The Impact of Eye Gaze
on Communication using Humanoid Avatars

Maia Garau +, **Mel Slater** +, **Simon Bee** * 1, **Martina Angela Sasse** +

+ Department of Computer Science	* MLB3/7
University College London (UCL)	BT Labs.
Gower St., London, WC1E 6BT, UK	Martlesham Heath
{m.garau | m.slater | a.sasse}@cs.ucl.ac.uk	Ipswich, IP5 3RE, UK

ABSTRACT

In this paper we describe an experiment designed to investigate the importance of eye gaze in humanoid avatars representing people engaged in conversation. We compare responses to dyadic conversations in four mediated conditions: video, audio-only, and two avatar conditions. The avatar conditions differed only in their treatment of eye gaze. In the random-gaze condition the avatar's head and eye animations were unrelated to conversational flow. In the informed-gaze condition, they were related to turn-taking during the conversation. The head animations were tracked and the eye animations were inferred from the audio stream. Our comparative analysis of 100 post-experiment questionnaires showed that the random-gaze avatar did not improve on audio-only communication. The informed-gaze avatar significantly outperformed the random-gaze model and also outperformed audio-only on several response measures. We conclude that an avatar whose gaze behaviour is related to the conversation provides a marked improvement on an avatar that merely exhibits liveliness.

Keywords

Avatars, gaze, collaborative virtual environments (CVEs), computer-mediated communication (CMC), mediated communication, nonverbal behaviours.

INTRODUCTION

This paper describes an experiment to assess the extent to which avatars are important in enhancing the quality of communication in remote meetings. In particular we consider the importance of gaze behaviour in increasing the expressive potential of avatars.

Research in videoconferencing has been driven by the premise that adding visual information should increase the communicative bandwidth of mediated interaction [23].

Though media spaces in the form of videoconferencing and groupware systems enable us to share visual information from our physical environment [10], the disadvantage is that the 3D context of each user's physical environment is lost [22]. While there are many advantages to using video for remote communication, there are certain collaborative situations - such as remote acting rehearsals [19] - in which it is essential to preserve spatial relationships. Collaborative Virtual Environments (CVEs) can begin to address this need by placing remote users within a shared computer-generated space where they can interact with the environment and with other users represented by avatars. However, the significant reduction involved in replacing video-mediated conferencing with current CVEs is in the paucity of avatar expression compared with the expression possible with a live human face on video.

One of the challenges in developing CVEs is therefore the creation of expressive avatars. Avatars in existing graphical chat environments have been criticized for acting merely as placeholders and not contributing meaningfully to the communication process [24]. While it is more straightforward to simply program avatars to exhibit random "liveliness" [24], our results show that the inclusion of even a single expressive behaviour that reflects the conversation can make a significant impact on the perceived quality of communication.

Here we isolate one expressive behaviour, eye gaze. Gaze is a richly informative behaviour in face-to-face interaction. It serves at least five distinct communicative functions [3, 12]: regulating conversation flow, providing feedback, communicating emotional information, communicating the nature of interpersonal relationships and avoiding distraction by restricting visual input. The perception of eye gaze depends on a combination of head and eye orientation [2, 3, 9]. In this study we compare an avatar with random head and eye movements to a visually

---

1 Author's present address: Parametric Technology Corp.
19 Apex Court, Woodlands, Almondsbury Park
Bristol, BS32 4JT, UK
sbee@ptc.com

identical avatar that combines simple head tracking with "while speaking" and "while listening" eye animations inferred from the audio stream. The design of these eye animations was informed by social psychology research on the differences in gaze patterns while speaking and while listening in face-to-face interaction [3, 4, 12]. Both avatar conditions are then compared to video (with audio) and audio-only baseline conditions.

We assess the impact of each condition on the quality of communication by comparing participants' subjective responses along four dimensions: how natural the conversation felt, degree of involvement in the conversation, sense of co-presence and positive or negative evaluation of the conversation partner.

In the next section we discuss related research on eye gaze in media spaces and CVEs. We then discuss the design and our prior expectations for the experiment. We conclude with a discussion of our findings and suggestions for future work.

RELATED WORK ON GAZE IN MEDIATED COMMUNICATION

Research on gaze in mediated communication has been concerned mainly with issues of conversation management in multiparty interaction. One of the perceived limitations of telephony-based videoconferencing systems is that they do not support selective gaze [7, 16, 23]. Various media space systems have attempted to address this limitation by distributing individual audiovisual units in physical space to represent each user [see 23 for a review].

