Animating Conversation in Online Games

Hannes Högni Vilhjálmsson

USC Information Sciences Institute[1]
4676 Admiralty Way, Marina del Rey, CA 90292, USA
hannes@isi.edu

Abstract. When players in online games engage in conversation with each other, often through a chat window, their graphical avatars typically do not exhibit interesting behavior. This paper proposes using a model of face-to-face communication to automatically generate realistic nonverbal behavior in the avatars based on what is going on in the conversation. It describes Spark, a flexible XML based architecture that makes this possible and reports on a user study that shows how such avatars could significantly improve online conversations.

1 Introduction

The term *avatar* refers to the graphical embodiment of a user in a virtual environment. In the same way that the avatars of Hindu mythology allowed the gods to walk among the humans by giving them corporeal form, graphical avatars allow human players to navigate and interact with electronic game worlds. To provide the illusion of presence in those worlds, the avatars need to obey the laws of the game, often including those of physics. For example avatars explode when hit or fall when driven off a ledge. As game worlds get richer, the avatars interact with it in richer ways, while keeping the player controller reasonably simple. For example, to make the avatar leap into a nearby vehicle, close the door and start the engine, a simple click of the mouse on the vehicle should suffice. Avatars have become more than simple puppets under the direct control of players and instead require intelligent behavior to carry out the player's intent and react to the world around them.

Different kinds of game environments require different kinds of intelligent avatar behavior. Shooter avatars know how to switch to the next best weapon when the current weapon is out of ammo, and adventure avatars know how to navigate around furniture to reach a chest across the room. In general, avatars are pretty good at handling dangerous situations, including dexterous locomotion and all manners of dying. But what about social interactions? When people meet face-to-face in the real world, their bodies spontaneously exhibit a range of highly coordinated behaviors that carry out their communicative intent. Yet, when players meet in an online game, their ava-

[1] The paper describes research done by the author at the MIT Media Lab

In M.Rauterberg (Ed.): Entertainment Computing ICEC 2004, LNCS 3166, pp. 139-150, 2004. © IFIP

tars typically cease all motion because they don't know what behaviors to exhibit and their players are too busy typing messages to each other to actually direct them.

This paper describes how avatars can be brought to live during conversations between online players. By deriving communicative intent from the text that is typed using natural language techniques and consulting empirical models of typical human conversational behavior, avatars can generate something that approximates what the conversation would look like if the players were having it face-to-face. By providing behavior that supports the interaction, rather than using no motion or randomly generated motion, the avatars could even make in-game communication between players more effective.

The next section will discuss relevant background and then the paper will propose an approach and a system architecture called Spark to automatically generate conversational behavior in avatars. The following section describes how an implementation of this architecture was evaluated and finally the paper wraps up with conclusions.

2 Background

2.1 Face-to-face Conversation

When people meet in the real world, they follow certain protocols without even thinking about it. These protocols both help them coordinate their actions and effectively communicate what they want to say. The former process has been referred to as interactional, while the latter has been referred to as propositional in nature [6]. There is certainly more going on in interactions face-to-face than these two types of processes, but they provide and maintain the underlying channel for communication and are therefore fundamental to human conversation.

On the interactional side, two important functions are turn management and feedback. Properly managing turns is necessary to ensure that everyone is not speaking at the same time, and can therefore be clearly heard. Turns are requested, taken, held and given using various signals, often exchanged in parallel with speech over nonverbal channels such as gaze, intonation, and gesture [10, 15]. Taking or requesting turn most often coincides with breaking eye contact [2] while raising hands into gesture space [17]. A speaker gives the turn by looking at the listener, or whoever is meant to speak next, and resting the hands [10, 14, 22].

