The Role of Eye Gaze in Avatar Mediated Conversational Interfaces

R. Alex Colburn
Michael F. Cohen
Steven M. Drucker

7/31/2000

Technical Report
MSR-TR-2000-81

Microsoft Research
Microsoft Corporation
One Microsoft Way

Keywords: Animation, Character Behavior, Facial Animation, HCI (Human-Computer Interface), Multi-User, Networked Apps.

As we begin to create synthetic characters (avatars) for computer users, it is important to pay attention to both the look and the behavior of the avatar's eyes. In this paper we present behavior models of eye gaze patterns in the context of real-time verbal communication. We apply these eye gaze models to simulate eye movements in a computer-generated avatar in a number of task settings. We also report the results of an experiment that we conducted to assess whether our eye gaze model induces changes in the eye gaze behavior of an individual who is conversing with an avatar.

Avatar Mediated Conversational Interfaces

R. Alex Colburn
Michael F. Cohen
Steven M. Drucker

7/31/2000

Technical Report
MSR-TR-2000-81

Microsoft Research
Microsoft Corporation
One Microsoft Way

ABSTRACT
As we begin to create synthetic characters (avatars) for computer users, it is important to pay attention to both the look and the behavior of the avatar's eyes. In this paper we present behavior models of eye gaze patterns in the context of real-time verbal communication. We apply these eye gaze models to simulate eye movements in a computer-generated avatar in a number of task settings. We also report the results of an experiment that we conducted to assess whether our eye gaze model induces changes in the eye gaze behavior of an individual who is conversing with an avatar.

Keywords
Animation, Character Behavior, Facial Animation, HCI (Human-Computer Interface), Multi-User, Networked Apps

INTRODUCTION

The eyes may be the windows into the soul, but they are also much more. Our eyes express our emotions and intentions and they help us direct attention. Social aspects of eye gaze have been mentioned as early as on clay tablets from Mesopotamia. Cultural norms dictate when, for how long and in what situations it is appropriate to gaze into another person's eyes. Psychologists [19] have found high correlations between changes in eye gaze and specific conversational actions. For example, speakers may shift their gaze towards a listener as they get to the end of a thought. Thus, as we begin to create synthetic characters (avatars) for computer users, either to act on our behalf or for use in automated interfaces, it will be important to pay attention to both the look and the behavior of the avatar's eyes.

In this paper we present behavior models of eye gaze patterns in the context of real-time verbal communication. We apply these eye gaze models to simulate eye movements in a computer-generated avatar in a number of task settings. We also report the results of an experiment that we conducted to assess whether our eye gaze model induces changes in the eye gaze behavior of an individual who is conversing with an avatar.

Our work is, in part, motivated by results of studies on the advantages and disadvantages of video-based teleconferencing [11,12,18,23,25,34]. These studies have identified a number of limitations inherent to this means of communication, ranging from bandwidth and other technical issues to social issues. A limitation that is of particular interest in our context is the teleconference participants' lack of opportunity to connect via eye gaze. More specifically, if an individual who participates in a video-based teleconference watches the other participant's image on a monitor, they cannot (without special mirrors) also be looking at the camera. Thus, two individuals will not be able to gaze into each other's eyes during a video-based teleconference. As noted by Ishii *et al*. [18] and Dourish *et al*. [10], this problem may diminish over time as users adapt their understanding of gaze awareness by reinterpreting gaze direction in the in this new context. Mutual eye gaze difficulties are even more apparent in multi-person settings. The literature reports that this limitation leads to less satisfying and thus perhaps to less productive conversations.

A further motivation for the work presented here is a desire to create better user interfaces. It has been hypothesized that if the computer interface includes an anthropomorphic graphical representation, the user can interact more naturally with the computer and hence the user interface can be simplified. Although such representations have been tried for a number of years, so far they have not reached their goals.

The Role of Eye Gaze in Conversation

There is a broad body of literature on the role of eye gaze in human interactions. This literature investigates questions such as whether people tend to look at each other more when speaking versus listening, and how eye gaze is used to regulate turn taking in conversations. Most of these studies focus on eye gaze patterns in one-on-one face-to-face conversations. The best single source is "Gaze and Mutual Gaze" by Argyle and Cook [1] in which they bring together many individual studies into a common context.

Results from the studies suggest that eye gaze helps control the flow of turn taking in conversations. For example, the person who is listening uses eye gaze to indicate whether they are paying attention, while the person who is speaking uses it to track whether the listener is still engaged in the conversation. Eye gaze is also a means to create subtle breaks in conversation to allow for a change of speaker [4, 19].

