In *Proceedings of the2nd Annual ACM International Conference on Autonomous Agents,* Minneapolis, 1998

BodyChat: Autonomous Communicative Behaviors in Avatars

Hannes Högni Vilhjálmsson and Justine Cassell
MIT Media Laboratory
Gesture and Narrative Language Group
20 Ames Street, Cambridge, MA 02139-4307 USA
{hannes, justine}@media.mit.edu

Abstract

Although avatars may resemble animated communicating interface agents, they have for the most part not profited from recent research into autonomous systems. In particular, even though avatars function within conversational environments (for example, chat or games), and even though they often resemble humans (with a head, hands, and a body) they are incapable of representing the kinds of knowledge that humans have about how to use the body during communication. Their appearance does not translate into increased communicative bandwidth. Face-to-face conversation among humans, however, does make extensive use of the visual channel for interaction management where many subtle and even involuntary cues are read from stance, gaze and gesture. We argue that the modeling and animation of such fundamental behavior is crucial for the credibility and effectiveness of the virtual interaction in chat. By treating the avatar as a communicative agent, we propose a method to automate the animation of important communicative behavior, deriving from work in context analysis and discourse theory. BodyChat is a system that allows users to communicate via text while their avatars automatically animate attention, salutations, turn taking, back-channel feedback and facial expression, as well as simple body functions such as the blinking of the eyes.

1. BEHAVIORS IN AVATARS

One type of embodied agent that has received much airplay but little serious research attention in the agent community, is the avatar in a graphical chat. An avatar represents a user in a distributed virtual environment, but has until now not been autonomous. That is, it has not had knowledge to act in the absence of explicit control on the part of the user. In most current graphical chat systems the user is obliged to switch between controlling the avatar behavior and typing messages to other users. While the user is creating the message for her interlocutor, her avatar stands motionless or repeats a selected animation sequence. This fails to reflect the natural relationship between the body and the conversation that is taking place, potentially giving misleading or even conflicting visual cues to other users. Some voice-based systems offer simple lip synching, which greatly enhances the experience, but actions such as gaze and gesture have not been incorporated or are simply produced at random to create a sense of "liveliness".

The development of graphical chat environments from text-based IRCs indicates an awareness of the importance of the body, and of different communication modalities. More recently, the creators of multi-user environments have realized that avatars need to be animated in order to bring them to life, but their approach has not taken into account the number of different communicative functions of the body during an encounter. They provide menus where users can select from a set of animation sequences or switch between different emotional representations. The largest problem with this approach is that the user has to explicitly control every change in the avatar's state. In reality however, many of the visual cues important to conversation are spontaneous and even involuntary, making it impossible for the user to explicitly select them from a menu. Furthermore, the users are often busy producing the content of their conversation, so that simultaneous behavior control becomes a burden.

In addition, when people looked at the stiff early versions of avatars and considered ways to make them more life-like, generally they came to the conclusion that they were lacking *emotions*. However, lively emotional expression in interaction is in vain if mechanisms for establishing and maintaining mutual focus and attention are not in place (Thórisson and Cassell 1996). We tend to take communicative behaviors such as gaze and head movements for granted, as their spontaneous nature and non-voluntary fluid execution makes them easy to overlook when recalling a previous encounter (Cassell, forthcoming). This is a serious oversight when creating avatars or humanoid agents since emotion displays do not account for the majority of displays that occur in a human to human interaction (Chovil 1992).

2. AUTOMATING AVATAR BEHAVIOR

Many believe that employing trackers to map certain key parts of the user's body or face onto the graphical representation will solve the problem of having to explicitly control the avatar's every move. As the user moves, the avatar imitates the motion. This approach, when used in a non-immersive setting, shares a classical problem with video conferencing: The user's body resides in a space that is radically different from that of the avatar. This flaw becomes particularly apparent when multiple users try to interact, because the gaze pattern and orientation information gathered from a user looking at a monitor does not map appropriately onto an avatar standing in a group of other avatars. Thus whereas tracking may be appropriate for Virtual Reality applications where head mounted displays are employed, it does not lend itself well to Desktop Virtual Environments.

