The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment

Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, M. Angela Sasse

Department of Computer Science, University College London (UCL), Gower St., London WC1E 6BT
{m.garau, m.slater, v.vinayagamoorthy, a.brogni, a.steed, a.sasse}@cs.ucl.ac.uk

ABSTRACT

This paper presents an experiment designed to investigate the impact of visual and behavioral realism in avatars on perceived quality of communication in an immersive virtual environment.

Participants were paired by gender and were randomly assigned to a CAVE™-like system or a head-mounted display. Both were represented by a humanoid avatar in the shared 3D environment. The visual appearance of the avatars was either basic and genderless (like a "match-stick" figure), or more photorealistic and gender-specific. Similarly, eye gaze behavior was either random or inferred from voice, to reflect different levels of behavioral realism.

Our comparative analysis of 48 post-experiment questionnaires confirms earlier findings from non-immersive studies using semi-photorealistic avatars, where inferred gaze significantly outperformed random gaze. However responses to the lower-realism avatar are adversely affected by inferred gaze, revealing a significant interaction effect between appearance and behavior. We discuss the importance of aligning visual and behavioral realism for increased avatar effectiveness.

Keywords

Virtual Reality, immersive virtual environments, avatars, mediated communication, photo-realism, behavioral realism, social presence, copresence, eye gaze.

INTRODUCTION

This paper presents an experiment that investigates participants' subjective responses to dyadic social interaction in a shared, immersive virtual environment (IVE). It focuses on the impact of avatar realism on perceived quality of communication. Specifically, it explores the relative impact of two logically distinct aspects of avatar realism: appearance and behavior.

One of the chief appeals of IVEs as a medium of communication is that they enable remotely located people to meet and interact in a shared 3D space. This is of particular benefit for tasks such as remote acting rehearsal [19], where preserving spatial relationships among participants is paramount. However, one significant limitation is low avatar expressiveness as compared with the rich feedback available through live human faces on video.

Improving avatar expressiveness poses complex challenges. There are technical limitations as well as theoretical goals to consider. Technically, one of the central constraints is the tension between "realism and real time" [20]. In terms of an avatar's appearance, increased photo-realism comes at the expense of computational complexity, introducing significant and unwanted delays to real-time communication. In terms of behavior, if the goal is to replicate each person's real movement, tracking can seem an attractive solution. Systems such as Eyematic [10] have shown compellingly that it is possible to track eye movement and drive an avatar in real time using a simple desktop camera. However, in immersive CAVE™-like systems[1] where users wear stereoscopic goggles and move freely about the space, it can be difficult to provide a robust solution. At the same time, tracking other body and facial behaviors can be invasive, as well as expensive in terms of rendering.

Research on nonverbal behavior in face-to-face communication [1] can offer valuable leads on how to improve avatar expressiveness without resorting to full tracking. In the study presented in this paper, we focus on a single behavior, eye gaze. We investigate whether it is possible to make an improvement to people's communication experience by inferring their avatar's eye movements from information readily available from the audio stream. We build on previous research conducted in a non-immersive setting [14] [17], where random eye gaze was compared with gaze that was inferred based on speaking and listening turns in the conversation.

[1] CAVE™ is a trademark of the University of Illinois at Chicago. In this paper we use the term 'Cave' to describe the generic technology as described in [9] rather than to the specific commercial product.

We further extend this previous research by varying the appearance to investigate the impact of behavioral realism with different levels of visual realism.

Our goal is to understand how these varying levels of realism impact on people's responses to their communication experience in the IVE. For each pair of participants taking part in the experiment, one experienced the IVE though a Cave and the other through a head-mounted display (HMD). We assess the impact of avatar realism by comparing participants' subjective responses along the four dimensions considered in previous research [14]: how natural the conversation seemed (in terms of how similar it was to a real face-to-face conversation), degree of involvement in the conversation, sense of copresence, and evaluation of the conversation partner.

In the following section we discuss related work on social responses to avatars that have varying degrees of realism. We then describe the design and running of our experiment and discuss our findings. We conclude with suggestions for continuing work needed to optimize users' experience in avatar-mediated communication.