Kendon [19] reports that a typical pattern of interaction when two people converse with each other consists of the listener maintaining fairly long gazes at the speaker, interrupted by short glances away. In contrast, the speaker makes longer gazes away from the listener with shorter gazes at the listener. In addition, particularly in longer utterances (more than 5 seconds in Kendon's work), the speaker tends to look at the listener as he or she is about to finish the utterance. It is hypothesized that this serves as a mechanism to "hand over the floor" to the listener. The usefulness of this last observation in an application will depend on whether there is any knowledge of the future, in other words, whether the avatar is pre-scripted or whether action and speech are generated on the fly.

Work by Novick et al. [22] supports Kendon's results. In addition, Novick et al. report that a "mutual break" pattern occurs in 42% of turns at speech. The authors define a mutual break as a pattern in which the speaker looks towards the listener as an utterance comes to a close. Gaze is then momentarily mutual, after which the listener breaks mutual gaze and begins to speak. Novick et al. furthermore found that in 29% of verbal exchanges a "mutual hold" took place, that is, the listener maintained gaze for a short time after beginning to speak.

Mutual gaze occurs when two individuals look simultaneously into each other's eyes. Argyle [1] discusses findings that indicate the length of mutual gaze varies depending on the following:

- age (longer in adults than young children)
- gender (longer in women than men)
- the distance between speakers (mutual gaze is longer the more widely separated the conversants are) [17]
- culture

In adults in western cultures, mutual gaze tends to be rather short with a mean duration of 1 to 2 seconds.

There are considerable cultural differences related to eye gaze patterns. Elzinga [14] reported that study subjects from Japan had "more frequent and shorter lasting other-directed gazes" than study participants from Australia. He also found that English-speaking study participants showed a tendency to look at the other person to signal speaker transitions, while Japanese speakers did not. Another study found that Swedes gaze at their counterpart more than English (50% vs. 38% of the time) [1]. Others have found similar differences across other cultural groups.

Novick et al. [22] also used the results of their eye gaze study to build a computer based simulation from their findings. Comparisons of the simulated eye gaze transcripts to eye gaze patterns recorded from real conversations showed a high degree of fidelity in their models. In this paper, we describe a similar simulation model that we have developed and applied to a graphical avatar representation. We have relied heavily on the literature to develop the eye gaze model we will demonstrate. We will discuss experiments in the use of the eye gaze models in one-on-one conversation, as part of an offline user interface, and in avatar based virtual group teleconferences.

Eye Gaze in Computer Applications

A considerable body of work has been carried out to track a computer user's eye gaze and subsequently use this information as an input modality, typically to replace mouse motion as a pointing device [16,31]. This is of use by, for example, disabled persons or for applications that require a hands-free setting. While we track the user's gaze in our study and draw on this information as input to the simulated gaze model for the avatar, gaze tracking as an input modality is not the focus of this study.

Little work has been reported on the topic of this paper, that is, simulating eye gaze patterns to drive the eyes on computer generated characters, or avatars. A few studies do employ eye gaze input to indicate where participants are looking in a virtual collaborative setting. For example, Vertegaal et al. [34] provide no direct eye gaze feedback from the participant's graphical embodiment, but rather place a "lightspot" in the virtual scene to indicate each participant's gaze target. [32] rely on the gaze direction of teleconference participants to choose between multiple static images of each participant to display to other participants. Each image indicates a different viewing direction. Thus, each teleconference participant receives some feedback about the direction in which other participants are looking. In contrast to the above work we will be using the eye gaze model to directly drive an avatars gaze.

Avatars

In recent years, avatars have been used in a variety of interactive computer-based experiences. The popular use of the term "avatar" originates from Neil Stephenson's novel Snowcrash [28] in which graphical representations of characters converse with each other in a virtual world. A variety of computer systems have incorporated avatars for graphical chats (Worlds Away, Worlds Inc) [36,24] and graphical role playing games (Ultima Online and Everquest) [15,30]. In these systems, a graphical representation for characters offers users the opportunity to customize their virtual appearance and behavior. Studies have shown that users of avatars mimic real life behavioral qualities when interacting with other users of avatars [5,27]. Perhaps most importantly, virtual conversations have shown to benefit from the use of gestures, from adjusting the physical distance between two avatars, and from adjusting the avatars' relative orientation to each other [5,27].

As noted, a limitation of video-based teleconferences is the participants' lack of opportunity to connect via eye gaze. It has been suggested that the use of avatars instead of video for person-to-person voice-based interaction may be able to address this limitation. A particular advantage of avatars over video is the avatar's capability to connect via eye gaze. As the visual realism and behavior of avatars increases, they may become accepted modes for many types of real-time interaction.

Cassell et al. [6,9,29] have explored automating expression and gesture in computer driven conversational agents. Cassell and Thorisson [8] have worked on trying to embody agents with physical human characteristics including eye-gaze and nodding. Their work was also primarily devoted to computer representations that were driven solely by artificial intelligence. Cassell and Vilhjalmsson [7] build upon some of this work to create "fully embodied, conversational avatars". They reason as we do, that the user's behavior in the real world differs from the avatar's behavior in the virtual world and thus must include some autonomous models. Their work has focused on full-body behaviors but has included eye-glances as a means for determining whether other users are willing to be included in a conversation.