The approach to avatar design adopted here, in contradistinction to explicit control, treats the avatar as an autonomous agent acting of its own accord in a world inhabited by other similar avatars. However the autonomy is limited to a range of communicative expressions of the face and head, leaving the user in direct control of navigation and speech content. The avatar shows appropriate behavior based on the current situation and user input. One can think of this as control at a higher level than in current avatar-based systems. This approach starts to addresses the following problems:

- **Control complexity:** The user manipulates a few high-level parameters, representing the user's current intention with respect to conversational availability, instead of micromanaging every aspect of animating a human figure.

- **Spontaneous reaction:** The avatar shows spontaneous and involuntary reactions towards other avatars, something that a user would not otherwise initiate explicitly.

- **Discrete user input:** By having the avatar update itself, carry out appropriate behaviors and synchronize itself to the environment, the gap between meaningful occurrances of user input or lag times is bridged to produce seamless animation.

- **Mapping from user space into Cyberspace:** The user and the user's avatar reside in two drastically different environments. Direct mapping of actions, such as projecting a live image of the user on the avatar's face, will not produce appropriate avatar actions. Control at an intentional level and autonomy at the level of involuntary communicative behaviors may however allow the avatar to give the cues that are appropriate for the virtual situation.

3. HUMAN COMMUNICATIVE BEHAVIOR

In order to automate communicative behaviors in avatars, one has to understand the basic mechanisms of human to human communication. A face-to-face conversation is an activity in which we participate in a relatively effortless manner, and where synchronization between participants seems to occur naturally. This is facilitated by the number of channels or modalities we have at our disposal to convey information to our partners. These channels include the words spoken, intonation of the speech, hand gestures, facial expression, body posture, orientation and eye gaze. For example, when giving feedback one can avoid overlapping a partner by giving it over a secondary channel, such as by facial expression, while receiving information over the speech channel (Argyle and Cook 1976). The channels can also work together, supplementing or complementing each other by emphasizing salient points (Chovil 1992, Prevost 1996), directing the listener's attention (Goodwin 1986) or providing additional information or elaboration (McNeill 1992, Cassell forthcoming). When multiple channels are employed in a conversation, we refer to it as being multimodal.

The current work focuses on gaze and communicative facial expression mainly because these are fundamental in establishing and maintaining a live link between participants in a conversation. The use of gesture and body posture is also very important, but the required elaborate articulation of a human body is beyond the scope of this current work and will be pursued later.

To illustrate what is meant by communicative behavior, the following section describes a scenario where two unacquainted people meet and have a conversation. The behaviors employed are referenced to background studies with relevant page numbers included.

Paul is standing by himself at a cocktail party, looking out for interesting people. Susan (unaquainted with Paul) walks by, mutual glances are exchanged, Paul nods smiling, Susan looks at Paul and smiles [distance salutation] (Kendon 1990, 173; Cary 1978, 269) Susan touches the hem of her shirt [grooming] as she dips her head, ceases to smile and approaches Paul (Kendon 1990, 186, 177). She looks back up at Paul when she is within 10' [for initiating a close salutation], meeting his gaze, smiling again (Kendon 1990, 188; Argyle 1976, 113). Paul tilts his head to the side slightly and says "Paul", as he offers Susan his hand, which she shakes lightly while facing him and replying "Susan" [close salutation] (Kendon 1990, 188, 193). Then she steps a little to the side to face Paul at an angle (Kendon 1990, 193; Argyle 1976, 101). A conversation starts.

During the conversation both Paul and Susan display appropriate gaze behavior, such as looking away when starting a long utterance (Kendon 1990, 63; Argyle 1976, 115; Chovil 1992, 177; Torres et al. 1997), marking various syntactic events in their speech with appropriate facial expressions, such as raising their eyebrows while reciting a question or nodding and raising eyebrows on an emphasized word (Argyle 1973; Chovil 1992, 177; Cassell et al. 1994a), giving feedback while listening in the form of nods, low "mhm"s and eyebrow action (Chovil 1992, 187; Schegloff 1968; Cassell et al. 1994a) and finally giving the floor to the other person using gaze (Kendon 1990, 85; Chovil 1992, 177; Argyle 1973; Argyle 1976, 118).