RELATED WORK ON SOCIAL RESPONSES TO AVATARS AND AGENTS WITH DIFFERING LEVELS OF REALISM

Social responses to virtual humans have been studied in contexts ranging from small-group interactions in shared VEs [7] [21], to interactions with interface agents capable of gaze behavior [26], to fully embodied conversational agents [8]. Both objective and subjective methods have been employed. Bailenson et al. [6] studied the impact of gaze realism on an objective social response, proxemic behavior. They report results consistent with expectations from Argyle's intimacy equilibrium theory [2] that participants would maintain greater interpersonal distance when an agent engaged in mutual gaze. The remainder of this section will focus specifically on a selection of studies centering on subjective responses to visual realism and eye gaze in agents and avatars.

Tromp et al. [25] describe an experiment where groups of three human participants met in a shared VE. Two users were represented by simple "blocky" avatars with little visual detail, while the third was represented by a more realistic one. Analysis showed that even though all three avatars had the same limited functionality, the person represented by the more realistic avatar was seen as "standoffish" and "cold" because of a lack of expression. Slater et al. [21] argue that higher realism in an avatar's appearance may lead to heightened expectations for behavioral realism. This crystallizes the need to further explore the relationship between the appearance of an avatar and its behavior.

Fukayama et al. [13] describe a study on the impact of eye animations on the impressions participants formed of an interface agent. Their gaze model consists of three parameters: amount of gaze, mean duration of gaze and gaze points while averted. Their comparative analysis of responses to nine different gaze patterns suggests that agent gaze can reliably influence impression formation. For this particular study they isolated the agent's eyes from any other facial geometry. Elsewhere, they investigate whether the impact of the gaze patterns is affected by the facial realism of the agent [12]. They conclude that varying the appearance from visually simplistic to more realistic has no effect on the impressions produced.

In terms of behavioral realism, and specifically eye gaze, two additional studies are directly relevant to the experiment discussed in this paper. Garau et al. [14] investigated the impact of avatar gaze on participants' perception of communication quality by comparing a random-gaze and inferred-gaze avatar. In the inferred-gaze condition, the avatar's head movement was tracked and its eye movement was driven by the audio stream based on "while speaking" and "while listening" animations whose timings were taken from research on face-to-face dyadic interaction [3] [4] [16]. In the random-gaze condition, the participant's head was not tracked, and both the avatar's head and eye movement were random. The results showed the inferred-gaze avatar significantly outperformed the random-gaze one on several response measures.

Lee et al. [17] present a similar experiment comparing random, static and inferred eye animations. Their inferred animations were based on the same theoretical principles as in [14], but were further refined using a statistical model developed from their own gaze tracking analysis of real people. Their results were consistent with Garau et al.'s findings that inferred gaze significantly outperforms random gaze. However, they do not report specifically on two-way verbal communication with the agent.

One aspect of studies to date is that participants were shown a limited, head-and-shoulders view of the virtual human, and that the spatial relationship was fixed by the 2D nature of the interaction. They leave open the question of how these gaze models might hold up in an immersive situation where participants are able to wander freely around a shared space, and where the avatar is seen as an entire body.

EXPERIMENT GOALS AND HYPOTHESES

Our goal for this experiment was threefold. Firstly, to disambiguate between the effect of inferred eye movements and head-tracking, both of which may have contributed to the results reported in [14]. Secondly, to test how the inferred-gaze model performs in a less forgiving immersive setting where it is not desirable to attempt to control the participant's gaze direction. Finally, to explore the combined impact on quality of communication of eye gaze model and visual appearance.

Our initial hypothesis was that behavioral realism would be independent in its effects on quality of communication from the impact of visual realism, and that the behavioral realism would be of greater importance. We expected the inferred-gaze model to outperform the random-gaze one for both the higher-realism and lower-realism avatar. We were not sure the extent to which the gaze animations would impact on the lower-realism avatars, or how the two avatars would perform in comparison with each other.

EXPERIMENTAL DESIGN

Independent Variables

A between-groups, two-by-two factor design was employed with the two factors being the degree of avatar photo-realism and behavioral realism, specifically in terms of eye gaze behavior.