The Eye Gaze Model

The eye gaze models described below are based primarily on the psychological literature. We have not yet attempted to model differences in age, gender or culture, although the model could be easily extended to do so. The specific model is based on average values for western adults. These values are supplemented with the authors' own informal observations in cases in which the psychological literature did not provide necessary detail.

Hierarchical State Machine

The stochastic eye gaze models we have developed are summarized in the hierarchical state machine diagrams shown in Figures 1 through 4. Figures 1 and 2 represent the model for one-on-one conversation, while Figures 3 and 4 represent multi-party interactions.

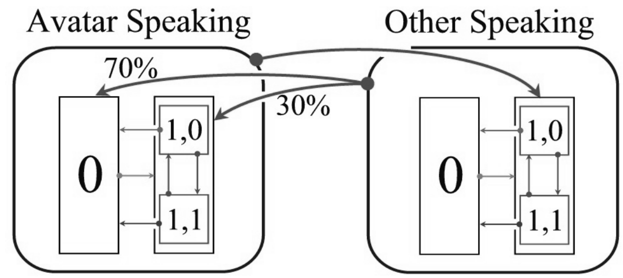


Figure 1: Hierarchical state diagram for two people talking.

In the one-on-one interaction, the two participants in the conversation are labeled avatar and other. The avatar is the one controlled by the model. The avatar may be a pre-scripted character, an automaton, or a character that is puppeted in real-time by a user. Likewise, the other may be an avatar or a live person whose eyes are either tracked or for whom there is an assumed eye gaze pattern. In our experiment described later, the interaction takes place between a loosely puppeted avatar and a live person.

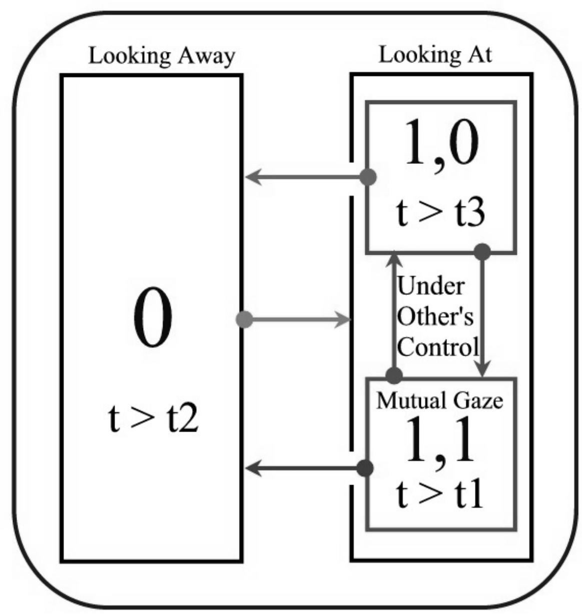


Figure 2: Detail of eye gaze state transitions.

The highest level of our eye gaze model is depicted in Figure 1. Figure 1 displays two states depending on whom is talking. Transitions between states are triggered when there is a change in which participant is talking. If no conversation takes place for a short amount of time, the model remains in its most recent state. If both participants talk simultaneously, the model moves to, or remains in, the "other talking" state.

Built into each major state are lower level states. Each of these sub-states is labeled with either one or two numbers. The (0) state indicates that the avatar is gazing away from the other. State (1,0) indicates the avatar is looking at the other, but that the other is looking away. State (1,1) is one of mutual gaze. The lack of a second index for state 0 is due to the fact that the avatar cannot "see" the other if the other is looking away. The avatar always looks at the other when the other begins to speak. When the avatar begins to speak it will stochastically look at the other only some of the time. To mimic results reported in the literature, we have the avatar look at the other 30% of the time when the avatar begins to speak.

Timing of Transitions Between Sub-States

The decision to transition the avatar's eye gaze, that is whether it looks at or away from the other, is triggered primarily by the passing of time in the current sub-state. One exception is the occurrence and timing of "vertical" transitions, that is transitions between states (1,0) and (1,1). These depend solely on the other glancing at or away from the avatar and are, thus, not under the control of our model.

In the figures, the passing of time in a particular sub-state is denoted as t. Each time a new sub-state is entered t is set to 0. When a sub-state is entered, in addition to t being set to zero, a transition time is chosen for when the sub-state will be left. Transitions are triggered when t surpasses the transition time. Transition times, denoted as t1, t2 and t3 in the state diagrams, are chosen stochastically. The means and variances of these transition times depend on the specific sub-state. For example, the mutual gaze sub-state has shorter transition times than either of the other sub-states. The transition times also depend on which major state the avatar is in. In general, transition times are longer in the listening state than in the speaking state.