Studies in CVEs have attempted to address the problem of how to support selective gaze in multiparty interaction within a shared 3D space. The GAZE groupware system [23] is designed to ease turn taking by conveying gaze direction in a shared virtual space using VRML2. This system uses an advanced desk-mounted eyetracking system to measure where each person is looking. The gaze information is then represented metaphorically in the form of a 2D texture-mapped "persona" which moves about its own x- and y-axis in the 3D environment.

Taylor and Rowe [22] argue that the GAZE groupware system is problematic for two reasons. First, using a snapshot instead of video precludes any possibility of expressing other nonverbal cues through the persona. Second, the use of a plane makes it difficult to generate the kinds of profile views useful in multiparty communication. They address these limitations by rendering video of the facial region on a generic 3D model of a face. Their system animates the head movement by tracking the two earphones and microphone to obtain head position information for each user. The eye movement is contained in the video image. Their system renders avatars from an asymmetric viewpoint that corresponds to the position of the real participant, who typically sits 20" away from a 14"

desktop screen. They conclude that this system improves group interaction by preserving the semantic significance of gaze. However, integrating video as a part of gaze animation fails to address the needs of users who prefer to remain visually anonymous behind a synthetic avatar.

Both of the above studies are concerned with supporting selective gaze in groups of three or more. In their SIGGRAPH sketch, Colburn et al. [8] present findings from an experiment comparing visual attention to the screen during 20 dyadic conversations using an avatar. Participants were presented with three 3-minute visual stimuli in random order: a blank screen, a fixed-gaze avatar and an avatar with a functioning eye gaze model, based on who was speaking and whether or not the participant was looking at the screen. Participants looked at the screen more when the avatar was present and most of all when the gaze model was active. These results would suggest that an eye gaze model informed by social psychology research might motivate participants to pay more attention to the avatar during conversation.

In our experiment we wished to make an in-depth analysis of the impact of inferred virtual gaze on the perceived quality of communication.

EXPERIMENT GOALS AND EXPECTATIONS

The purpose of the experiment was two-fold: first to test whether having an avatar improved the quality of the communication compared to not having a visual image at all (i.e. compared to audio-only). The second, more specific goal was to examine the role of gaze: when the avatar's gaze was directly related to the conversation, did this improve the quality of communication compared to the visually identical random-gaze avatar. Our expectation was that the inferred-gaze condition would lead to an improvement in communication quality; however, we were not sure whether having an avatar with random gaze would be better or worse than not having one at all. The video condition was expected to always outperform the other conditions in terms of quality of communication.

EXPERIMENTAL DESIGN

100 participants were randomly assigned to one of four conditions representing different methods of mediated communication. The conversations took place over a video tunnel link, which offered a face-on, head-and-shoulders view of their conversation partner. We deliberately did not make use of the 3D potential of the avatar, as we wished to abstract away from everything but the presence of the avatar and the effect of the head and eye animations. Therefore, only a head-and-shoulders view of each avatar was displayed. Each group was randomly divided into pairs of participants. The two people in each pair did not know each other. They were of the same gender, and were matched approximately for age (20s, 30s, 40s or 50s). They performed a ten-minute role-playing task in which

they were randomly assigned to play one of two roles - a mayor or a baker.

Independent Variables

The four conditions were as follows.

1. Audio-only: there was an audio connection and only a flat gray image on the video tunnel monitor.

2: Random-gaze avatar: the avatar's head and eye animations were random. We wished to avoid using a fixed-gaze avatar as there is evidence [6] that continuous gaze can result in negative evaluation of a conversation partner. It was therefore necessary to use an avatar whose head and eyes moved, but in a way that was not related to the conversation except by accident.

3. Inferred-gaze avatar: the avatar's head movement was determined by tracking and the eye movement was determined by inference from conversational turn taking, as described below.

4. Video: the video tunnel monitor showed a head-and-shoulders view of the remote participant with no perceptible audio or video delay. We chose high-quality video as a baseline condition because in same-room interaction, factors such as interpersonal distance, posture and gesture might have confounded results [2].

Population

All participants were employees of a Telecommunications company and the majority therefore had a technical background. Our original goal was to have 12 pairs per condition. We ran a total of 15 per condition and discarded data from sessions where we experienced technical difficulties. All usable data was kept in the sample. Hence the inferred-gaze and audio conditions have 13 pairs and the other conditions have 12. All pairs were matched for gender and age, with the exception of a pair in the audio-only condition, where a man in his 30s was paired with one in his 40s. There were 37 male and 13 female pairs.