Speakers often request feedback while speaking and expect at the very least some sign of attention from their listeners. Feedback requests typically involve looking at the listeners and raising eyebrows [9]. To request a more involved feedback, this behavior can be supplemented with pointing the head towards the listener or conducting a series of low amplitude head nods ending with a head raise [22]. Listener response to feedback requests can take on a variety of forms. Brief assertion of attention can be given by the dropping of the eyelids and or slight head nod towards the speaker. A stronger cue of attention may involve a slight leaning and a look towards the speaker along with a short verbal response or a laugh. A speaker's ability to formulate messages is critically dependent on these attentive cues, and therefore, even if only one person is speaking, everyone present is engaged in some kind of communicative behavior.

The propositional side deals with how we make sure those who are listening pick up what we are saying. Three types of communicative function play an important role here: emphasis, reference and illustration. Emphasis signals to listeners what the speaker considers to be the most important contribution of each utterance. It commonly involves raising or lowering of the eyebrows and sometimes vertical head movement as well [3, 9]. A short formelss beat with either hand, striking on the stressed syllable, is also common [20]. Reference is most commonly carried out by a pointing hand. The reference can be made to the physical surroundings such as towards an object in the room, or to imaginary spots in space that for example represent something previously discussed [20]. References through pointing are also made towards the other participants of a conversation when the speaker wishes to acknowledge their previous contributions or remind them of a prior conversation [5]. Illustration is the spontaneous painting with the hands of some semantic feature of the current proposition. The particular features may lend themselves well to be portrayed by a visual modality, such as the configuration or size of objects, or the manner of motion [18, 20].

2.2 Avatars

Avatars in some form have been around in games since the virtual paddles in the original Pong, but the more human-like avatars used for online social interaction date back to the 2D sprites used in the first online graphical recreational world of Habitat in 1985 [21]. Since then online avatars have appeared in a wide variety of applications ranging from graphical chat rooms and networked games, to collaborative virtual environments and military training simulations. The model for generating communicative behaviors in avatars has generally relied on explicit user input, such as key presses or menu selections. For example, the pioneering online 3D communities *Blaxxun Contact* and *Active Worlds* allowed users to gesture with their avatars by clicking on labeled buttons. In popular online shooters such as *Unreal Tournament*, players can exhibit nonverbal taunts by pressing a key and in massively multiplayer games like *EverQuest*, special emote commands in the text chat window produce corresponding behaviors in the avatars.

This approach is fine for acting out deliberate displays of affect or insult, or to issue orders, but as was argued in [26], requiring the users to think about how to coordinate their virtual body every time they communicate places on them the burden of too much micromanagement. When people communicate face-to-face, they are not used to think about gesture production because it is something that happens spontaneously without conscious effort [17]. In the same way that avatars should animate walk cycles so that players won't have to worry about where to place their virtual feet, avatars should also provide the basic nonverbal foundation for communication.

Several innovative avatar interfaces have been developed, though few have made it into the gaming mainstream. Perhaps the best known technique for animating avatars during player conversation, pioneered by *Oz Virtual* and *OnLive!*, is to generate mouth movement based on the amplitude or even phoneme analysis of a speech signal. While this is a good way to see who is talking and get some idea of how intense the speaker is, it requires a voice link between players and focuses on very limited movement, leaving out important communicative behaviors such as gesture and gaze.

Animating gesture in avatars has been attempted through special input devices including a pen [4] and physical puppets [16], which may provide more intuitive interfaces than menus, and even generate personal motion nuances, but they still require deliberate effort to use. Bringing the avatar control to a much higher level, to avoid micromanagement of behavior, has been suggested by several researchers including [23]. This is a very good idea which still needs work, since it is not clear exactly how the player would best indicate their intent while playing a game, perhaps calling for ways to somehow automatically derive intent.

Automating the generation of communicative nonverbal behaviors in avatars was first proposed in BodyChat where avatars were not just waiting for their own users to issue behavior commands, but were also reacting to events in the online world according to preprogrammed rules based on a model of human face-to-face behavior [26]. The focus was on gaze cues associated with the interactional communicative processes. A study showed that the automation did not make the users feel any less in control over their social interactions, compared to using menu driven avatars. In fact, they reported they felt even more in control, suggesting that the automated avatars were providing some level of support [7].