Perhaps the most important transition is the one that provides for the breaking of mutual gaze. The timing of this transition depends on a stochastic value that is chosen each time state (1,1) is entered. According to the literature, this value has a mean of 2 seconds. There is disagreement, however, over whether variations in this mean are more attributable to specific individuals or to the dynamics between a particular pair of people [1]. As mentioned, there is evidence that the length of gaze increases as physical proximity between individuals decreases. Put differently, individuals tend to shy away from looking directly into another person's eyes, the closer they are [1]. Gaze is also averted more with increased cognitive activity [2] Finally, we have observed that gestures, such as a nod, increase the duration of mutual gaze. In our model, a transition from (1,0) to (1,1), indicating that the other has just looked at the avatar, invokes a nod from the avatar. In our model the timer does not begin until the nod ends, increasing the overall length of t1.

Breaking gaze while the other is not looking occurs less often than during mutual gaze. The time to this transition is represented by t3 in Figure 2. Although no pertinent values are reported in the literature, one can deduce this time to have a mean of approximately 6 seconds while listening and 4 seconds while speaking.

The length of time a participant looks away from the other person during a conversation, t2 in Figure 2, is highly dependent on whether the participant is speaking or listening. Listeners are more responsible for looking at the speaker, to show that they are paying attention. In contrast, the speaker often only periodically checks in to see that the listener is still attentive. In Kendon's work, the mean time a participant looks away from the other conversant was about 4 seconds. Kendon mentions that there is a significant difference in the amount of time a speaker versus a listener looks away. However, from his tabulated data it is difficult to separate these two situations. We chose a mean of 3 seconds while in listening mode and of 4 seconds while in speaking modes.

As one might expect, each of these timing values varies both within a particular person and across persons and pairs of persons. Kendon has hypothesized that pairs of individuals come to a tacit agreement each time they converse and adapt their gaze patterns to each other. However, he found the general pattern and proportions to be fairly consistent. Thus, while some individuals and pairs of individuals tended to move their eyes less often, they still maintained similar patterns proportionally.

Gaze Direction

Our model distinguishes between whether the other is a person or another avatar. In the case of interactions with a human, when the avatar looks at the other its gaze is directed at the virtual camera that is used to render the avatar. When the conversation partner is another avatar, the gaze is directed towards the other avatar's eyes. When the avatar looks away from the other, its eyes are directed at a location that is chosen stochastically. Studies have shown that patterns of gaze often linger on the mouth and ears of the other conversant or they follow other objects of interest in the environment [1]. We did not implement such a sophisticated away-gaze positioning model.

Multi-party Interactions

An eye gaze model for multi-party interactions is slightly more complex than a model for one-on-one interaction. There are now opportunities to look at more than one non-speaking person and to have them look back, as well as to look at the speaker or at no one. At the highest level (Figure 3) one new transition is needed from other speaking back to itself, triggered when there is a change in speaker between two non-avatar participants. Within the other speaking state there are now three targets for gaze: away from anyone, towards the speaker, and towards a non-speaker. Figure 4 illustrates the details of possible transitions between these states. Transitions are made in a two-step process. As before, a transition time is stochastically chosen each time a state is entered and t is set to 0. The means of the transition times are given in each state in the Figure 4. When t passes the chosen transition time, next, a new state is chosen stochastically from all available outward flowing arrows. For example, when looking at a non-speaker, there is a 60% chance the next gaze will be at the speaker, a 30% chance of looking at another non-speaker, and a 10% chance of looking away from anyone. The specific percentage assignments have been drawn form personal observations as there is little literature on such group interaction and eye gaze.

As in the one-on-one case, when the speaker is being looked at and it shifts its gaze to look at the avatar (the downward green arrow transition between states (1,0) and (1,1)) a gesture such as a nod is often (70% of the time) triggered to acknowledge the entry

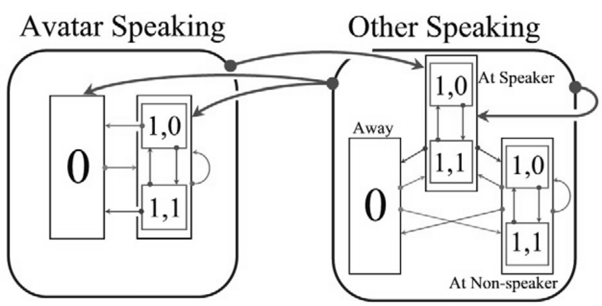


Figure 3: Hierarchical state diagram for multi-party talking.