Population

48 participants were paired with someone of their own gender and assigned randomly to one of the four conditions. They did not know their conversation partner prior to the experiment, and were not allowed to meet beforehand. A gender balance was maintained across the four conditions, as illustrated in Table 1. The reason for this is that there is evidence [3] that males and females can respond differently to nonverbal behaviors, particularly in the case of eye gaze cues.

Table 1: Factorial Design

	Random gaze	Inferred gaze
Lower-realism avatar	3 male pairs 3 female pairs	3 male pairs 3 female pairs
Higher-realism avatar	3 male pairs 3 female pairs	3 male pairs 3 female pairs

Participants were recruited from the university campus using an advertising poster campaign. They were paid $8 for the one-hour study.

Apparatus

ReaCTor: The Cave used was a ReaCTor made by Trimension, consisting of three 3m x 2.2m walls and a 3m x 3m floor. It is powered by a Silicon Graphics Onyx2 with 8 300MHz R12000 MIPS processors, 8GB RAM and 4 Infinite Reality2 graphics pipes. The participants wore CrystalEyes stereo glasses which are tracked by an Intersense IS900 system. They held a navigation device with 4 buttons and an analogue joystick that is similarly tracked; all buttons except for the joystick were disabled to stop participants from manipulating objects in the virtual room. The joystick was used to move around the VE, with pointing direction determining the direction of movement enabled for the horizontal plane only.

Head-mounted Display (HMD): The scenarios were implemented on a Silicon Graphics Onyx with twin 196 MHz R10000, Infinite Reality Graphics and 192M main memory. The tracking system has two Polhemus Fastraks, one for the HMD and another for a 5 button 3D mouse. The helmet was a Virtual Research V8 which has true VGA resolution with 640x480x3 color elements for each eye. The V8 has a field of view of 60 degrees diagonal at 100% overlap. The frame rate was kept constant for both the Cave and the HMD.

Both participants had wireless microphones attached to their clothing. These were activated only for the duration of the conversation.

Figure 1: Participants in the Cave could see their own bodies

Figure 2: Participants in the HMD could not see their own bodies or physical surroundings while in the IVE. The image of the IVE visible on the screen was for the benefit of the researchers.

Software

The software used was implemented on a derivative of DIVE 3.3x [11]. This was recently ported to support spatially immersive systems [23]. DIVE (Distributed Interactive Virtual Environment) is an internet-based multi-user virtual reality system in which participants can navigate in a shared 3D space and interact with each other.

Plugins make DIVE a modular product. A plugin was developed in C to animate the avatar body parts as discussed below. Since DIVE also supports the import and export of VRML and several other 3D file formats, it was possible to import ready-made avatars from other projects [19]. DIVE reads the user's input devices and maps physical actions to logical actions in the DIVE system. In this case the head and the right hand were tracked.

At the start of each session, the avatars were moved to their correct starting positions with the aid of Tcl script. A separate Tcl script was used to open the doors separating the virtual rooms at the end of the training period.

Virtual Environment

The shared IVE in which the participants met consisted of two spacious "training" rooms connected to a smaller "meeting" room in the center. The doors separating the virtual rooms were kept closed during the training session to avoid participants seeing each other's avatar before the conversation task. All rooms were kept purposefully bare so as to minimize visual distraction.

Avatars

Each participant was represented by a visually identical avatar as we wished to avoid differences in facial geometry affecting the impact of the animations. Each avatar was independently driven for each user. The participants in the HMD, who were visually isolated from the physical surroundings of the lab, could see the hands and feet of their avatar when looking down; the participants in the Cave could only see their own physical body. This means that participants never saw their own avatar in full, so they were unaware that both were visually identical. In the lower-realism condition a single, genderless avatar was used to represent both males and females (Figure 3). For the higher-realism avatar, a separate male and female avatar were used, as shown in Figure 3.

Figure 3: Lower-realism avatar, higher-realism male avatar, higher-realism female avatar

All avatars used in the experiment were made H-Anim compliant [15] and had identical functionality. A plugin was used to animate the avatar's body in order to maintain a visually consistent humanoid. This included inferring the position of the right elbow using inverse kinematics when the user's tracked hand moved, and deducing the position of the avatar's knees when the user bent down. There were also some deductions involved in the rotation of the head and body. The body was not rotated to the same direction as the head unless there was some translation associated with the user. This was to enable the user to nod, tilt and shake their head in the VE whilst in conversation.

