaker [3]. Among embodied conversational agents, only Steve includes this act.

We use *Initiative* to model how the agent should plan contributions. Even though the turn may shift from speaker to speaker, in many parts of a dialogue, a single agent is controlling the flow of the contributions, while others only respond to the initiative of the other. For some *mixed-initiative* dialogues, initiative may shift from one participant to another. Initiative is sometimes pre-allocated by role for specific tasks. Otherwise, it starts with the agent who opened the conversation, and can be managed with **take-initiative**, **hold-initiative**, and **release-initiative** actions. These acts can often be signalled by performing (only) appropriate core-speech acts in context, e.g., as proposed by [41, 40]. We are not currently considering nonverbal signals of initiative.

We use the grounding acts from [39, 37]: **initiate**, **continue**, **repair**, **request-repair**, **display**, **acknowledge**, **request-acknowledge**, and **cancel**. See previous work for details of all but display, which is an explicit signal of what was understood (e.g., repeating a word, performing an action), leaving it to the original speaker to decide if this act functions as a repair, request-repair, or acknowledge [20]. Embodied conversational agents typically include nonverbal actions for request-acknowledge (e.g., gaze at listener at grammatical pauses) and acknowledge (e.g., gaze at speaker and nod), and some include request-repair (e.g., when speech recognition fails, Peedy the parrot cups his wing to his ear and says "Huh?" [5]).

Topic actions include **start-topic** and **end-topic**. Topic structure can also be complex, when new topics are started before old ones have completed. Topic shifts of various sorts can be signalled explicitly with cue phrases (e.g., "now," "anyway"), but also with nonverbal cues. Head movements can signal topic shifts; a related sequence of utterances by a speaker typically uses the same basic head movement, and the speaker will often employ a new type of head movement to mark the start of a new topic [21]. Topic shifts are also frequently accompanied by shifts in the speaker's body posture [21].

Our model also includes layers for rhetorical structure of topical elements within a conversation [25], obligations and commitments [38, 1], and negotiation [4, 35]. We will, however, skip detailed discussion of these layers for the present because they have received a fair amount of attention in the previous literature.

5. PLANS AND EVALUATION

Aside from analysis of the scripted interaction, we also have two venues for evaluating the multi-modal, multi-party, multi-conversation dialogue model presented above. First, we are currently developing a prototype end-to-end spoken-language dialogue agent (including also the emotion models of [18]), to be deployed within the immersive virtual world. The goal for the prototype is to handle at least the fragment presented in Figure 2, and suitable variations. Given appropriate knowledge, goals, and connections to the virtual body, this agent system should be able to function as either the sergeant, medic, or mother (we assume, at least for the present, a human lieutenant, the subject of the training exercise).

We are also gathering actual performance data to complement the Hollywood-scripted interaction. We have designed a variant of the mission rehearsal exercise with a combination of human participants and virtual characters — humans portray not just the lieutenant (who is an external test-subject), but also the Sergeant, and other agents who communicate by radio (command post, other platoon, medevac pilot). Moreover, there is a "Wizard" controlling the computational agents (and communicating with the live ones), using a fixed palette of behaviors. Initial pilots have demonstrated the viability of the approach, and we will soon be gathering usable data on how the human participants interact with each other and with the virtual characters. We will use this data to evaluate the dialogue models and inform future system development.

Acknowledgements

We would like to thank members of ICT and ISI working on the MRE project for interesting discussions about the complexities in this scenario. The work described in this paper was supported in part by the U.S. Army Research Office under contract #DAAD19-99-C-0046. The content of this article does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred.

6. REFERENCES

[1] J. Allwood. Obligations and options in dialogue. *Think Quarterly*, 3:9–18, 1994.

[2] J. Allwood, J. Nivre, and E. Ahlsen. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9, 1992.

[3] M. Argyle and M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press, Cambridge, 1976.

[4] M. Baker. A model for negotiation in teaching-learning dialogues. *Journal of Artificial Intelligence in Education*, 5(2):199–254, 1994.

[5] G. Ball, D. Ling, D. Kurlander, J. Miller, D. Pugh, T. Skelly, A. Stankosky, D. Thiel, M. van Dantzich, and T. Wax. Lifelike computer characters: the persona

project at microsoft. In J. Bradshaw, editor, *Software Agents*. AAAI/MIT Press, Menlo Park, CA, 1997.

[6] Boston Dynamics. *PeopleShop 1.4 User Manual*, 2000.

[7] H. Bunt. Interaction management functions and context representation requirements. In *Proceedings of the Twente Workshop on Language Technology: Dialogue Management in Natural Language Systems (TWLT 11)*, pages 187–198, 1996.

[8] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan. Conversation as a system framework: Designing embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000.

[9] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proceedings of ACM SIGGRAPH '94*, pages 413–420, Reading, MA, 1994. Addison-Wesley.

[10] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors. *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000.

[11] J. Cassell and K. R. Thórisson. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13:519–538, 1999.

[12] J. Cassell and H. Vilhjálmsson. Fully embodied conversational avatars: Making communicative behaviors autonomous. *Autonomous Agents and Multi-Agent Systems*, 2:45–64, 1999.