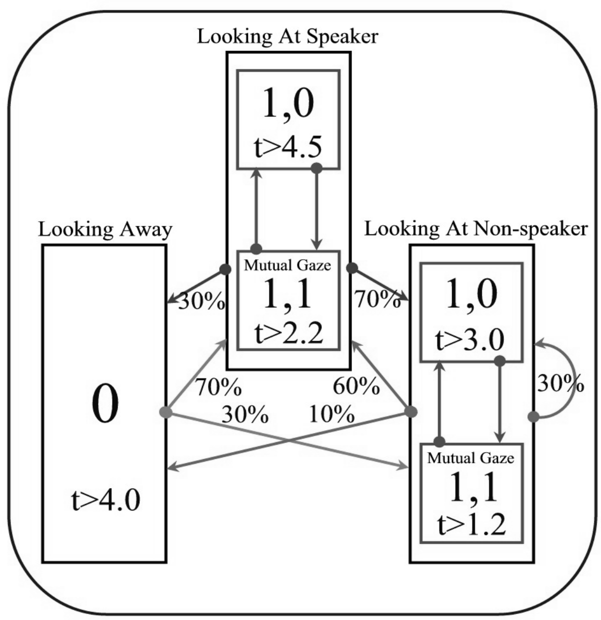


Figure 4: Detail of eye gaze state transitions and timings.

into a mutual gaze state[1]. Proximity likely plays an even larger role in group settings than in one-on-one conversation. Those closer together spend less time gazing at one another compared to those further away. This is implemented by simply adding a multiplier on the transition times inversely proportional to distance.

Applications and Experiments

One-on-one Avatar-Human Interaction

A first question that arises is whether users interacting with an avatar will act in a way that resembles interaction with a human. Or will the knowledge that this is not a real person counteract natural reactions. More specifically, in the context of the topic of this paper, will the use of an eye gaze model elicit more natural eye gaze patterns in a user? To investigate this question we conducted the following experiment.

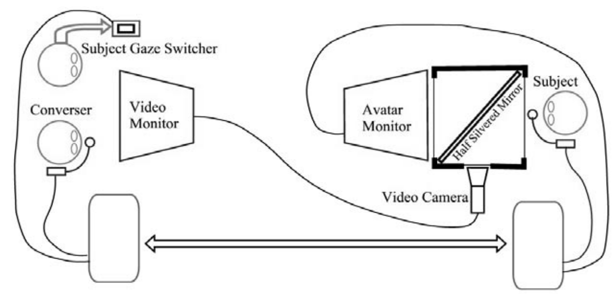


Figure 5: Diagram of experimental setup.

Figure 6: Photographs of the experiment in action.

The experimental system is illustrated in Figures 5 and 6. A subject sits in front of a monitor depicting the front facing

[1] The reader can try this the next time addressing a group. By shifting your eyes to a listener while speaking you can almost force a small nod in response.

avatar head and shoulders (Figure 6). The subject views the monitor through a box containing a half-silvered mirror. This allows a video camera to, via the mirror, directly face the user. The subject speaks via a headset and microphone to a conversant who is part of the experimental team sitting in front of a video monitor viewing the subject. A second experimenter watches the monitor and records when the subject is gazing at the avatar and when they are not. The microphones of the subject and conversant are fed into computers that record when each is speaking based on a simple volume threshold. The microphone input along with the subject's eye gaze if fed into the eye gaze model described above to drive the eye gaze of the avatar. The voice of the experimenter also drives the avatar's mouth.

The subject and experimenter conduct a casual conversation for 9 minutes on topics such as the Y2K problem, vacation plans, etc. The experimenter was asked to try to create a give and take conversation with about equal speaking and listening roles. Other than that, there was no specific task. During the nine minutes the subject was shown three 3-minute segments in random order. One 3-minute segment had a blank screen; one had an avatar without the eye gaze model that simply looked straight ahead at the subject, and one 3-minute segment had the eye gaze model functioning. At the end of the experiment, subjects were asked for general comments about the avatar, and were asked whether they noticed any changes in the way the avatar shifted its eye gaze during the experiment.

The main question of interest is whether there were detectable differences in the eye gaze patterns of the subject between the avatar without the eye gaze model and the one with the eye gaze model driving the avatar's eyes. If differences are found, were they towards a more "natural" pattern as found in the literature? Notably, only one subject noticed a change in the avatars eye gaze pattern. Thus, if any changes found in the subject's eye gaze were driven subliminally.

Measures of interest are what percentage of time did the subject look at the avatar while speaking and similarly what percentages of the time did they look while listening. Argyle [1] reports averages 41% and 75% respectively. Also, to account for differences across individuals, did the ratio of these percentages change based on the eye gaze model the avatar used? One would expect in a natural setting to see more other-directed gaze while listening than speaking.

To answer these questions, the speaking data were first filtered to extract utterances of more than 4 seconds. Short gaps in speech less than one second were first deleted since these generally represented gaps between words and sentences. Then, utterances of less than 4 seconds were eliminated since these corresponded primarily to "uh-huh" type utterances typical during speech. The eye gaze pattern was then intersected with the remaining utterances to determine the fraction of time the subject was gazing at the avatar during speaking and listening, broken down by which condition was active (no avatar, no eye gaze model, with eye gaze model). A typical plot of speech and eye gaze events is shown below.