Eye animations

One of the fundamental rules of gaze behavior in face-to-face communication is that in dyadic interaction, people gaze at their communication partner more while listening than while speaking [3] [4] [5].

Garau et al. [14] drew on this principle, implementing a "while speaking" and "while listening" eye animation model based on timing and frequency information taken from face-to-face studies [3] [4] [5]. More recently, Lee et al. [17] refined the animations based on their own empirical gaze tracking research. Their model is consistent with timing expectations from the literature, but adds valuable new probabilities for gaze direction during "away" fixations that were absent in [14]. In a pre-experiment, we implemented and compared the models used by [14] and [17]. The more detailed model by Lee was selected for this study as it yielded more satisfying results in the immersive setting. Full details of this model can be found in [17].

Both previous models assumed a non-immersive setting where the participant was seated in front of a screen. The avatar's "at partner" gaze was therefore always straight ahead. In this new study, a decision was made not to automatically target "at partner" eye direction at the other avatar. Rather, "at partner" gaze was kept consistent with the position and orientation of the head. In this way, the avatar could only seem as if it was looking "at partner" if the participant was in fact looking directly at the other avatar's face (based on head-tracking information).

Task

The same role-playing negotiation task as described in [14] was used. Each participant was randomly assigned to play either a mayor or a baker, whose families were involved in a potentially volatile situation. It was within both their interests to avoid a scandal breaking out in their small town. The task was to come to a mutually acceptable conclusion within ten minutes. It has been argued that it is when performing equivocal tasks with no single "correct" outcome that people stand to profit from having visual feedback [18] [24]. We wanted to test the impact of the different avatars in a context where high demands would be placed on their contributing role in the communication process.

Procedure

Participants did not meet prior to the experiment, to avoid the possibility of any first impressions influencing the role of the avatar in the conversation. The first person to arrive was assigned to the Cave, the second to the HMD in an adjacent room. Since there were two different roles in the scenario, the role played by the participant in each interface was randomized to avoid introducing constant error.

After filling out a background questionnaire, participants read the scenario. They then each performed a navigation training task in the Cave or HMD. When they felt comfortable, the doors separating the virtual training rooms from the central meeting room were opened simultaneously. At the same time, the microphones were activated and they were given a maximum of 10 minutes for the conversation. The session concluded with a post-questionnaire and a semi-structured interview conducted individually with each participant.

Response Variables

The primary variable of interest was *perceived quality of communication*, divided into four broad indicators. n is the number of questions on with the construct is based.

1. *Face-to-face:* The extent to which the conversation was experienced as being like a real face-to-face conversation. (n=6)

2. *Involvement*: The extent to which the participants experienced involvement in the conversation. (n=2)
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. (n=2)
4. *Partner Evaluation*: The extent to which the conversational subjects positively evaluated their partner, and the extent to which the conversation was enjoyed. (n=5)

Whilst [14] used a 9-point Likert scale, each questionnaire response in this study was on a 7-point Likert-type scale, where 1 was anchored to strong disagreement and 7 to strong agreement. For the purposes of analysis some questionnaire anchors needed to be swapped so that all "high" scores would reflect a high score of the response variable being studied.

Explanatory Variables

As well as the independent variables (two visual and two behavioral conditions) there were a number of explanatory variables in the analysis. These included gender, age, and status. In addition, data was collected on their technical expertise in terms of computer use and programming, as well as experience with interactive virtual reality systems and computer games. Another important explanatory variable was the degree of participants' social anxiety in everyday life, as measured by the standardized SAD questionnaire [27] where a higher score reflects greater social anxiety. This final variable was employed in order to take account of different types of subject responses to the interaction, for example the tendency to approach or avoid the avatar during the conversation.

Method of Analysis

The same logistic regression method was used as in [14] and other previous analyses [22]. This is a conservative method of analysis, and has the advantage of never using the dependent variable ordinal questionnaire responses as if they were on an interval scale. Each response variable is constructed from a set of n questions. For each question we count the number of "high responses" (that is, responses of 6 or 7 on the Likert Scale). Therefore each response variable is a count out of n possible high scores. For example, for the face-to-face variable, n = 5, so the response is the number of "high scores" out of these 5 questions.