Speakers choose conversational partners but do not choose to raise their eyebrows along with an emphasis word, or to look at the other person when giving over the floor. Yet we attend to these clues as listeners, and are thrown off by their absence. In BodyChat, we have implemented these communicative behaviors as a function of their volitional status. That is, we distinguish between user choices, such as who to speak to and when to end the conversation, and body behaviors, such as meeting the gaze of somebody one has chosen to converse with.

4. RELATED WORK

Embodiment in Distributed Virtual Environments has been a research issue in systems such as MASSIVE at CRG Nottingham University, UK where various techniques and design issues have been proposed (Benford et al. 1995). There it is made clear that involuntary facial expression and gesture are important but hard to capture. Avatar autonomy however is not suggested. Popular Internet based chat systems that connect a number of users to graphical multi-user environments, such as the early WorldChat from Worlds Inc., have shown that graphical representation of users is a compelling alternative to purely text-based systems. However these systems have not been able to naturally integrate the graphics with the communication that is taking place.

Studies of human communicative behavior have seldom been considered in the design of believable avatars. Significant work includes Judith Donath's Collaboration-at-a-Glance (Donath 1995), where on-screen participant's gaze direction changes to display their attention, and Microsoft's Comic Chat (Kurlander et al. 1996), where illustrative comic-style images are automatically generated from the interaction. In Collaboration-at-a-Glance the users lack a body and the system only implements a few functions of the head. In Comic Chat, the conversation is broken into discrete still frames, excluding possibilities for things like real-time backchannel feedback and subtle gaze behaviors.

Creating fully autonomous agents capable of natural multi-modal interaction entails integrating speech, gesture and facial expression. By applying knowledge from discourse analysis and studies of social cognition, systems like Animated Conversation (Cassell et al. 1994b) and Gandalf (Thórisson 1997) have been developed. Animated Conversation renders a graphical representation of two autonomous agents engaged in conversation. The system's dialogue planner generates the conversation and its accompanying communicative signals, based on the agent's initial goals and knowledge. Gandalf is an autonomous agent capable of carrying out a conversation with a user and employing a range of communicative behaviors that help to manage the conversational flow. Both these systems are good examples of discourse theory and studies of human communication applied to computational environments, but neither is concerned with representations of user embodiment and issues of avatar control.

The real-time animation of lifelike 3D humanoid figures has been greatly improved in recent years. The Improv system (Perlin and Goldberg 1996) demonstrates a visually appealing humanoid animation and provides tools for scripting complex behaviors, ideal for agents as well as avatars. Similarly the Humanoid 2 project deals with virtual actors performing scripts as well as improvising role-related behavior (Wavish and Connah 1997). However, automatically generating the appropriate communicative behaviors and synchronizing them with an actual conversation between users has not been addressed yet in these systems. Real-time external control of animated autonomous actors has called for methods to direct animated behavior on a number of different levels such as in ALIVE (Blumberg and Galyean 1995) and in the OZ Project (Bates et al. 1991). In this sense, the goals of BodyChat are similar, but the set of behaviors is different. Here we focus on those behaviors that accompany language. We have also introduced, for the first time, a distinction between conversational Phenomena and Communicative Behaviors.

5. BODYCHAT

BodyChat is a system that demonstrates the automation of communicative behaviors in avatars. The system consists of a Client program and a Server program. Each Client is responsible for rendering a single user's view into the Distributed Virtual Environment (see figure 7). All users connected to the same Server see each other's avatars as a 3D model representing the upper body of a cartoon-like humanoid character. Users can navigate their avatars using the cursor keys, give command parameters to their avatar with the mouse and interact textually with other users through a two-way chat window.

5.1. User choices

The avatar's communicative behavior reflects its user's current intentions and the avatar's knowledge of communicative rules. The user's intentions are described as a set of control parameters that are sent from the user's Client to all connected Clients, where they are used to produce the appropriate behavior in the user's remote avatars. BodyChat implements three control parameters as described in Table 1.

Parameter	**Type**	**Description**
Potential Conversational Partner	Avatar ID	A person the user wants to chat with
Availability	Boolean	Shows if the user is available for chatting
Breaking Away	Boolean	Shows if the user wants to stop chatting

Table 1: Control Parameters that reflect the user's intention

The ***Potential Conversational Partner*** indicates whom the user is interested in having a conversation with. The user chooses a Potential Conversational Partner by clicking on another avatar visible in the view window. This animates a visual cue to the chosen Avatar that in turn reacts according to that user's *Availability*.