Apparatus

The lab consisted of a reception room, a central control area from which the experiment could be monitored, and two separate soundproofed rooms in which the participants sat.

Audiovisual Link between the Soundproofed Rooms

The two rooms were audiovisually linked using a video tunnel [7]. This symmetrical setup is designed to enable mutual gaze. In each room a camera is placed behind a half-silvered mirror that reflects the image from an upturned monitor. Thus, while the participant looks at the reflected image, the hidden camera captures a face-on view that is sent to the remote participant's monitor. Participants sat 2m from a 21-inch Sony PVM-2130QM video monitor.

In the video condition each monitor was directly linked to the camera output from the other room. In the avatar conditions an S-video link with Vinegen Pro scan rate converters was used to connect each computer with the other person's monitor, enabling participants to see each other's avatar. Participant 1's avatar was run on a Dell Dimension XPST 550 (Pentium III), Windows 98, with a GeForce 256 chipset, Gulillemot 3D Prophet video card and Creative AWE32 sound card. Participant 2's avatar was run on a Compaq AP400 PIII 500, Windows 98, with a GeForce 256 chipset, Elsa Gloria2 video card and integrated sound.

Figure 1: Video tunnel setup showing the video condition

The sound was recorded using an AKG C747 microphone placed on the desk. As the audio stream drove the avatar's lip and eye movements, we needed to isolate each microphone from incoming sound from the other room. We therefore equipped participants with Senneheiser HD265 headphones. In the inferred-gaze condition a copy of the audio stream was sent to the computer, along with head position information from the Polhemus Isotrak II tracking system.

Avatars

We used a female avatar for female participants, and a male avatar for male participants. Both participants in each pair were represented by a visually identical avatar as differences in facial geometry and texture mapping could potentially impact on the visual effect of the animations. The participants never saw their own avatar, so the fact that both were identical was unknown to them. Each avatar was independently controlled for each user.

Eye animations

In the random-gaze condition the timings and directions for both head and eye movement were determined using the computer's pseudo random number generator function and therefore had no relationship with the ongoing conversation. In the inferred-gaze condition the eye animations were driven by the audio stream based on

"while speaking" and "while listening" modes whose timings were taken from research on face-to-face dyadic conversations [3, 4, 12]. For the "while speaking" mode, mean duration of gaze was 1.8 seconds for "at partner" gaze, and 2.1 seconds for "away" gaze, with an average frequency of 14 "at partner" glances per minute. For the "while listening" mode, mean duration of gaze was 2.5 seconds for "at partner" gaze and 1.6 seconds for "away" gaze, with an average frequency of 17 "at partner" glances per minute. For "at partner" gaze, the avatar's eyes focused directly ahead. The values for vertical and horizontal angles of "away" gaze were chosen randomly from a uniformly distributed range of 0 to 15 degrees. The sign of the angle was random. In order to avoid repeating identical animation loops the duration of "at" and "away" gaze was randomised using the waiting time exponential distribution.

Figure 2: Face-on view of male and female avatars looking "at" and "away" from partner

Task
In order to assess whether the presence of the avatar improved the quality of communication, we needed a task in which participants would benefit from having visual feedback. It has been suggested [16, 21] that users benefit most from having visual feedback when performing equivocal tasks that have no "correct" outcome and require negotiation. We developed a negotiation task specifically for the study. Participant 1 was asked to represent a mayor and participant 2 a baker. For female pairs both the mayor and baker were described as female, and for male pairs both characters were described as male. In the fictional scenario the baker's daughter was pregnant, allegedly by the mayor's son. The son had refused to take responsibility, prompting the baker's irate spouse to draft a letter to the press. It was in the interests of both to avoid a scandal. Each participant was given slightly different goals and the task was to reach a mutually acceptable decision within ten minutes in order to prevent the letter from being mailed. It was stressed that this was not an acting task and that the participants should be themselves and should feel free to improvise.

Procedure
Participants were met in a reception area by two researchers. Conversation between participants was avoided or kept to a minimum as they were led through to the lab and into separate soundproofed rooms. Here they were given ten minutes to read through the scenario. The video tunnel monitors were already switched on when participants arrived. When both participants were ready the piece of paper describing the scenario was taken away to avoid visual distraction during the conversation. Participants were told that the doors would be shut simultaneously and that the conversation would be timed from that point.

