High-Fidelity Avatar Eye-Representation

William Steptoe*  Anthony Steed†

Department of Computer Science
University College London

Abstract

In collaborative virtual environments, the visual representation of avatars has been shown to be an important determinant of participant behaviour and response. We explored the influence of varying conditions of eye-representation in our high-fidelity avatar by measuring how accurately people can identify the avatar’s point-of-regard (direction of gaze), together with subjective authenticity assessments of the avatar’s behaviour and visual representation. The first of two variables investigated was socket-deformation, which is to say that our avatar’s eyelids, eyebrows and surrounding areas morphed realistically depending on eye-rotation. The second was convergence of our avatar’s eyes to the exact point-of-regard.

Our results suggest that the two variables significantly influence the accuracy of point-of-regard identification. This accuracy is highly dependent on the combination of viewing-angle and the point-of-regard itself. We found that socket-deformation in particular has a highly positive impact on the perceived authenticity of our avatar’s overall appearance, and when judging just the eyes. However, despite favourable subjective ratings, overall performance during the point-of-regard identification task was actually worse with the highest quality avatar. This provides more evidence that as we move forward to using higher fidelity avatars, there will be a trade-off between supporting realism of representation and supporting the actual communicative task.

Keywords: Virtual reality, collaborative virtual environments, avatars, social presence, behavioural realism, representation, eye-gaze.


1 INTRODUCTION

1.1 Background

Collaborative Virtual Environments (CVEs) connect co-located or remote users of immersive display systems within a spatial, social and information context [13], with the aim of supporting high-quality interaction. Co-presence is the extent to which the computer becomes transparent and there is a sense of being present with other people in the CVE, and there is a direct working with the other people [15]. CVEs often combine high degrees of presence and co-presence, because the sense of being in another place and of being there with other people reinforce each other [14]. Consequently, user embodiment is a fundamental issue when designing CVEs [4], and this is typically maintained using an avatar - a graphical representation of a human.

Virtual humans are capable of eliciting appropriate responses from people, and it has been observed that unwritten social norms such as proxemics and unease from close-range mutual eye-contact with unknown partners occur in CVEs as they do in real-life. Bailenson and colleagues tested Argyle and Dean’s (1965) equilibrium theory’s specification of an inverse relationship between mutual gaze and interpersonal distance [3]. They found that all participants maintained more space around virtual humans than they did around similarly sized and shaped but nonhuman-like objects. Correspondingly, people’s perception of their own virtual representation has been found to have a significant and instantaneous impact on user behaviour, with more attractive avatars prompting an increase in self-disclosure, and relatively taller avatars raising confidence during negotiation tasks [19].

Murray and Roberts [12] presented a study assessing people’s ability to judge an avatar’s direction of gaze (point-of-regard) between head-tracked movement (static centred eyes) and combined head and eye-movement. Results indicated that the addition of eye-movement significantly improved accuracy of point-of-regard (POR) identification. Svanfield and colleagues [17] measured sensitivity to avatar-gaze by asking people to judge when the avatar was looking directly into their eyes. Results showed that it was possible to achieve a state where participants perceive this to be the case. It was also found that people were less sensitive to changes in POR depth than in the horizontal and vertical dimensions, and sensitivity diminished when the avatar head was turned to the side.

Avatars exhibiting higher levels of visual and behavioural fidelity can potentially communicate the subtleties of human nonverbal communication more successfully, thereby enhancing the perceived authenticity of the interaction [18, 7]. A consistent interaction effect between visual and behavioural realism has been found [6], indicating that the effect of identical behavioural traits change in relation to the avatar’s appearance: higher visual fidelity benefits from consistently realistic behaviour, while the converse is true for lower fidelities. Therefore, it is the combination of how realistic an avatar is perceived to be (termed social presence) and the degree to which an avatar behaves as would be expected in real life (termed behavioural realism) which determines the level of social influence it can have on people [5].

1.2 Motivation

As we start to use highly realistic avatars in CVEs, it is important to investigate representational and behavioural facets which are likely to be influential in shaping people’s perception, and consequently the overall effectiveness of an avatar. In this paper, we limit our investigation to the eyes, as they are of central importance to social behaviour and nonverbal communication, and are considered to be the most intense social signals on the human face [1]. Uses of eye-communication which are particularly relevant to CVEs include indicating and determining focus of attention [2], indicating and determining actions [8], addressing and prompting [11], and when describing and handling objects [16].