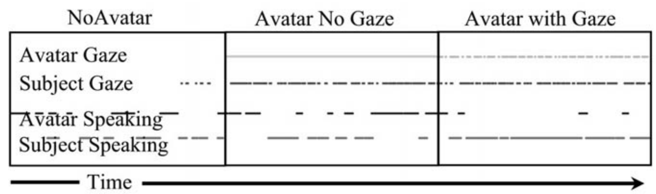


Figure 7: Typical plot of speech and eye gaze events.

A total of 20 subjects participated in the study drawn from a convenient population of computer scientists and administrative aides. The subjects did not know the purpose or topic of the study. The data for a subject was eliminated if there were less than 20 seconds of either the subject or conversant speaking during each condition.

Not surprisingly the subjects looked at the center of monitor more when there was an avatar vs. when there was nothing to look at; 38% vs. 64% of the time. Little change was detected in the eye gaze patterns of the subjects while they were speaking with an avatar present. In both cases they looked at the avatar about 56% of the time. The results did show an increase in the amount of time the subject looked at the avatar while listening when the gaze model was active vs. when there was no gaze model. The increase was from 67% to 74% of the time. This difference was surprisingly consistent. T-Test results were 0.1 indicating a 90% chance that a significant difference was detected. There was also a difference in the ratio of time spent looking while listening vs. looking while speaking. As expected in both cases the ratio was above 1.0. The ratios were 1.2 without the eye gaze model and 1.3 with the model. These results were also quite consistent across subjects. A T-test of .17, although not conclusive is marginally significant. It is of interest that we found these results even though only one subject had consciously noted any change in the eye gaze pattern even when asked specifically if they had noticed it. Thus any detected differences can be attributed to subliminal behavior.

Clearly more subjects would help disambiguate the results. We also suspect the subjects looked at the avatar more often than they might a real person. The average while speaking was 56% of the time versus 41% reported in the literature [1]. The averages for while listening were consistent with the literature. This increase in time looking while speaking may be because there is less shyness looking at a video monitor than a real person, or may be that the subjects suspected they were supposed to be looking (they looked 38% of time when there was no avatar to see). Many subjects also complained that the lip-sync was not accurate and was distracting. We would hope that more realistic avatars with more timely and accurate lip movement would generate more consistent results.

Virtual Conferencing

A second system in which the eye gaze model was used is in a virtual conferencing setting. In a one-on-one phone call we feel obliged as we listen to, say every few seconds, “uh-huh” or give some other verbal acknowledgement that we are listening and perhaps signal our state of agreement with the speaker. In conference calls of more than 3 people, this type of group awareness is lacking to an even greater extent. It is difficult to know who is listening, who is agreeing, when it is your turn to speak. Video based teleconference systems have been developed for many years to try to over come these and other problems. However, video based systems are expensive and require special rooms. They cannot afford such visual cues as eye gaze without additional cameras and special setups since if one is looking at the monitor, one cannot (without a special setup as in the previous experiment) also be looking at the camera. In addition, for multiple participants one would need a separate camera/monitor setup per participant. Such systems have been explored [26] but these are too cumbersome to be of general use.

In one such system, Muhlbach et al [21] found that eye contact, although not a critical factor in one-on-one conversation, was very important in multi-party communication. Without eye contact “there was significantly less feeling of being looked at, more uncertainty as to who was being addressed (both in terms of knowing that oneself is addressed and that someone else is addressed), and more uncertainty about knowing who was listening and about observed difficulties in turn taking.”

Avatar based systems can provide many of the kinds of social cues needed to provide group awareness including eye gaze albeit with the loss of some fidelity in non-verbal communication. We will be reporting the details of the avatar-based system along with study findings in another paper. Here we will briefly demonstrate the use of the eye gaze model described in section 3.1 above within this context.

Figure 8: concept study of a virtual conference room application.

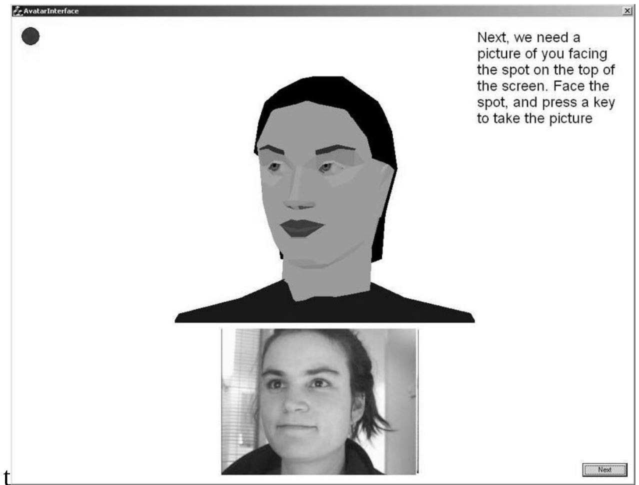
Figure 9: Current virtual conference room prototype.