Conversations were stopped at the end of the 10 minutes whether or not a conclusion was reached. Participants were then asked to fill out a questionnaire. We concluded with a semi-structured interview, conducted individually with each participant. All of the conversations and interviews were videotaped with participants' written consent.

A slight change of procedure was necessary in the inferred-gaze condition, as we needed to calibrate the head tracking for each participant. In order to avoid participants seeing their partner's avatar being adjusted, we only switched on the video tunnel monitors after calibration was complete in both rooms. We then found a pretext to leave the room for a few minutes to let participants become accustomed to their partner's avatar.

Response Variables
The fundamental variable of interest was quality of communication. This was subdivided into four broad indicators.

1. *Face-to-face:* The extent to which the conversation was experienced as being like a real face-to-face conversation.

2. *Involvement*: The extent to which the participants experienced involvement in the conversation.

3. *Co-presence*: The extent of co-presence between the participants - that is, the sense of being with and interacting with another person rather than with a computer interface.

4. *Partner Evaluation*: The extent to which the conversational subjects positively evaluated their partner, and the extent to which the conversation was enjoyed.

The responses to these variables were elicited by means of the post-experiment questionnaire, each response being on a 9-point Likert scale, where 1 was anchored to 'Strongly disagree' and 9 to 'Strongly agree'. The questions were partly based on previous questionnaires designed to elicit subjective responses to mediated communication [16, 21] and were grouped as follows:

Face-to-face:
1. I could readily tell when my partner was listening to me.

2. I was able to take control of the conversation when I wanted to.

3. It was easy for me to contribute to the conversation.

4. The conversation seemed highly interactive.

5. There were frequent and inappropriate interruptions.

6. This felt like a natural conversation.

Involvement:
7. I found it easy to keep track of the conversation.

8. I felt completely absorbed in the conversation.

Co-presence:
9. I had a real sense of personal contact with my conversation partner.

10. I was very aware of my conversation partner.

Partner-evaluation:
11. My partner was friendly.

12. My partner did NOT take a personal interest in me.

13. I trusted my partner.

14. I enjoyed talking to my partner.

15. I would be interested in meeting my partner face-to-face.

For the purposes of analysis the questionnaire anchors were swapped when coding responses to negative questions such as 12.

Explanatory Variables
Data was collected on three auxiliary variables that could be used as explanatory variables in the analysis: gender, age, and whether or not the subject was a native English speaker.

Method of Analysis
The Likert scales are of course ordinal and it is not appropriate to use these as interval responses. The method was used as in [18] which provides a conservative analysis of the responses based only on count data. We select a cut-off value of 7, and classify as a positive response one which is 7, 8 or 9. Then for each set of questions we count the number of positive responses (r) out of the n possible positive responses. For example, n=6 for the face-to-face variable, and n=5 for the partner-evaluation variable. Under the null hypothesis of randomly and independently assigned responses, r has a binomial distribution and therefore logistic regression can be used for the analysis as to how r varies across the main condition and with respect to other variables [15]. In the case where the right-hand-side of the regression consists of only one factor (for example, the condition) this is equivalent to a one-way ANOVA but using the more appropriate binomial distribution rather than the Normal. The *deviance* is the appropriate goodness of fit measure for this regression model, and has an approximate chi-squared distribution with degrees of freedom depending on the number of fitted parameters. A rule-of-thumb is that if the deviance is less than twice the degrees of freedom then the model is good fit (at the 5% significance level). The change in deviance as variables are deleted from or added to the current model is especially useful, since this indicates the significance of that variable in the fitted model. Here a large change of deviance indicates the degree of significance, i.e., the contribution of the variable to the overall fit.

RESULTS
Figure 3 shows the means of the raw questionnaire responses. It is noteworthy that in every case but for co-presence, the video results in the highest score, followed by the inferred gaze avatar condition.