We selected two variables which, based on our preliminary avatar observation tests, we expected to have a significant impact on the accuracy of POR identification and the perceived authenticity of representation:
• **Socket deformation** denotes that we deform our avatar’s eyelids, eyebrows and surrounding areas as the eyes rotate, and includes blinking. This deformation is based upon human physiology that the eyelids, during vision must act bilaterally, and in following the elevation and depression of the eyeballs must not cover the pupils [9].

• **Vergence** indicates that rather than looking ‘through’ the focal point, our avatar’s eyes converge to the precise POR; a behaviour predicted (but not investigated) as being an influential factor benefitting the accuracy of POR identification [12].

Our main hypothesis was that, by more closely mimicking natural human behaviour, the inclusion of socket-deformation and vergence would enhance the behavioural and representational properties of our avatar. Therefore, we expected superior levels of POR identification accuracy and (particularly) perception of authenticity under these conditions.

2 Experimental Design

A within-subjects, repeated measures design was conducted over four conditions to determine the impact of the two variables on POR identification accuracy and perception of authenticity. The conditions were: 1) No Socket-deformation, no Vergence, 2) No Socket-deformation, Vergence, 3) Socket-deformation, no Vergence, and 4) Socket-deformation and Vergence. 48 (24 male), normal or corrected-to-normal vision participants were recruited and paid to perform the study. Conditions were resequenced to negate learning bias, with two participants completing each order. Our avatar was developed using Poser (e frontier) and 3DS Max (Autodesk), while our experiment system was implemented in Flash (Adobe).

2.1 Stimuli

The scene presented to participants consisted of a clock-face positioned directly in front of our avatar and viewed from three camera angles: centre and (symmetrical) left and right. The virtual cameras were positioned to be consistent with common two and three-way conversational situations as defined by proxemics grammar, and illustrated in Figure 1. Our avatar gazed at the 12 numbers around the clock-face, thereby varying POR laterally, vertically and to a lesser extent in depth. To satisfy the objectives of the study, it was essential to isolate results to the eyes and immediate surrounding areas, so these were the sole animated parts of our avatar. Hence, we did not consider head-movement, which despite (and due) to being closely linked to eye-movement [1], is highly likely to impact data and thus a topic to be addressed in future work.

Figure 1: Experiment setup from above. A 1.0m separation between bodies is in the far-phase of personal distance [10], promoting affinity between participant and avatar and maintaining visual clarity for the task. Side-cameras were offset to +/-40° from the 0° central camera, remaining consistent with common conversational situations.

2.2 Procedure

The experiment was performed on a 42” LCD screen, thereby displaying the avatar’s head approximately life-size. Participants were seated 1.0m from the screen at avatar eye-level. Our experiment was divided into two sections, performed under each condition:

- **POR identification**: From the looking straight-ahead rest-position, our avatar’s eyes would rotate to look at one of the numbers around the clock-face for 2 seconds, return to the central POR for one second, and then return to the previous POR for another 2 seconds before returning to the rest-position. Each POR shift was completed in a single saccade (a small rapid jerky movement of the eye as it jumps from fixation on one point to another). Participants would then respond with their perceived POR using the input panel to the left of the avatar. This process would continue, with the viewing camera alternating in a circular sequence (centre, left, right) and the avatar looking at another number. For each condition, our avatar gazed in a random order at each POR three times (once per camera). Thus, there were 36 data points per condition and a total of 144 per participant over the four conditions.

- **Questionnaire**: After each condition’s POR identification task, participants were presented with a questionnaire eliciting responses on a 1.7 Likert scale. Questions related to their perception of the naturalness and authenticity of our avatar’s behaviour and appearance, together with self-performance ratings of the POR identification task.

Therefore, for each condition, the data we collected was accuracy of POR identification (objective measure of perceived POR compared to actual POR), together with the subjective questionnaire responses.