Availability indicates whether the user welcomes other people that show interest in having a conversation. This has an effect on the initial exchange of glances and whether salutations are performed that confirm the newcomer as a conversational partner. Changing Availability has no effect on a conversation that is already taking place. The user switches Availability ON or OFF through a toggle switch on the control panel (see Figure 7).

During a conversation, a user can indicate willingness to ***Break Away***. The user informs the system of his or her intention to Break Away by placing a special symbol (a forward slash) into a chat string. This elicits the appropriate diverted gaze, giving the partner a visual cue along with the words spoken. For example, when ready to leave Paul types "/well, I have to go back to work". The partner will then see Paul's avatar glance around while displaying the words (without the slash). If the partner replies with a Break Away sentence, the conversation is broken with a mutual farewell. If the partner replies with a normal sentence, the Break Away is cancelled and the conversation continues. Only when both partners produce subsequent Break Away sentences, is the conversation broken (Kendon 1990, Schegloff and Sacks 1973).

5.2. Generated behaviors

When discussing the communicative signals, it is essential to make clear the distinction between the *Conversational Phenomena* on one hand and the *Communicative Behaviors* on the other. Conversational Phenomena describe an internal state of the user (or avatar), referring to various conversational events. For example, a *Salutation* is a Conversational Phenomenon. Each Phenomenon then has associated with it a set of *Communicative Behaviors*, revealing the state to other people. For example, the Salutation phenomenon is associated with the *Looking*, *Head Tossing*, *Waving* and *Smiling* Behaviors.

The avatars in BodyChat react to an event by selecting the appropriate Conversational Phenomenon that describes the new state, initiating the execution of associated Communicative Behaviors. Essentially the avatar's behavior control consists of four tiers, where the flow of execution is from top to bottom (see Figure 1).

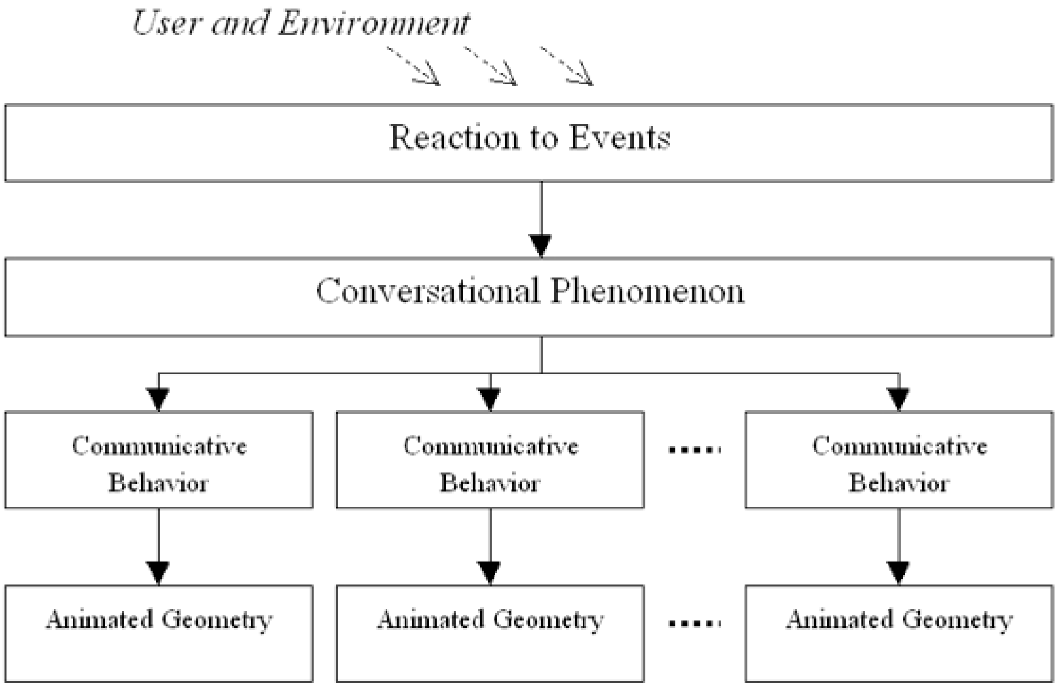

Figure 1: The avatar's behavior control consists of four tiers, where the flow of execution is from top to bottom.