If the avatar behavior is to be automated, one can ask how important it is that the behavior be carefully modeled to reflect the communicative intent of their users. Wouldn't it be enough to automatically generate some randomized nonverbal behavior to appear life-like? After all, that is an approach that many of today's systems take. According to several studies that compare algorithmically generated gaze behavior, taking into account speaker and listener roles, to randomized gaze behaviors in avatars, the algorithmic behaviors have proven more effective [11, 24].

So far, the automated behaviors have still mostly been restricted to gaze and mouth movement, and supported interactional management only. While some interesting work is being done on posture generation as a function of attitudes between users [12], no work to date has attempted to support both the interactional and the propositional procceses of conversation through real-time analysis of communicative intent.

3 The Spark Architecture

3.1 Overview

Spark is a new online messaging architecture that builds both on the previous BodyChat work [26] and the work on generating gesture from text in the BEAT system [8]. Lets first look at an example interaction to explain what Spark does.

Two players are having a conversation through their avatars, a wizard and a druid, when a third player comes running with a ranger avatar. As the ranger approaches, the wizard and druid look at her, the wizard raising his eyebrows in recognition and smiling as an invitation for her to join the conversation. The player of the ranger moves her avatar closer to the others and starts typing a message while her avatar nods in salutation and raises her hands in the air in preparation to speak, momentarily glancing away as if she needed to gather her thoughts. When the ranger starts speaking, voicing the words her user just typed, she looks at the wizard and begins gesticulating: "I just came back from the mountain pass". A brief nod of emphasis occurs on the coming back part and a glance and a hand wave towards the mountains behind her when mentioning the pass. The wizard and druid both glance up towards the mountains and then look back at the ranger nodding. She continues with eyes wide and raised eyebrows: "I noticed an Ogre headed this way, he had a group of goblins following him!". Her animated hands go up to indicate the tallness of the Ogre and then sweep low when mentioning the smaller goblins.

The players here did nothing but maneuver their avatars into place and then type what they wanted to say. Their avatars synthesized both their voices and nonverbal behavior not only to make it look like a believable conversation was taking place, but also to add some visual emphasis to what was being said. Spark accomplishes this by analyzing written messages in terms of their communicative intent, annotate them using tags that the avatars will understand, and then pass the message back to all the avatars engaged in the conversation so they can all generate associated communicative behaviors. The speaking avatar needs to generate behavior that communicates the intent while the listening avatars need to generate proper reactive behavior. The following sections describe the parts of this process in greater detail.

3.2 Frames

All interaction related events that pass through the Spark architecture are represented by an XML data structure termed a frame. A frame holds the actual event description, such as a written message, along with some additional context that can be used to interpret the event. A frame is a dynamic entity that can have its event description expanded and an interpretation added as it passes through a sequence of analyzing processes (see Figure 1). There are action frames and utterance frames. An action is a communicative event that occurs in the absence of a verbal message, such as when someone starts to type (but hasn't pressed ENTER yet) or selects something on the screen (perhaps picking an object or a person). An utterance contains words and eventually any nonverbal behavior that are associated with the delivery of those words.