3 Analysis

3.1 Point-of-Regard Identification

A repeated measures two-way analysis of variance (ANOVA) was performed on our entire data-set for the POR identification task, with the four eye-representation conditions and the twelve POR angles (numbers) as factors. This overall measure combined data from all three cameras, and indicated significant differences between the four conditions ($F(3,11)=4.19; \ P<0.01$), and significant differences between POR angles ($F(3,11)=11.26; \ P<.001$). Post-hoc tests determined that the significant differences lay between the
Socket-deformation and Vergence condition and two other classes: Socket-deformation, no Vergence \( (F(1,11)=11.73; P<.001) \) and No Socket-deformation, no Vergence \( (F(1,11)=4.03; P<.005) \), with an interaction effect found between condition and POR angle during the latter comparison \( (F(1,11)=2.02; P<.05) \). Table 1 shows varying decreases in overall POR identification accuracy levels as we introduce conditions featuring socket-deformation and vergence. This corresponds to the above ANOVA calculations, suggesting that when analysed on this macro level, the inclusion (and in particular the combination) of the two methods negatively impact POR identification accuracy, thereby opposing part of our original hypothesis.

We then analysed the impact of camera-angle on the accuracy of POR identification. Our three experimental camera-angles can be evaluated as two \( (1\% \text{ centre-camera and 2}. \% \text{ symmetrical left and right side-cameras}) \) granted identical states for opposite POR angles. Starting with the centre camera, repeated measures two-way ANOVA was performed, again with the four conditions and the twelve POR as factors. A significant difference was found between conditions \( (F(3,11)=3.26; P<.05) \) with an interaction effect between treatments \( (F(3,11)=1.73; P<.01) \). Post-hoc tests then exposed that the differences lay between the Socket-deformation, no Vergence condition and two other classes: Socket-deformation and Vergence \( (F(1,11)=7.64; P<.01) \) and No Socket-deformation, no Vergence \( (F(1,11)=6.99; P<.01) \). We then performed identical tests on the side-camera data and again found a significant difference between conditions \( (F(3,11)=4.02; P<.01) \). Similar to overall camera analysis, differences were found between the Socket-deformation and Vergence condition and the Socket-deformation, no Vergence \( (F(1,11)=6.92; P<.01) \) and No Socket-deformation, no Vergence \( (F(1,11)=11.83; P<.001) \) classes, with an interaction effect found during the latter comparison \( (F(1,11)=1.98; P<.05) \). Table 1 details accuracy of POR identification, highlighting consistently high accuracy from the centre-camera, and lower, more variable levels from the side-cameras.

### Table 1: POR identification mean accuracy and (standard deviation) for conditions and camera-angles.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Overall</th>
<th>Centre-camera</th>
<th>Side-cameras</th>
</tr>
</thead>
<tbody>
<tr>
<td>No S-d, No V</td>
<td>80.7% (0.72)</td>
<td>89.2% (0.53)</td>
<td>76.5% (0.79)</td>
</tr>
<tr>
<td>No S-d, V</td>
<td>79.4% (0.73)</td>
<td>92.2% (0.43)</td>
<td>73.0% (0.83)</td>
</tr>
<tr>
<td>S-d, No V</td>
<td>78.9% (0.58)</td>
<td>93.8% (0.24)</td>
<td>71.4% (0.67)</td>
</tr>
<tr>
<td>S-d, V</td>
<td>73.0% (0.64)</td>
<td>89.8% (0.34)</td>
<td>64.7% (0.67)</td>
</tr>
</tbody>
</table>

Throughout analysis, two-way ANOVA consistently exposed significant differences between POR angles, indicating that accuracy was not only dependent on POR itself, but also on the combination of POR and camera-angle. Therefore, we performed analysis between symmetrical POR ‘pairs’ in order to isolate scenarios and expose specific conditions with high impact. Our experimental design (clock-face with one central and two symmetrical side-cameras) granted identical states for opposite POR angles depending on side-camera angle: POR 1 right-camera = POR 11 left-camera, 2 left = 10 right, 3 left = 9 right, 4 left = 8 right, 5 left = 7 right. POR 6 and 12 were treated independently as appearance differences largely greatly. These logical pairings of symmetrical POR allow us to refer to the left and right side-camera viewing angles as near and far, thus permitting us to highlight dependencies using terms more appropriate to a conversational or object-focused CVE context. Before analysis, we verified the validity of the pairings by performing repeated measures two-way ANOVA on the data for the two POR within each pair, and uncovered no significant differences. Overall accuracy levels for POR pairs can be seen in Figure 3 for centre, near and far camera-angles, showing a significant negative impact of certain combinations of POR and camera-angle; in particular, 30° offset from vertical (pairs 1 and 11, 5 and 7) when viewing from the near camera, and the 60° offset from vertical (pairs 2 and 10, 4 and 8) when viewing from the far camera.