The response variables may be thought of as counts of "successes" out of n trials, and therefore naturally have a binomial distribution, as required in logistic regression. In the case where the right-hand-side of the regression consists of only one two factors (in the case the type of avatar and the type of gaze animation) this is equivalent to a two-way ANOVA but using the more appropriate binomial distribution rather than the Normal. Of course other co-variates may be added into the model, thus being equivalent to two-way ANOCOVAR.

In this regression model the *deviance* is the appropriate goodness of fit measure, 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 overall is a good fit to the data (at the 5% significance level). More important, *the change in deviance* as variables are deleted from or added to the current fitted model is especially useful, since this indicates the significance of that variable in the model. Here a large change of deviance indicates the degree of significance, i.e. the contribution of the variable to the overall fit.

RESULTS

In this section we report the results of a logistic regression analysis on the independent variables for perceived quality of communication.

Table 2: Mean ± Standard Errors of Count Responses

Response	Type of avatar	Random Gaze	Inferred Gaze
Face-to-face	Lower-realism	4.2±0.5	2.9±0.5
	Higher-realism	2.2±0.4	3.9±0.6
Involvement	Lower-realism	1.3±2.9	1.3±0.2
	Higher-realism	0.9±0.2	1.2±0.2
Copresence	Lower-realism	1.2±0.2	0.7±0.2
	Higher-realism	0.3±0.1	1.1±0.3
Partner Evaluation	Lower-realism	2.6±0.5	2.2±0.4
	Higher-realism	1.8±0.5	2.8±0.5

Table 2 shows the raw means of the count response variables. An inspection of the *face-to-face* response suggests that there is a strong interaction effect - that within each row and column there is a significant difference between the means, but that there is no significant difference between the top left and bottom right cells.

Table 3: Fitted Logistic Regression for the Count Response Variables

Fitted Variable	Face-to-face Deviance χ^2	Involvement	Co-presence	Partner evaluation
Type avatar • type gaze	22.03 (+)	-	9.7 (+)	5.0 (+)
Age	7.8 (+)	16.9 (+)	14.1 (+)	-
Role (baker)	10.0 (-)	-	-	6.2 (-)
SAD	15.7 (-)	-	-	-
Overall deviance	79.9	67.7	60.5	125.0
Overall d.f.	40	46	43	44

Again, we consider the results for *face-to-face* as the response variable to illustrate the analysis. In Table 3 above, the deviance column shows the increase in deviance that would result if the corresponding variable were deleted from the model. The tabulated χ^2 5% value is 3.841 on 1 d.f. and all d.f.'s below are 1. The sign in brackets after the χ^2 value is the direction of association of the response with the corresponding variable (i.e., positively or negatively correlated).

Each of these terms is significant at the 5% level of significance (i.e., none can be deleted without significantly reducing the overall fit of the model). *Type of avatar* and *type of gaze* were significant for 3 of these 4 response variables. The participant *age*, *role* and *SAD* score were significant for some of them (*role* refers whether they played the mayor or baker in the negotiation task). Just as in [14], for this response variable, the person who played the role of the baker tended to have a lower *face-to-face* response count than the person who played the mayor.

The type of *interface* (Cave or HMD) did not have a significant effect on responses. However, *age* was found to be significant, and positively associated with the response: older people are more likely to have rated their experience as being like a face-to-face interaction.

The formal analysis demonstrates the very strong interaction effect between the *type of avatar* and the *type of gaze* (denoted by the • symbol in Table 3). In other words the impact of the gaze model is different depending on which type of avatar is used. For the lower-realism avatar, the (more realistic) inferred-gaze behavior reduces *face-to-face* effectiveness. For the higher-realism avatar the (more realistic) inferred-gaze behavior increases effectiveness. This is illustrated by Figure 4 and Figure 5 below, showing the means of raw questionnaire responses for each avatar.