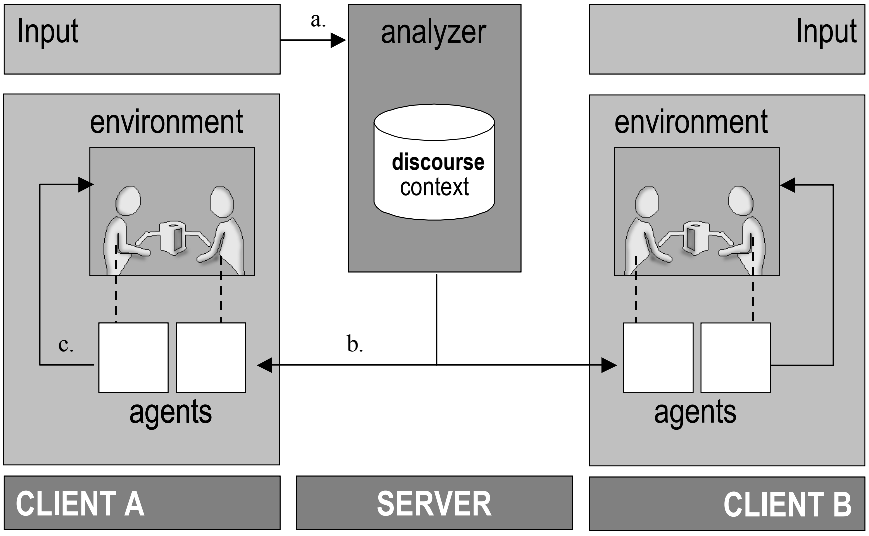

Fig. 1. Overview of the Spark architecture. A frame with raw input (a) enters an analyzer on the server-side, which annotates the frame with XML tags describing communicative function. The functionally annotated frame (b) is distributed to all clients where avatar agents transform those annotations into behavior annotations according to rules. The behaviorally annotated frame (c) is finally used to render the whole synchronized performance in the environment

3.3 Analyzer

The Analyzer sits on the server side and interprets all incoming frames, annotating their communicative function. It consists of two modules, one to process action frames and one to process utterance frames. Both modules have access to a special data structure called discourse context to help with the interpretation.

Action Module The action module interprets the action described in the frame and maps it from an interface event to a communicative action. For example it would map the event when a player starts to type a message into an action requesting a turn in a conversation.

Discourse Module The discourse module carries out a series of linguistic and discourse analyses to identify and label how the various units of discourse within the utterance text, such as words, phrases and clauses, contribute to the conversation processes described in 2.1. It passes utterance frames to specialized methods that each focuses on one communicative function and adds only one new type of XML tag to the frame. This way new or improved methods can easily be plugged in without changes to the rest of the module. The methods currently implemented in Spark represent some known heuristic methods as well as some new ones. These are all described in more detail in [8] and [25], but are summarized here in Table 1.

Discourse Context An action or words alone don't tell the full story, they always occur in some context crucial for determining their communicative function. The discourse context in Spark is represented by three main data structures: Discourse Model, Domain Knowledge and Participation Framework. The Discourse Model describes how information gets shared over the course of an entire conversation [1]. It contains a Discourse History that holds all the utterances exchanged so far. Discourse entities, corresponding to the unique objects referred to by noun phrases, are also kept track of in a list where the last entity referred to is moved to the top. The Domain Knowledge is a static structure that describes the ontology of a particular domain that relates to the conversation. It describes visual features of objects and actions that can be used for gestural illustration. The Participation Framework is a dynamic structure that keeps track of who is present and what their roles currently are in relation to the current utterance [13].

Table 1. A summary of methods that annotate frames with communicative function

Method	Description
Information Structure	Finds the most information rich part of a clause
Discourse Entities	Matches noun phrases with objects in ontology
Reference	For each discourse entity marks a possible visual (can be seen) or textual (has been mentioned before) reference
Emphasis	Uses information structure to mark special focus
Contrast	Uses a dictionary to tag contrasting words like antonyms
Illustration	Tags visual features of identified entities using ontology
Topic shifts	Places a tag where a new topic starts
Grounding	Finds places where listener feedback should be requested
Turn taking	Inserts floor negotiation markers

3.4 Avatar Agents

Now that a frame has been analyzed and annotated with a rich description of a communicative event in the context of the current encounter, it can be sent to all the clients where the delivery itself can draw from this representation to coordinate a believable and effective performance by all the avatars present.