The ***Reaction to Events*** tier defines the entry point for behavioral control. This tier is implemented as a set of functions that get called by the Client when messages arrive over the network or by the avatar as the environment gets updated. These functions are listed in Table 2. This tier is the heart of the avatar automation, since this is where it is decided how to react in a given situation. The reaction involves picking a Conversational Phenomenon that describes the new state of the avatar. This pick has to be appropriate for the situation and also reflect, as closely as possible, the user's current intentions.

Function	Event
ReactToOwnMovement	User moves the avatar
ReactToMovement	The conversational partner moves
ReactToApproach	An avatar comes within reaction range
ReactToCloseApproach	An avatar comes within conversational range
ReactToOwnInitiative	User shows interest in having a conversation
ReactToInitiative	An avatar shows interest in having a conversation
ReactToBreakAway	The conversational partner wants to end a conversation

ReactToSpeech	An avatar spoke
Say (utterance start)	User transmits a new utterance
Say (each word)	When each word is displayed by the user's avatar
Say (utterance end)	When all words of the utterance have been displayed

Table 2: The Behavior Control functions that implement the Reaction to Events

The Conversational Phenomena tier implements the mapping from a state selected by the Event Reaction, to a set of visual behaviors (see Table 3). This mapping is based on previous work in human communicative behavior.

Conversational Phenomena	Communicative Behavior
Approach and Initiation:	
Reacting	ShortGlance
ShowWillingnessToChat	SustainedGlance, Smile
DistanceSalutation	Looking, HeadToss/Nod, RaiseEyebrows, Wave, Smile
CloseSalutation	Looking, HeadNod, Embrace or OpenPalms, Smile
While chatting:	
Planning	GlanceAway, LowerEyebrows
Emphasize	Looking, HeadNod, RaiseEyebrows
RequestFeedback	Looking, RaiseEyebrows
GiveFeedback	Looking, HeadNod
AccompanyWord	Various
GiveFloor	Looking, RaiseEyebrows (followed by silence)
BreakAway	GlanceAround
When Leaving:	
Farewell	Looking, HeadNod, Wave

Table 3: The mapping from Conversational Phenomena to visible Behaviors

Finally, each Communicative Behavior starts an animation engine that manipulates the corresponding avatar geometry in order change the visual appearance.

5.3. Sample interaction

Overview

This section describes a typical session in BodyChat, illustrated with images showing the various expressions of the avatars. The images are all presented as sequences of snapshots that reflect change over time.

No interest

User A is scouting out the scene, seeking out someone interested in chatting. After awhile A spots a lone figure that is apparently not occupied. A clicks on the other avatar, choosing a potential conversational partner (see 5.1). The other Avatar reacts with a brief glance without a change in expression. This lack of sustained attention signals to A that the other user is not Available (see 5.1). The automated sequence of glances is shown in figure 2.

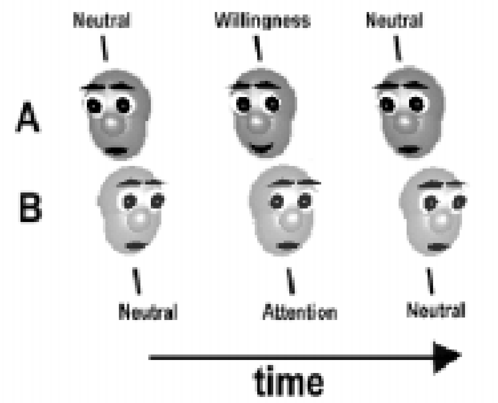

Figure 2: The sequence of glances when user A clicks on avatar B to express willingness to chat while user B is not available.

Partner found

User A continues to scout for a person to chat with. Soon A notices another lone figure and decides to repeat the attempt. This time around the expression received is an inviting one, indicating that the other user is Available. The automated sequence of glances can be seen in figure 3.