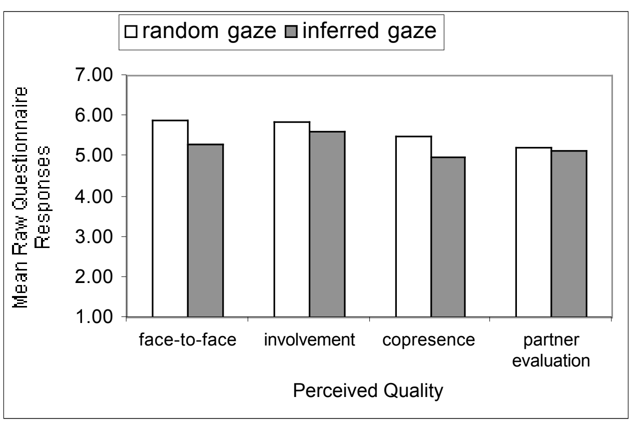

Figure 4: Means of Raw Questionnaire Responses for Lower-Realism Avatar

For the lower-realism avatar, the inferred-gaze model has a consistently negative effect on each response variable (Figure 4). The opposite is true of the higher-realism avatar (Figure 5). Consistency between the visual appearance of the avatar and the type of behavior that it exhibits seems to be necessary; low fidelity appearance demands low fidelity behavior, and correspondingly higher fidelity appearance demands a more realistic behavior model (with respect to eye gaze).

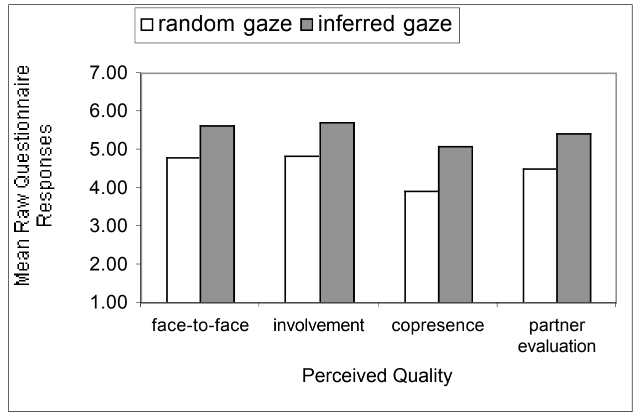

Figure 5: Means of Raw Questionnaire Responses for Higher-Realism Avatar

The logistic regression analysis suggests that for 3 out of the 4 response variables, there is a significant interaction effect between *type of avatar* and *type of gaze*. The exception is *involvement*, for which there is no significant effect of either avatar or gaze type; this is consistent with the findings of [14]. However, the *copresence* and *partner evaluation* variables illustrate the same strong interaction effect as *face-to-face*. In each of the three cases, the higher-realism avatar has a higher response when used with the inferred-gaze model. The implications of these findings are discussed in the following section.

In addition to perceived quality of communication, other social responses were captured by the questionnaire. These included the extent to which participants had a sense of being in a shared space (spatial copresence), the extent to which the avatar was perceived as real and like a human, and the degree to which the avatar helped participants to understand aspects of their partner's behavior and attitude. Our analysis indicates that there is an overwhelmingly cohesive model, where the same interaction effect between the *type of avatar* and the *type of gaze* holds for all of these other measures. The findings related to these additional measures will be reported in detail elsewhere.

DISCUSSION

The findings in [14] were that the inferred-gaze avatar consistently outperformed the random-gaze avatar, and that for several of the response measures this difference was significant. However, the results confounded head tracking with the inference about the avatar's eye movement based on face-to-face dyadic research [3] [4] [16]. The present result resolves the ambiguity, since head-tracking was kept identical in all conditions. Independently of head tracking, the inferred-gaze model has a significant positive impact on perceptions of communication in the case of the higher-realism avatar.

Our second aim was to compare gaze models within an immersive setting. Recall that previous studies [12] [13] [14] [17] were carried out in a non-immersive setting where the participants' point of view was controlled by the experimental setup. How would the eye gaze models perform in a communication context where participants were able to control their point of view within a shared 3D space? The results presented here suggest that in the case of the higher-realism avatar, the pattern of results reported in [14] holds for 3 of the 4 response variables: namely, that in the case of *face-to-face*, *copresence* and *partner evaluation*, the inferred-gaze model significantly outperforms the random-gaze model. This is consistent with our initial hypothesis that the inferred-gaze model should have a significant and positive impact on participants' responses to the communication experience in the IVE. The fact that this was not the case for the lower-realism avatar is very interesting and is addressed below.