When a communicative frame arrives at a client, it is first handed to the Avatar Agent that represents the player that originated the communicative event, the *actor*. The actor's job is to now annotate the frame with actual behaviors that nonverbally carry out the communicative functions described by the XML tags. This is a matter of translating functional annotations into behavior annotations according to a set of translation rules that represent empirical findings in the human communicative behavior literature reviewed in section 2.1.. Some example mapping rules are shown in Table 2.

The frame is then passed around to all other Avatar Agents, participating in the scene, that then get a chance to add reacting behaviors. This way, everyone represented in a scene can have their avatars spontaneously react to what is going on. In fact, other agents than Avatar Agents could also participate in this reaction phase, for example a special Camera Agent could add camera moves, based on what is happening.

Table 2. An example how avatar agents map communicative function tags into behavior tags

Function Tag	Speaker Behavior	Listener Behavior
EMPHASIS_WORD	HEADNOD	
	GESTURE_BEAT	
EMPHASIS_PHRASE	EYEBROWS_RAISE	
GROUNDING_REQUEST	GLANCE(addressee)	GLANCE(speaker)
	EYEBROWS_RAISE	HEADNOD
TURN_GIVE	LOOK(addressee)	LOOK(addressee)
TURN_TAKE	GLANCE_AWAY	LOOK(speaker)
TOPICSHIFT	POSTURESHIFT	
REFERENCE_TEXTUAL	GLANCE(last ref. speaker)	
REFERENCE_VISUAL	GLANCE(object)	GLANCE(object)
	GESTURE_POINT(object)	
CONTRAST	GESTURE_CONTRAST	
ILLUSTRATE	GESTURE_FEATURE	

The output from the Avatar Agents is a frame that now is a detailed description of a performance that involves one or more avatars in the environment. This performance has to be carried out in the game world. The frame is given to a scheduler that converts any utterance to speech, retrieves the onset time for each word and thereby the onset and duration of each behavior spanning one or more words (see [8] for details) and then hands out timed behavior scripts to the individual avatars for execution.

4 Evaluation

The Spark approach was evaluated by implementing a scenario where three players had to get together in a special map room after being told they were prisoners in an enemy stronghold and had to plan their escape. This scenario was chosen because it involved conversation about a complex visual object, the map, and because it involved collaboration, which could provide some insight into whether the avatar behaviors contributed to the conversation process and a successful outcome. Besides, players coming together to strategize, is something that is a common occurrence in today's multiplayer games, but something that hasn't been animated before.

First, to evaluate how well the rules that the avatars possessed mimicked the real world, a physical mockup of the scene was created and videotaped with three live subjects performing the task. The utterances from a random 40-second segment of the video were fed through the Spark architecture and the resulting avatar behavior compared to that of the real people, which was annotated using Anvil [19]. The analysis involved emphasis gestures, pointing gestures, gaze direction and head movements. Overall Spark did well, making exact predictions of behavior more than half the time, with the wrong predictions mainly due to excessive emphasis and feedback generation. That makes sense, because the avatars in Spark generate a behavior every time a rule indicates a logical opportunity. If Spark took into account factors such as personality or affect, it might be able to use that as a principled way of reducing overall avatar liveliness. This is something to consider as future research.

50 subjects were signed up for 2 sessions each with the Spark system, where a session involved the group selecting what they believed to be the quickest escape route on the map in front of them. Each subject was briefed on a different set of helpful facts about the map prior to a session to ensure they needed to work together. In half of the sessions the subjects would just see the room and the map, and receive each other's messages without the avatars. In the other half, everything would be the same except they would see each other as animated avatars standing around the map see Figure 2. In both kinds of sessions the subjects could highlight parts of the map to indicate their choice of path. Two other conditions, crossed with the avatar versus no-avatar conditions, were the use of synthesized speech versus scrolling text. Apart from noting that people typically didn't like the synthesized voices, this part of the study won't be discussed further here.

The fact that each subject was assigned to 2 instead of all conditions (although balanced for order effects and only assigned to adjacent cells) of the 4 conditions in this 2x2 design, made the analysis of the data more difficult and contributed to lower power than with standard within-subject experiments, which suggests a simpler design for future studies. But nevertheless, some clear results emerged.