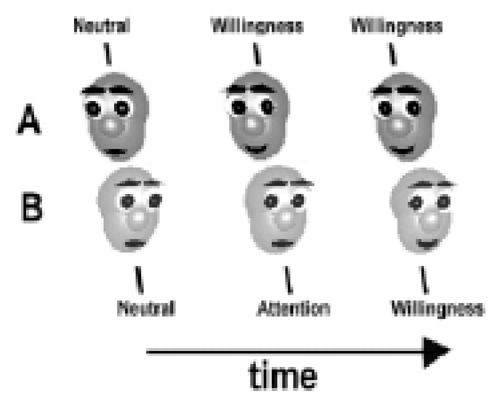

Figure 3: The sequence of glances when user A clicks on avatar B to express willingness to chat and user B is available

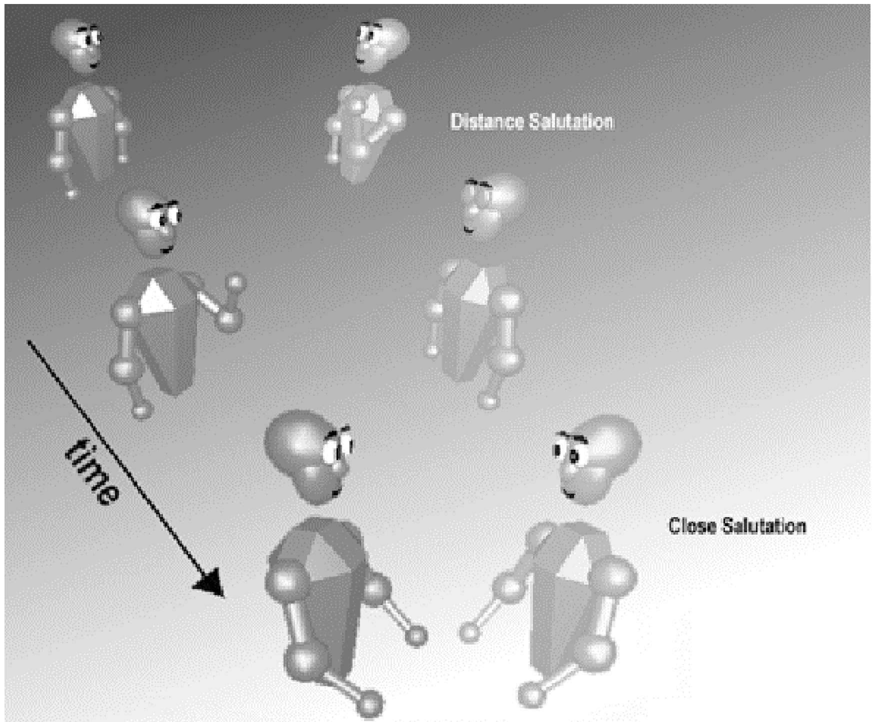

Figure 4: Avatars A and B exchange Distance Salutations when the system registers them as conversational partners. When they get within a conversational range, Close Salutations are exchanged.

Immediately after this expression of mutual openness, both avatars automatically exchange Distance Salutations to confirm that the system now considers A and B to be conversational partners. Close Salutations are automatically exchanged as A comes within B's conversational range. Figure 4 shows the sequence of salutations.

A conversation

So far the exchange between A and B has been non-verbal. When they start chatting, each sentence is broken down into words that get displayed one by one above the head of their avatar. As each word is displayed, the avatar tries to accompany it with an appropriate expression. An example of an animated utterance can be seen in figure 5.

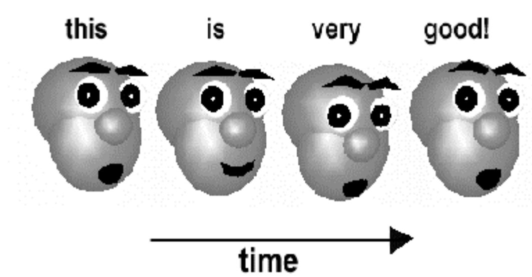

Figure 5: Some words are accompanied with a special facial expression. Here "very" is being emphasized with a nod. The exclamation mark elicits raised eyebrows at the end of the utterance.