One response variable, *involvement*, was not affected by either *type of avatar* or *type of gaze*. This variable referred to sense of absorption and the ability to keep track of the conversation. The overwhelming majority of participants stated that the focus of their attention was on their partner's voice, as the avatar did not give them the rich visual feedback they required in the conversation. The deliberate reduction of the avatar's expressive repertoire to minimal behaviors (eye, head and hand movement) may partly explain why *involvement* was not affected.

Despite the limited feedback offered, other aspects of the communication experience were significantly affected, as illustrated by the comments of one participant in the lower-realism, random-gaze condition: "Even if it is not a very realistic avatar, it helps a little. It gives you something to focus on. Although you do not think of it as a person, strangely it does stop you turning away or doing anything inappropriate. Also your mind does not wander as much as it might on the telephone. You are immersed in the environment." Many participants mentioned that the avatar helped to give them a strong sense of being in a shared space with their partner. Without exception, all participants stood facing their partner's avatar throughout the entire conversation. They took care to maintain a suitable interpersonal distance and felt compelled to display polite attention.

Our third and final question concerned the appearance of the avatars. In [14], both eye gaze conditions were implemented with the same relatively photorealistic avatar. In the present research we wanted to investigate whether higher-quality avatar behavior could compensate for a lower-realism appearance. It is clear that there is a highly consistent pattern of responses amongst many of the response variables that make up our notion of quality of communication. The overall conclusion must be that for the lower-realism avatar, the inferred-gaze model may not improve quality of communication, and may in some instances make things worse. However, for the higher-realism avatar, the inferred-gaze model improves perceived quality of communication. The evidence suggests that there should be some consistency between the type of avatar and the type of gaze model that is used: the more realistic the avatar appearance, the better the gaze model that should be used.

Contrary to Fukayama et al. [12], we found a significant difference in the way our lower-realism and higher-realism avatars were affected by the different gaze models. The divergence in our findings may be at least partially explained by two factors. Firstly, their gaze model was based on different parameters to ours. Secondly, their communication context was fundamentally different to ours: where theirs concerned one-way interaction from an agent to a human, ours concerned two-way communication between immersed human participants who were engaged in a delicate negotiation task. For this reason, it is likely that the demands placed on the virtual human were fundamentally different.

One other interesting finding is that in absolute terms, the higher-realism avatar did not outperform the lower-realism avatar. This lends weight to the hypothesis in [21] that the higher the photo-realism of the avatar, the higher the demands for realistic behavior. It would be interesting to further explore this notion in future work.

CONCLUSIONS AND FUTURE WORK

This study sought to investigate the impact of visual and behavioral realism in avatars on perceived quality of communication between participants meeting in a shared IVE. In terms of appearance, the avatar was either visually simplistic or more realistic; in terms of behavior, we singled out eye gaze, comparing inferred-gaze and random-gaze models previously tested in a non-immersive setting. Our results clear up an ambiguity from previous research regarding whether the significant differences in performance between the gaze models were due to head-tracking or avatar eye animations inferred from the audio stream. We conclude that independent of head-tracking, inferred eye animations can have a significant positive effect on participants' responses to an immersive interaction. The caveat is that they must have a certain degree of visual realism, since the lower-realism avatar did not appear to benefit from the inferred gaze model. This finding has implications for inexpensive ways of improving avatar expressiveness using information readily available in the audio stream. It suggests avenues for interim solutions for the difficult problem of providing robust eyetracking in a Cave.

In this study we have taken eye gaze animation as a specific (though important) instance of avatar behavior. We cannot claim, of course, that results will generalize to other aspects of avatar behavior, but findings for eye-gaze will generate hypotheses for studies of further aspects of avatar animation. In future work we aim to investigate the impact of other behaviors such as facial expression, gesture and posture, and to expand the context to include multi-party groups of 3 or more. We also aim to further explore the complex interaction effect between an avatar's appearance and behavior by investigating additional social responses such as spatial copresence, with a view to understanding how to make avatars more expressive for communication in shared IVEs.