The 14 subjects that tried both an avatar system and a system without avatars were asked to compare the systems on a 9 point likert scale from a high preference for no avatars to a high preference for avatars along 6 dimensions including which system was "more useful", "more fun", "more personal", "easier to use", "more efficient" and "allowed easier communication". One-tailed t-tests showed that the preference for avatars was significant ($p<0.05$) for all but the "easier to use" question where no significant preference either way was found. These subjective results clearly indicate that people find the avatars compelling and helpful.

To test the hypothesis that the avatars automating communicative intent would improve the overall process of conversation, compared to non-embodied communication, 11 different measures of quality of conversation process were taken. 7 were objective behavioral measures from the chat logs, including the portion of utterances without explicit grounding (i.e. verbal verification of reception), portion of utterances that got requested replies, portion of non-overlapping utterances and portion of on-task utterances. 4 were subjective likert scale questionnaire measures, including sense of ability to communicate and sense of control over conversation. All but one measure was found higher in the avatar condition and a t-test of the grand mean (across all 11 normalized measures) showed that indeed it was significantly higher ($p<0.02$) in the avatar condition than in the non-avatar condition, supporting the hypothesis.

To test the hypothesis that the avatars automating communicative intent would improve the actual outcome of player collaboration, compared to the non-embodied communication, 8 different measures of the quality of task outcome were taken. 2 were objective measures, one being the quality of the escape route that the subjects chose together and the other being the completion time (which ranged from 5 to 40 minutes). 6 were subjective likert scale questionnaire measures including "How well did you think the group performed on the task?", "How strong do you think the group's consensus is about the final solution?" and "How much do you think you contributed to the final solution?". Again, all but one measure was higher in the avatar condition, and again, a t-test of the grand mean (across all 8 normalized measures) showed that it was significantly higher ($p<0.02$) in the avatar condition than in the non-avatar condition, supporting the hypothesis.

5 Conclusions

The Spark architecture provided a flexible framework for animating the avatars in a collaborative game setting, using behavior generation rules that provided a reasonably realistic performance by drawing from research into human communicative behavior. A study showed that these avatars compel players and that the avatar behavior may in fact be supporting the conversational activity. While the overall results were encouraging, a follow-up study is needed to really look at the effect the avatars had on individual measures. Work remains to be done in extending the set of rules to include more aspects of social interaction and to embrace research into personality and emotion to broaden the range of expression and avatar intelligence.

Acknowledgements

The author would especially like to thank Justine Cassell for her guidance throughout the project and Amy Bruckman, Cynthia Breazeal, Bruce Blumberg and Dan Ariely for feedback that helped shape this work. Many thanks go to Joey Chang, Kenny Chang, David Mellis, Ivan Petrakiev, Vitaly Kulikov, Alan Gardner, Timothy Chang and Baris Yuksel for technical assistance on the character animation system. The author is grateful to Ian Gouldstone and Nina Yu for beautiful artwork and to the entire GNL team and Deepa Iyengar for input along the way. Last but not least, the author would like to thank the many sponsors of Digital Life at the MIT Media Lab for their support.