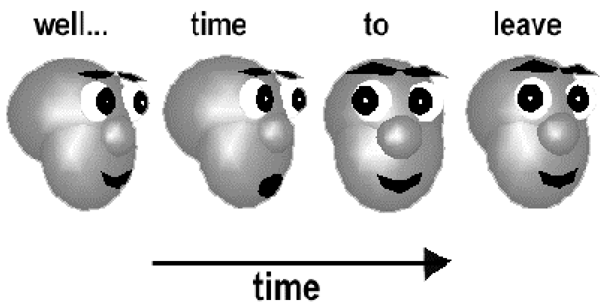

Figure 6: When the user marks a sentence as a Break Away utterance, the avatar displays diverted gaze while reciting the words to give subtle cues to the conversational partner.

Finally, after A and B have been chatting for awhile, A produces a Break Away utterance by placing a forward slash at the beginning of a sentence (see 5.1). This makes A's avatar divert its gaze while reciting the words as shown in figure 6. User B notices this behavior and decides to respond similarly, to end the conversation. The avatars of A and B automatically wave farewell and break their eye contact.

6. RESULTS

BodyChat presents a new approach that takes avatars from being a mere visual gimmick to being an integral part of a conversation, from allowing shear and mere co-presence to allowing embodied communication. The interaction between user choices and autonomous communicative behaviors allows the user to concentrate on high level control and locomotion, while depending on the avatar to convey the communicative signals that represent the user's communicative intentions.

Regarding the approach in general, a few limitations should be considered. The first thing to keep in mind is that although communicative non-verbal behavior adheres to some general principles, it is far from being fully understood. Any computational models are therefore going to be relatively simplistic and constrain available behavior to a limited set of displays devoid of many real world nuances. This raises concerns about the system's capability to accurately reflect the user's intentions under unforeseen circumstances or resolve issues of ambiguity. If the avatar makes a choice that conflicts with what the user had in mind, reliability is severely undermined and the user is left in an uncomfortable skeptical state. The balance between autonomy and direct user control is a really tricky issue.

7. FUTURE WORK

The issue of the relationship between autonomy and user control is far from trivial and presents many interesting problems, some of which are shared with task assisting software agents (Friedman and Nissenbaum 1997). The current work introduces autonomy as an approach to animating avatars. This invites further research, both to see how far we can take the autonomous behavior before the user no longer feels in control and, how we can integrate this technique with other methods and possibly other user input devices.

This work only starts to build a repertoire of communicative behaviors, beginning with the most essential cues for initiating a conversation. It is important to keep adding to the modeling of conversational phenomena, both drawing from more literature and, perhaps more interestingly, through real world empirical studies conducted with this domain in mind. Behaviors that involve more than two people have to be examined and attention should be given to orientation and the spatial formation of group members. The humanoid models in BodyChat are simple and not capable of carrying out detailed, co-articulated movements. In particular, the modeling of the arms and hands needs more work, in conjunction with the expansion of gestural behavior. It is to these behaviors that we will turn in our next version of BodyChat.

8. CONCLUSIONS

This paper has introduced a novel approach to the design and implementation of avatars, drawing from literature in context analysis, discourse theory, and autonomous communicating agents. It was argued that today's avatars merely serve as presence indicators, rather than actually contributing to the experience of having a face-to-face conversation. In order to understand the important communicative functions of the body, we relied on previous research on multi-modal communication among humans. We used that research to develop BodyChat, a system that employs those findings in the automation of communicative behaviors in avatars.

Because of the richness of involuntary behavior in a social situation, relying only on explicit user control will not exploit the function of embodiment in the construction of animated avatars. Regarding an avatar as a personal conversational agent that together with the user is capable of naturally initiating and sustaining a conversation provides a valuable perspective, contributing both to research on avatars, and on communicative autonomous agents.

Figure 7: Looking at another user's avatar in BodyChat

9. ACKNOWLEDGEMENTS

We would like to thank Bruce Blumberg and Carl Malamud for input. Thanks to the members of the Gesture and Narrative Language group and the Media Lab community for valuable discussions. This research was supported in part by the Media Lab *Digital Life* and *Things That Think* consortia.