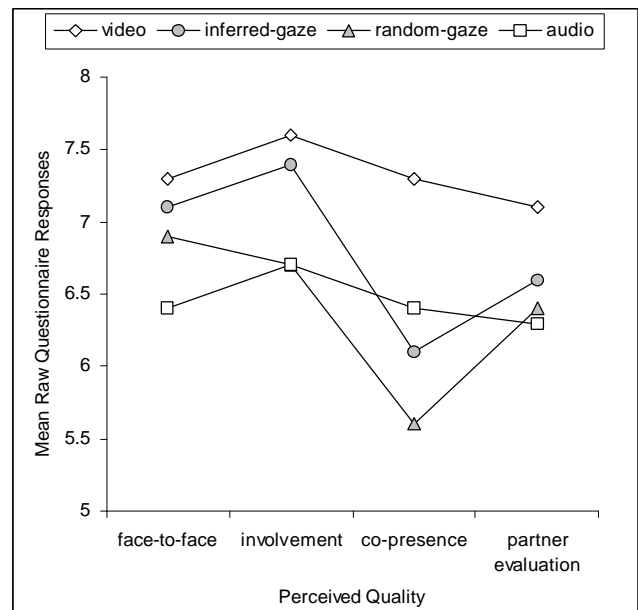**Figure 3: Means of Raw Questionnaire Responses**

Table 1 below shows the mean and standard deviations of the counts of 'high' responses (7, 8 or 9) across the n questions in each condition.

Table 1: Mean ± standard deviations of Count Response Variables.

N = number of subjects in each condition.

n = number of questions on which the count is based.

	Video (N=24)	Inferred gaze (N=26)	Random gaze (N=24)	Audio (N=26)
Face-to-face n=6	4.8±1.4	4.5±1.3	4.0±1.6	3.4±1.8
Involvement n=2	1.7±0.4	1.6±0.6	1.4±0.7	1.3±0.6
Co-presence n=2	1.6±0.6	1.1±0.8	0.7±0.7	1.2±0.7
Partner-evaluation n=5	3.7±1.4	3.2±1.4	2.7±1.5	2.8±1.4

The results in Table 1 are consistent with the expectation that video will tend to produce the highest responses, and audio-only the worst (except for the co-presence variable). In each case the avatar with the inferred gaze direction results in a higher count than the avatar with random gaze. We use the logistic regression analysis to check for significance, and also to test whether other exogenous variables should be included in the model.

Table 2: Logistic Regression with Face-to-face as the Response Variable

Overall deviance = 214, d.f. = 9

The deviance column shows the increase in deviance that would result if the corresponding variable were deleted from the model.

Variable	Deviance c^2	d.f.	5% c^2
Condition	23.8	3	7.815
Role	8.4	1	3.841

Table 2 shows the results of the regression analysis for face-to-face as the response variable. Condition and role were significant. Role refers to whether the subject is playing the mayor or the baker. There is no significant difference between the video and inferred-gaze avatar conditions. The random-gaze avatar and audio methods are significantly different from the video and inferred-gaze methods. The person who played the role of the baker tended to have a lower response count than the person who played the mayor.

A similar analysis was carried out for each of involvement and co-presence as the response variables. For both the main-effect for condition was the only significant factor. For involvement the inferred-gaze avatar is not significantly different from the video condition. The random-gaze avatar is significantly less than the video condition, but not significantly different from the audio. For co-presence video is significantly higher than all the others, and the worst case is the random-gaze-avatar.

Table 3: Logistic Regression with Partner-evaluation as the Response

Overall deviance = 182, d.f. = 94

Variable	Deviance c^2	d.f.	5% c^2
Condition	13.77	3	7.815
Role	12.22	1	3.841
Native Speaker	5.9	1	3.841

Table 3 shows the results of the regression for partner-evaluation. In this case both role and whether or not the person is a native English speaker are significant. Video again produces the highest count, and each of the other three conditions are significantly lower than video. The audio and random-gaze-avatar are not significantly different from each other, and the inferred-gaze avatar count is significantly higher than the audio and random-gaze. The baker role again led to a lower count than the mayor role. A person who was not a native English speaker tended to have a higher count than a native English speaker.

If all of these response variables are combined, then we get the result shown in Table 4.

Table 4: Logistic Regression for the Sum of all Response Variables

Overall deviance = 329, d.f. = 95

Variable	Deviance c^2	d.f.	5% c^2
Condition	53.4	3	7.815
Role	28.5	1	3.841

Once again condition is highly significant with video resulting in the highest count, the inferred-gaze avatar is lower (the difference from video is significant) and the audio and random-gaze avatar are significantly lower than the others, but not different from each other. The baker role has a lower overall count than the mayor role.