[13] H. H. Clark. Managing problems in speaking. *Speech Communication*, 15:243 – 250, 1994.

[14] H. H. Clark. *Using Language*. Cambridge University Press, Cambridge, England, 1996.

[15] P. Dillenbourg, D. Traum, and D. Schneider. Grounding in multi-modal task-oriented collaboration. In *Proceedings of the European Conference on AI in Education*, 1996.

[16] Discourse Resource Initiative. Standards for dialogue coding in natural language processing. Report no. 167, Dagstuhl-Seminar, 1997.

[17] S. Duncan, Jr. Some signals and rules for taking speaking turns in conversations. In S. Weitz, editor, *Nonverbal Communication*, pages 298–311. Oxford University Press, 1974.

[18] J. Gratch and S. Marsella. Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. In *Proceedings of the Fifth International Conference on Autonomous Agents*, 2001.

[19] W. L. Johnson, J. W. Rickel, and J. C. Lester. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11:47–78, 2000.

[20] Y. Katagiri and A. Shimojima. Display acts in grounding negotiations. In *Proceedings of Gotalog 2000, the 4th Workshop on the Semantics and Pragmatics of Dialogue*, pages 195–198, 2000.

[21] A. Kendon. Some relationships between body motion and speech. In A. Siegman and B. Pope, editors, *Studies in Dyadic Communication*, pages 177–210. Pergamon Press, New York, 1972.

[22] A. Kendon. A description of some human greetings. In R. Michael and J. Crook, editors, *Comparative Ecology and Behavior of Primates*, pages 591–668. Academic Press, New York, 1973.

[23] S. Larsson and D. Traum. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering*, 6:323–340, September 2000. Special Issue on Spoken Language Dialogue System Engineering.

[24] J. C. Lester, J. L. Voerman, S. G. Towns, and C. B. Callaway. Deictic believability: Coordinating gesture, locomotion, and speech in lifelike pedagogical agents. *Applied Artificial Intelligence*, 13:383–414, 1999.

[25] W. C. Mann and S. A. Thompson. Rhetorical structure theory: A theory of text organization. Technical Report ISI/RS-87-190, USC, Information Sciences Institute, June 1987.

[26] C. Matheson, M. Poesio, and D. Traum. Modelling grounding and discourse obligations using update rules. In *Proceedings of the First Conference of the North American Chapter of the Association for Computational Linguistics*, 2000.

[27] D. Novick. *Control of Mixed-Initiative Discourse Through Meta-Locutionary Acts: A Computational Model*. PhD thesis, University of Oregon, 1988. also available as U. Oregon Computer and Information Science Tech Report CIS-TR-88-18.

[28] C. Pelachaud, N. I. Badler, and M. Steedman. Generating facial expressions for speech. *Cognitive Science*, 20(1), 1996.

[29] M. Poesio and D. R. Traum. Towards an axiomatization of dialogue acts. In *Proceedings of Twendial'98, 13th Twente Workshop on Language Technology: Formal Semantics and Pragmatics of Dialogue*, 1998.

[30] J. Rickel, J. Gratch, R. Hill, S. Marsella, and W. Swartout. Steve goes to Bosnia: Towards a new generation of virtual humans for interactive experiences. In *AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment*, March 2001.

[31] J. Rickel and W. L. Johnson. Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13:343–382, 1999.

[32] J. Rickel and W. L. Johnson. Virtual humans for team training in virtual reality. In *Proceedings of the Ninth International Conference on Artificial Intelligence in Education*, pages 578–585. IOS Press, 1999.

[33] J. Rickel and W. L. Johnson. Task-oriented collaboration with embodied agents in virtual worlds. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents.* MIT Press, Cambridge, MA, 2000.

[34] E. A. Schegloff and H. Sacks. Opening up closings. *Semiotica*, 7:289–327, 1973.

[35] C. L. Sidner. An artificial discourse language for collaborative negotiation. In *Proceedings of the forteenth National Conference of the American Association for Artificial Intelligence (AAAI-94)*, pages 814–819, 1994.

[36] W. Swartout, R. Hill, J. Gratch, W. Johnson, C. Kyriakakis, K. Labore, R. Lindheim, S. Marsella, D. Miraglia, B. Moore, J. Morie, J. Rickel, M. Thiebaux, L. Tuch, R. Whitney, and J. Douglas. Toward the holodeck: Integrating graphics, sound, character and story. In *Proceedings of 5th International Conference on Autonomous Agents*, 2001.

[37] D. R. Traum. *A Computational Theory of Grounding in Natural Language Conversation.* PhD thesis, Department of Computer Science, University of Rochester, 1994. Also available as TR 545, Department of Computer Science, University of Rochester.

[38] D. R. Traum and J. F. Allen. Discourse obligations in dialogue processing. In *Proceedings of the* 32^{nd} *Annual Meeting of the Association for Computational Linguistics*, pages 1–8, 1994.

[39] D. R. Traum and E. A. Hinkelman. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3):575–599, 1992. Special Issue on Non-literal language.

[40] M. A. Walker and S. Whittaker. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proceedings ACL-90*, pages 70–78, 1990.

[41] S. Whittaker and P. Stenton. Cues and control in expert-client dialogues. In *Proceedings ACL-88*, pages 123–130, 1988.