Fig. 2. Subjects discussing their escape from the stronghold

References

1. Allen, J. *Natural Language Understanding*. The Benjamin/Cummings Publishing Company, Inc., Redwood City, CA, 1995.
2. Argyle, M. and Cook, M. *Gaze and Mutual Gaze*. *Cambridge University Press*, Cambridge, 1976.
3. Argyle, M., Ingham, R., Alkema, F. and McCallin, M. The Different Functions of Gaze. *Semiotica*., 1973
4. Barrientos, F., Continuous Control of Avatar Gesture. in *Bridging the Gap: Bringing Together New Media Artists and Multimedia Technologists, First International Workshop*, (Marina del Rey, CA, 2000), ACM.
5. Bavelas, J.B., Chovil, N., Coates, L. and Roe, L. Gestures Specialized for Dialogue. *Personality and Social Psychology*, *21* (4). 394-405.
6. Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H. and Yan, H., Embodiment in Conversational Interfaces: Rea. in *CHI*, (Pittsburgh, PI, 1999), ACM, 520-527.
7. Cassell, J. and Vilhjalmsson, H. Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous. *Autonomouse Agents and Multi-Agent Systems*, *2* (1). 45-64.
8. Cassell, J., Vilhjalmsson, H. and Bickmore, T., BEAT: the Behavior Expression Animation Toolkit. in *SIGGRAPH01*, (Los Angeles, CA, 2001), ACM, 477-486.
9. Chovil, N. Discourse-Oriented Facial Displays in Conversation. *Research on Language and Social Interaction*, *25* (1991/1992). 163-194.
10. Duncan, S. On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, *3*. 161-180.

11. Garau, M., Slater, M., Bee, S. and Angela Sasse, M., The Impact of Eye Gaze on Communication using Humanoid Avatars. in *CHI 2001*, (Seattle, WA, 2001), ACM, 309-316.
12. Gillies, M. and Ballin, D., A Model of Interpersonal Attitude and Posture Generation. in *Intelligent Virtual Agents*, (Irsee, Germany, 2003), Springer-Verlag.
13. Goffman, E. *Behavior in public places; notes on the social organization of gatherings*. Free Press of Glencoe, [New York], 1963.
14. Goffman, E. *Forms of Talk*. University of Pennsylvania Publications, Philadelphia, PA, 1983.
15. Goodwin, C. *Conversational Organization: Interaction between speakers and hearers*. Academic Press, New York, 1981.
16. Johnson, M.P., Wilson, A., Blumberg, B., Kline, C. and Bobick, A. Sympathetic Interfaces: Using a Plush Toy to Direct Synthetic Characters. *Proceedings of CHI'99*. 152-158.
17. Kendon, A. *Conducting Interaction: Patterns of behavior in focused encounters*. Cambridge University Press, New York, 1990.
18. Kendon, A. On Gesture: Its Complementary Relationship With Speech. in Siegman, A.W. and Feldstein, S. eds. *Nonverbal Behavior and Communication*, Lawrence Erlbaum Associates, Inc., Hillsdale, 1987, 65-97.
19. Kipp, M., Anvil - A Generic Annotation Tool for Multimodal Dialogue. in *Eurospeech*, (Aalborg, 2001), 1367-1370.
20. McNeill, D. *Hand and Mind*. The University of Chicago Press, Chicago and London, 1992.
21. Morningstar, C. and Farmer, F.R., The Lessons of Lucasfilm's Habitat. in *The First Annual International Conference on Cyberspace*, (1990).
22. Rosenfeld, H.M. Conversational Control Functions of Nonverbal Behavior. in Siegman, A.W. and Feldstein, S. eds. *Nonverbal Behavior and Communication*, Lawrence Erlbaum Associates, Inc., Hillsdale, 1987, 563-601.
23. Shi, J., Smith, T.J., Graniere, J. and Badler, N., Smart avatars in JackMOO. in *Virtual Reality*, (Houston, TX, 1999), IEEE.
24. Vertegaal, R. and Ding, Y., Explaining Effects of Eye Gaze on Mediated Group Conversations: Amount or Synchronization. in *CSCW 2002*, (New Orleans, LA, 2002), ACM, 41-48.
25. Vilhjalmsson, H., Augmenting Online Conversation through Automatice Discourse Tagging. in *HICSS Sixth Annual Minitrack on Persistent Conversation*, (Hawaii, Under Review).
26. Vilhjalmsson, H. and Cassell, J., BodyChat: Autonomous Communicative Behaviors in Avatars. in *Autonomous Agents*, (Minneapolis, MN, 1998), ACM, 269-276.