10. REFERENCES

Anderson, D.B., Barrus, J.W., Brogan, D., Casey M., McKeown, S., Sterns, I., Waters, R., Yerazunis, W. (1996). Diamond Park and Spline: A Social Virtual Reality System with 3D Animation, Spoken Interaction, and Runtime Modifiability. Technical Report at MERL, Cambridge.

Argyle, M., Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge University Press.

Argyle, M., Ingham, R., Alkema, F., McCallin, M. (1973). The Different Functions of Gaze. *Semiotica.*

Bates, J., Loyall, A.B., Reilley, W.S., (1991). Broad Agents. *SIGART Bulletin*, 4 (2).

Benford, S., Bowers, J., Fahlen, L.E., Greenhalgh, C., Snowdon, D. (1995). User Embodiment in Collaborative Virtual Environments. *In Proceedings of CHI'95*, 242-249.

Blumberg, B. M., Galyean, T. A. (1995). Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments. *Proceedings of SIGGRAPH '95.*

Cary, M. S. (1978). The Role of Gaze in the Initiation of Conversation. *Social Psychology*, 41(3).

Cassell, J. (forthcoming). *A Framework For Gesture Generation And Interpretation*. In R. Cipolla and A. Pentland (eds.), *Computer Vision in Human-Machine Interaction*. Cambridge University Press.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., Stone, M. (1994b). Animated Conversation: Rule-based Generation of Facial Expression, Gesture & Spoken Intonation for Multiple Conversational Agents. *Proceedings of SIGGRAPH '94.*

Cassell, J., Stone, M., Douville, B., Prevost, S., Achorn, B., Steedman, M., Badler, N., Pelachaud, C. (1994a). Modeling the Interaction between Speech and Gesture. *Proceedings of the Cognitive Science Society Annual Conference*

Chovil, N. (1992). Discourse-Oriented Facial Displays in Conversation. *Research on Language and Social Interaction*, 25, 163-194.

Donath, J. (1995). The Illustrated Conversation. *Multimedia Tools and Applications*, 1, 79-88.

Fiedman, B., Nissenbaum, H. (1997). Software Agents and User Autonomy. *Proceedings of Agents'97*, 466-469.

Goodwin, C. (1986). Gestures as a Resource for the Organization of Mutual Orientation. *Semiotica*, 62(1/2).

Kendon, A. (1990). Conducting Interaction: Patterns of behavior in focused encounters. Cambridge University Press. New York.

Kendon, A. (1992). The negotiation of context in face-to-face interaction. In A. Duranti and C. Goodwin (eds*.), Rethinking context: language as interactive phenomenon*. Cambridge University Press. New York.

Kurlander, D., Skelly, T., Salesin, D. (1996). Comic Chat. *Proceedings of SIGGRAPH '96.*

McNeill, D. (1992*). Hand and Mind: What Gestures Reveal about Thought*. University of Chicago.

Perlin, K., Goldberg, A. (1996). Improv: A System for Scripting Interactive Actors in Virtual Worlds. *SIGGRAPH 1996 Course Notes #25*.

Prevost, S. (1996). Modeling Contrast in the Generation and Synthesis of Spoken Language. *In Proceedings of ICSLP '96*.

Schegloff, E. (1968). Sequencing in Conversational Openings. *American Anthropologist*, 70, 1075-1095.

Schegloff, E., Sacks, H. (1973). Opening up closings. *Semiotica*, 8, 289-327.

Thórisson, K. R. (1997). . Gandalf: An Embodied Humanoid Capable of Real-Time Multimodal Dialogue with People. *Proceedings of Agents'97*, 536-537.

Thórisson, K.R., Cassell, J. (1996) Why Put an Agent in a Human Body: The Importance of Communicative Feedback in Human-Humanoid Dialogue. (abstract*) In Proceedings of Lifelike Computer Characters '96*, Snowbird, Utah, 44-45.

Torres, O., Cassell, J., Prevost, S. (1997). Modeling Gaze Behavior as a Function of Discourse Structure. *In Proceedings of the First International Workshop on Human-Computer Conversations 1997*. Bellagio, Italy.

Wavish, P., Connah, D. (1997). Virtual actors that can perform scripts and improvise roles. *Proceedings of Agents'97*, 317-322.