Our eventual goal is to build a system that might look like Figure 8. This image is from a concept study that suggests including a graphical representation of the conversants, plus a speech-to-text transcript on the left and possible relevant document icons on the right. A view of our current virtual teleconferencing system is shown in Figure 9. The system consists of a standard PC computer console with camera and microphone for each (of currently 6) participants, each in their own office. Voice is transmitted over an ordinary teleconference call. Each participant is represented by their own avatar seated around a 3D virtual conference table. In addition to the eye gaze model, we are developing models for the attention and mood of each participant that direct the avatars when and where to face their bodies and heads, when to nod, etc. Each participant also has a personal virtual cinematographer [13] selecting a sequence of camera positions to best show the conference. Thus, participants are freed to simply watch and talk and allow their avatars to act on their behalf. How this “hands-off puppeting” of the avatars effects conversation flow is a topic of further study.

The eye gaze model outlined above from subjective observation appears to fulfill most of its goals. It clearly helps direct attention to the speaker. It helps participants to know who is being addressed and who is listening. We will leave it to the readers’ assessment for now to judge the effectiveness based on the accompanying video. More quantitative analyses of these questions will follow in a further study.

Avatar as User Interface

Eye gaze was also used as part of an avatar based user interface. The application in this case is a system for automatically creating avatars that look like the user. The avatar acts as the narrator asking the user to perform a series of actions. In Figure 10, the user is being asked to look at a red dot on the screen and press a key to take a picture. The avatar’s eye gaze is hand coded to help direct attention and guide the user through the system. Since the



t

Figure 10: Avatar based user interface used in the avatar maker application.

avatar in this case is pre-scripted, we are able to feed the speech recognition system a text version of the dialog thus enabling enhanced lip-synch. The same system was used to generate the narrator in the accompanying video.

Conclusions

Of all the forms of non-verbal communication, eye gaze is perhaps the most powerful, yet it is one we are rarely consciously aware of. As we begin to build realistic virtual human characters, it will be increasingly important to have them behave in natural ways. This paper has focused on one aspect of that behavior, that of eye gaze.

A simple computational model for eye gaze has been shown to be effective at simulating what people do in ordinary conversations. Our experiments have indicated that having a natural eye gaze model on an avatar elicits changes in the viewers' eye gaze patterns. In a virtual teleconferencing setting, natural eye gaze models help direct attention from avatar to avatar and can indicate the levels of participation.

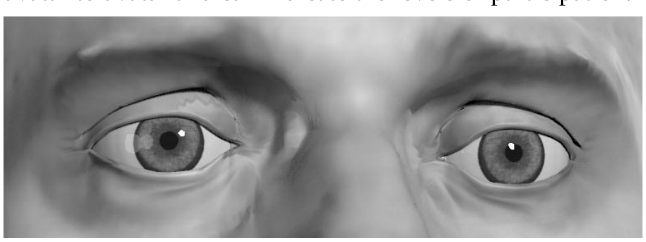


Figure 11: More realistic avatar model.

Clearly, there is still a long way to go to make real-time avatars look and behave like real people. Our avatar model is continuing to increase in visual realism. Figure 11 shows an example of a portion of a current avatar model we will be using in future experiments. We are also building personality models for avatars which will effect a variety of non-verbal communication ranging from body posture to eye gaze. We would hypothesize that the increased realism will add new demands to the eye gaze model itself.

We also intend to include expected variations in the eye gaze model due to perceived geometric distance between conversants, and due to gender and culture differences as reported in the literature. When using eye gaze models in avatar mediated interaction, cross-cultural interaction difficulties could potentially be minimized by having each user see a different gaze model conforming to local social expectations.

Gazing into another's eyes tells us a lot about that person. We hope that avatars will be able to communicate is ways just as rich in the near future.

REFERENCES

1. Argyle, M., Cook, M. Gaze and mutual gaze. Cambridge University Press, New York, NY, 1976.
2. Beattie, G.W. Contextual constraints on the floor-apportionment function of speaker-gaze in dyadic conversation. British Journal of Social & Clinical Psychology. Nov,1979, 391-392.
3. Beattie, G. W. A further investigation of the cognitive interference hypothesis of gaze-patterns during conversation. British Journal of Social Psychology. Nov 1981, 243-248.
4. Bull, M. The Timing and Coordination of Turn-Taking. Doctoral Thesis, University of Edinburgh 1998
5. Becker, B. Marks, G. Social Conventions in Collaborative Virtual Environments, Proceedings of Collaborative Virtual Environments 1998 (CVE'98), Manchester, UK, 17-19th June 1998.
6. Cassell, J., Pelachaud, C., Badler, N. Steednam, M., Achorn, B. Becket, T., Deouville, B., Prevost, S. and Stone, M. Animated conversation: Rule-based generation of facial expression, gesture, & spoken intonations for multiple conversation agents. Proceedings of SIGGRAPH 1994.
7. Cassell, J., and Vilhjalmsson, H. Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous, in Autonomous Agents and Multi-Agent Systems, 2: 45-64. 1999.
8. Cassell, J. Thorisson,J. The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents, Applied Artificial Intelligence, 1999.
9. Cassell, J., Torres, O. and Prevost, S. (1999). "Turn Taking vs. Discourse Structure: How Best to Model Multimodal Conversation." In Y. Wilks (ed.), *Machine Conversations.* The Hague: Kluwer.
10. Chopra-Khullar, S. and Badler, N. "Where to look? Automating visual attending behaviors of virtual human characters," Autonomous Agents Conf., Seattle, WA, May 1999.