ACKNOWLEDGMENTS

This research was possible thanks to a BT/EPSRC Industrial CASE award. It was funded by the EQUATOR Interdisciplinary Research Collaboration. We thank David

Swapp for his generous help with the audio, and Pip Bull for his help in adapting the avatars originally created by David-Paul Pertaub. Finally, we would like to thank the participants for their time and for sharing their thoughts.

REFERENCES

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

2. Argyle, M. *Bodily Communication.* 2nd ed., Methuen & Co., London, 1988.

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. Bailenson, J. N., Blascovich, J., Beall, A. C., and Loomis, J. M. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators and Virutal Environments 10*, 6 (2001), 583-598.

7. Benford, S., Bowers, J., Fahlen, L. E., Greenhalgh, C., and Snowdon, D. User Embodiment in Collaborative Virtual Environments, in *Proceedings of CHI'95: ACM Conference on Human Factors in Computing Systems* (Denver, CO, 1995), 242-249.

8. Cassell, J., Sullivan, J., Prevost, S., and Churchill, E., Eds., *Embodied Conversational Agents,* The MIT Press, Cambridge, MA, 2000.

9. Cruz-Neira, C., Sandin, D. J., and DeFanti, T. A. Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE, in *Proceedings of Computer Graphics (SIGGRAPH) Proceedings, Annual Conference Series* 1993), 135-142.

10. Eyematic. Available at http://www.eyematic.com.

11. Frecon, E., Smith, G., Steed, A., Stenius, M., and Stahl, O. An Overview of the COVEN Platform. *Presence: Teleoperators & Virtual Environments 10* (2001), 109 - 127.

12. Fukayama, A., Sawaki, M., Ohno, T., Murase, H., Hagita, N., and Mukawa, N. Expressing Personality of Interface Agents by Gaze, in *Proceedings of INTERACT* (Tokyo, Japan, 2001), 793-794.

13. Fukayama, A., Takehiko, O., Mukawa, N., Sawaki, M., and Hagita, N. Messages Embedded in gaze of Interface Agents - Impression management with agent's gaze -, in *Proceedings of SGICHI - Conference in Human factor in Computing Systems* (Minneapolis, USA, 2002), ACM Press, 41-48.

14. Garau, M., Slater, M., Bee, S., and Sasse, M.-A. The Impact of Eye Gaze on Communication using Humanoid Avatars, in *Proceedings of CHI'01: ACM Conference on Human Factors in Computing Systems* (Seattle, WA, 2001), 309-316.

15. H-Anim. *Humanoid Animation Working Group,* Available at http://www.hanim.org.

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

17. Lee, S. H., Badler, J. B., and Badler, N. I. Eyes Alive, in *Proceedings of 29th Annual Conference on Computer Graphics and Interactive Techniques* (San Antonio, TX, 2002), ACM Press, 637-644.

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

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

20. Slater, M., Steed, A., and Chrysanthou, Y. *Computer Graphics and Virtual Environments: From Realism to Real-Time.* Addison Wesley Publishers, Harlow, England, 2001.

21. Slater, M. and Steed, A. *Meeting People Virtually: Experiments in Virtual Environments.* In R. Schroeder, Ed., *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments,* Springer Verlag, Berlin, 2001.

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

23. Steed, A., Mortensen, J., and Frecon, E. *Spelunking: Experiences using the DIVE System on CAVE-like Platforms.* In B. Frohlicj, J. Deisinger, and H.-J. Bullinger, Eds., *Immersive Projection Technologies and Virtual Environments,* Springer-Verlag, 2001, 153-164.

24. 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.

25. Tromp, J., Bullock, A., Steed, A., Sadagic, A., Slater, M., and Frecon, E. Small Group Behaviour Experiments in the COVEN Project. *IEEE Computer Graphics and Applications 18*, 6 (1998), 53-63.

26. Vertegaal, R., Slagter, R., Van der Veer, G., and Nijholt, A. Eye Gaze patterns in Conversations: There is More to Conversational Agents than Meets the Eyes, in *Proceedings of CHI'01: ACM Conference on Human Factors in Computing Systems* (Seattle, WA, 2001), 301-307.

27. Watson, D. and Friend, R. Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology 33* (1969), 448-457.