The analysis above has relied on the more appropriate count variables for face-to-face, involvement, co-presence, partner-evaluation and the overall response. If, however, a standard normal regression is used then all the main effects for condition remain significant, and the conclusions do not change. The overall response remains significant at the 5% level. The fit for the overall response is significant but not high (r = 0.35, d.f. = 95), which indicates that there are certainly other explanatory factors responsible for variations in quality of communication, which were not included in this study.

DISCUSSION

Our expectation was that in the context of dyadic interaction using this particular setup, the video condition would outperform the other conditions. As we used an equivocal task in which participants would benefit from having visual feedback, we predicted that video would result in better perceived communication quality as it presented participants with complete, real-time nonverbal feedback from the head and shoulders. It was therefore more visually informative than both the audio-only condition, which had no visual feedback, and the avatar conditions, which only used head, eye and lip movement. The reason for comparing the avatar conditions to the audio and video baseline conditions was to understand whether an avatar can improve the perceived quality of mediated communication. More specifically, where does it fall on a continuum from no visual information (audio-only) to complete visual information from the head and shoulders (video)?

It should be stressed that this experiment was not designed to compare the relative merits of video and avatar-mediated communication. In this particular context it would undoubtedly be preferable to use video, as results attest. The purpose of this research is to understand how to create expressive behaviours for avatars so that we can enrich mediated communication in CVEs for those contexts, such as virtual acting rehearsals, where remote participants benefit from interacting in a shared 3D space. The study was deliberately not conducted in the context of a CVE because we wished to isolate gaze behaviour from any other factors, such as spatial, gestural or postural cues that might have confounded results.

We did not know whether having an avatar with random gaze would be better or worse than not having one at all. The overall analysis suggests that simply having an avatar whose head and eye movements are not related to the conversational flow does not improve communication when compared to audio-only. Indeed there is some evidence that in the case of co-presence, the random-gaze avatar is worse than the pure audio stream. Comments in the interviews supported this. One participant in this condition explained that *"I didn't feel it represented anything. It just kind of sat there doing something rather than helping. It didn't make me feel there was a presence of the other person there."*

We predicted that having an avatar whose gaze behaviour was directly related to the conversation would improve quality of communication compared to one whose gaze behaviour was random. This prediction is confirmed by the results, which show that the inferred-gaze avatar consistently and significantly outperforms the random-gaze avatar. This builds on Colburn et al.'s findings [8] that participants appeared to pay more attention to the avatar when the gaze model was active than when it was

not. This finding also supports the notion [24] that, for avatars to meaningfully contribute to communication, it is not sufficient for them to appear lively. Rather, their animation needs to reflect some aspect of the conversation that is taking place.

In the overall analysis the inferred-gaze avatar also significantly outperforms audio-only. This suggests that in contrast with a randomly animated avatar, one with meaningful animations can indeed contribute to perceived communication quality. However in the case of the co-presence variable it performs worse than pure audio. This might partly be explained by the familiarity of telephone conversations versus the novelty effect of interacting with a person represented by an avatar.

What is perhaps most surprising is that the inferred-gaze avatar is not significantly different from the video in terms of sense of involvement and the extent to which the conversation is likened to face-to-face. The avatar only used an approximation of a single nonverbal behaviour, gaze, whereas the video presented participants with full and accurate nonverbal feedback from the face. This is encouraging as it suggests that an avatar can begin to make a significant contribution to the positive perception of communication even without detailed facial expression. Future research will need to address the additive impact of combining gaze with other nonverbal behaviours in avatars.

CONCLUSIONS AND FUTURE WORK

This study deliberately isolated gaze animations by providing only a face-on, head-and-shoulders view of the avatar. In future we aim to expand the research to investigate the impact of gaze in combination with other nonverbal behaviours such as facial expression, gesture and posture in the context of immersive multiparty interaction in CVEs. We would also aim to use more precise instrumentation to investigate the impact of using an avatar whose eye movements are accurately tracked in addition to the head movements.

Through this study we sought to answer two questions. First, whether an avatar can improve the quality of communication between remote users. Second, whether an avatar whose gaze behaviour is directly related to the conversation can offer a significant improvement over a visually identical avatar with random gaze.

The random-gaze avatar does not provide a significant improvement over pure audio, suggesting that the simple introduction of an avatar does not automatically improve participants' perception of communication. Rather, the avatar must have certain behaviour characteristics in order to be useful. The inferred-gaze avatar outperforms the pure audio stream on several measures. This suggests that an avatar whose behaviours reflect an aspect of conversational flow can indeed make a contribution to

improving remote communication. Finally, the inferred-gaze avatar significantly outperforms the random-gaze avatar on all measures, indicating that an avatar whose behaviours are related to the conversation can present a marked improvement over an avatar that merely exhibits liveliness.