11. Cohen, K. Speaker interaction: video teleconferences versus face-to-face meetings. Proceedings of teleconferencing and electronic communications. 1982, 189-199.
12. Dourish,, Adler, Bellotti and Henderson Your Place or Mine? Learning from Long-Term Use of Video Communication EPC-1994-105
13. He, L. Cohen, M.F., Salesin, D,. The Virtual Cinematographer: A Paradigm for Automatic Real-Time Camera Control and Directing SIGGRAPH 96
14. Elzinga, R.H. Temporal Organization of Conversation, Sociolinguistics Newsletter, 1978, 9, 2 (Summer), 29-31.
15. EverQuest. http://www.everquest.com
16. EMED: Eye Movement Equipment Database. http://ibs.derby.ac.uk/emed/
17. Grayson, D. and Coventry, L. The effects of visual proxemic information in video mediated communication. SIGCHI, Vol.30 No.3, July 1998
18. Ishii, H., Kobayashi, M. & Grudin, J. Integration of inter-personal space and shared workspace: ClearBoard design and experiments. Proc. CSCW'92
19. Kendon, A. Some functions of gaze direction in social interaction. ACTA PSYCHOLOGICA, 26 (1967), 22-63.
20. Morikawa, O. & Maesako, T.HyperMirror: Toward pleasant-to-use video mediated communication system. CSCW'98.
21. Muhlbach,L., Bocker, M. and Prussog, A. Telepresence in videocommunications: A study on stereoscopy and individual eye contact. Human Factors. 1995, 37. 290-305.
22. Novick, D,G., Hansen, B. and Ward,K. Coordinating turn-taking with gaze. Proceedings of the International Conference on Spoken Language Processing (ICSLP'96), Philadelphia, PA, October, 1996, 3, 1888-91
23. Okada, K., Maeda, F., Ichikawaa, Y. & Matsushita, Y. Multiparty videoconferencing in virtual social distance: MAJIC design. Proc. CSCW'94
24. The Palace. http://www.thepalace.com
25. C. Pelachaud, N. Badler, and M. Steedman. ``Generating facial expressions for speech,'' Cognitive Science 20(1), pp. 1-46, 1996
26. Sellen, A., Buxton, W. & Arnott, J. Using spatial cues to improve videoconferencing. Proceedings of CHI '92, 651-652. 1992.
27. Smith, M. A., Farnham, S., and Drucker, S. M. The Social Life of Small Graphical Chat Spaces. To appear in SIGCHI 2000.
28. Stephenson, N. Snowcrash. Bantam Doubleday Dell Publishing Group, 1992.
29. Torres, 0. Cassell, J and Prevost S. (1997). Modeling gaze behavior as a function of discourse structure. First International Workshop on Human-Computer Conversations. Bellagio, Italy. (To appear as a chapter in Machine Conversation. Yorick Wilks, editor. Kluwer.)
30. Ultima Online http://www.ultimaonline.com
31. Velichkovsky, B.M. & Hansen, J.P. New Technological Windows into Mind: There is More in Eyes and Brains for Human-Computer Interaction. In CHI-96: Human factors in computing systems. NY: ACM Press.
32. Velichkovsky, B.M. Communicating attention: Gaze position transfer in cooperative problem solving. Pragmatics and Cognition. (1995) 3, 199-223.
33. Velichkovsky, B.M., Pomplun, M., Rieser. H. Attention and communication: Eye-movement-based research paradigms. In W.H.Zangemeister, S.Stiel & C.Freksa (Eds.), Visual attention and cognition. Amsterdam/New York: Elsevier Science Publisher. 1996.
34. Vertegaal, R., Velichkovsky, B. and van der Veer, G. Catching the eye. SIGCHI Bulletin, 1997, 29, 4, 87-92.
35. Whittaker, S. Rethinking video as a technology for interpersonal communications: theory and design implications. International Journal of Human-computer Studies. 1995 (42) 501-529
36. WorldsChat. http://www.worlds.net