ACKNOWLEDGMENTS

REFERENCES

1. Anstis, S.M., Mayhew, J.W., Morley, T. The Perception of Where a Face or Television 'Portrait' is Looking. *American Journal of Psychology 82* (1969), 474-489.

2. Argyle, M., *Bodily Communication*. Methuen & Co., London, 1975.

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

4. Argyle, M. and Ingham, R. Mutual Gaze and Proximity. *Semiotica 6* (1972), 32-49.

5. Argyle, M., Ingham, R., Alkema, F. and McCallin, M. The Different Functions of Gaze. *Semiotica 7* (1973), 10-32.

6. Argyle, M., Lefebvre, L. and Cook, M. The Meaning of Five Patterns of Gaze. *Eur. J. Soc. Psychol. 4, 2* (1974), 125-136.

7. Buxton, W. Telepresence: integrating shared task and person spaces, in *Proceedings of Graphics Interface '92*, 123-129. Earlier version appears in *Proceedings of Groupware '91* (Amsterdam, Oct. 29, 1991), 27-36.

8. Colburn, A., Drucker, S., Cohen, M. The Role of Eye-Gaze in Avatar-Mediated Conversational Interfaces, in *Sketches and Applications, Siggraph '00* (New Orleans LA, July '00), ACM Press, 253.

9. Gibson, J.J. and Pick, A. Perception of Another Person's Looking Behaviour. *American Journal of Psychology 46* (1963), 386-394.

10. Hindmarsh, J., Fraser, M., Heath, C., Benford, S. and Greenhalgh, C. Fragmented Interaction: Establishing Mutual Orientation in Virtual Environments, in *Proceedings of CSCW '98* (Seattle WA, November 1998), 48-53.

11. Isaacs, E. and Tang, J. What Video Can and Can't Do for Collaboration: A Case Study. *Multimedia Systems, 2* (1993), 63-73.

12. Kendon, A. Some Functions of Gaze-Direction in Social Interaction. *Acta Psychologica, 26* (1967), 22-63.

13. Kendon, M. and Cook, A. The Consistency of Gaze Patterns in Social Interaction. *Br. J. Psychol, 60, 4* (1969), 481-494.

14. Lord, C. and Haith, M.M. The Perception of Eye Contact. *Perception & Psychophysics 16, 3* (1974), 413-416.

15. McCullagh, P. and Nelder, J.A. *Generalised Linear Models*. Chapman and Hall, 1983.

16. Sellen, A. Remote Conversations: The Effects of Mediating Talk with Technology. *Human-Computer Interaction, 10(4)* (1995), 401-444.

17. Short, J., Williams, E. and Christie, B. *The Social Psychology of Telecommunications*. John Wiley & Sons, London, 1976.

18. Slater, M., Steed, A., McCarthy, J., Maringelli, F. (1998) The Influence of Body Movement on Subjective Presence in Virtual Environments. *Human Factors, 40(3)*, 469-477.

19. Slater, M., Howell, J., Steed, A., Pertaub, D-P., Garau, M. and Springel, S. Acting in Virtual Reality, in *ACM Collaborative Virtual Environments*, (San Francisco CA, Sept 2000), 103-110.

20. Steed, A. Slater, M., Sadagic, A. Tromp, J. and Bullock, A. Leadership and collaboration in virtual environments. In *IEEE Virtual Reality*, (Houston TX, March 1999), 112-115.

21. Straus, S. and McGrath, J.E. Does the Medium Matter: The Interaction of Task and Technology on Group Performance and Member Reactions. *Journal of Applied Psychology 79* (1994), 87-97.

22. Taylor, M.J. and Rowe, S.R. Gaze Communication Using Semantically Consistent Spaces, in *Proceedings of CHI 2000* (The Hague, Amsterdam, April 2000), ACM Press, 400-407.

23. Vertegaal, R. The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration, in *Proceedings of CHI '99* (Pittsburgh PA, May 1999), ACM Press, 294-301.

24. Vilhjálmsson, H. and Cassell, J. BodyChat: Autonomous Communicative Behaviours in Avatars, in *Autonomous Agents '98* (Minneapolis MN, May 1998), 269-276.