Table 2 illustrates this suggestion that socket-deformation greatly enhanced perceived authenticity, while vergence was not a determining factor.

### 3.2 Perception of Authenticity

The questionnaire sought to elicit judgments of our avatar regarding each condition’s POR identification task. Questions 1 (‘The behaviour of the avatar’s eyes appeared natural’) and 4 (‘The general appearance of the avatar was realistic’) directly addressed perception of authenticity. A critical requirement of our experimental design was to isolate animation to the eyes and surrounding areas, and thus any variation in reported levels of perceived authenticity could be solely put down to changes in eye-representation conditions. One-way ANOVA evaluation of combined data from questions 1 and 4 exposed a highly significant difference between conditions \( (P<.001) \). Post-hoc tests exposed a highly significant difference between the base condition (No Socket-deformation, no Vergence) and the Socket-deformation, no Vergence condition \( (P<.001) \), but an insignificant difference between the base condition and the No Socket-deformation, Vergence condition \( (P=.09) \). Table 2 illustrates this suggestion that socket-deformation greatly enhanced perceived authenticity, while vergence was not a determining factor.

Mean responses to question 4 indicate that the influence of socket-deformation was not limited solely to judgment of the eyes, and enhanced the perception of our avatar as a whole. There was no a significant difference found between Socket-deformation and Vergence and Socket-deformation, no Vergence conditions \( (P=.14) \), adding to the evidence that vergence was not a significant factor affecting our avatar’s authenticity. These results concur with our original hypothesis, which predicted the comparatively subtle impact of vergence conditions on eye-representation.
One-way ANOVA evaluations of responses from questions 2 and 3 (‘I could easily tell where the avatar was looking from the centre view / side views’) did not expose significant differences between conditions despite significant differences found during analysis of POR identification task data. However, a significant difference was found between conditions for the final self-assessment question 5 (’How well do you think you completed the identification task?’) (P<.05), with socket-deformation conditions earning superior assessment levels. This suggests that confidence when identifying our avatar’s POR was increased by the enhanced realism of the socket-deformation conditions, despite accuracy levels not always corresponding to this perception as discussed during analysis of POR identification task data and highlighted in Table 1.

Table 2: Mean response and (standard deviation) for questions 1, 4 and 5 on the 1–7 (negative–positive) Likert questionnaire.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Question 1</th>
<th>Question 4</th>
<th>Question 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>No S-d, No V</td>
<td>3.75 (1.48)</td>
<td>4.29 (1.32)</td>
<td>4.87 (1.24)</td>
</tr>
<tr>
<td>No S-d, V</td>
<td>3.52 (1.53)</td>
<td>4.35 (1.38)</td>
<td>4.53 (1.36)</td>
</tr>
<tr>
<td>S-d, No V</td>
<td>5.69 (1.06)</td>
<td>5.98 (1.13)</td>
<td>5.42 (0.82)</td>
</tr>
<tr>
<td>S-d, V</td>
<td>5.33 (1.17)</td>
<td>5.85 (1.14)</td>
<td>5.16 (1.10)</td>
</tr>
</tbody>
</table>

In future work we will continue to investigate the themes of perception and practicality of high-fidelity avatars. Next, we will implement this study’s conclusions in a CVE platform operating between three CAVEs, using mobile eye-trackers to drive avatar eye-gaze. The platform will allow us to investigate further within an immersive setting featuring several different high-fidelity avatar models and the addition of participant eye-gaze data. To view extra material relating to this paper, including the actual experiment, visit http://www.cs.ucl.ac.uk/staff/W.Steptoe.

ACKNOWLEDGEMENTS
This study is funded by the UK’s EPSRC. We would like to thank our partners involved in the Eye Catching project (EP/E010032/1) at the universities of Salford, Reading and Roehampton.

REFERENCES

4 CONCLUSIONS AND FUTURE WORK
We explored the influence of socket-deformation and vergence on the accuracy of POR identification and the perceived authenticity of our high-fidelity avatar. Overall analysis of the POR identification task showed that accuracy decreased under socket-deformation and vergence conditions. However, this is highly dependent on viewing camera angle and the POR itself as our post-hoc analysis explored. Judgment of our avatar’s authenticity both as a whole and when considering just the eyes was significantly enhanced by socket-deformation, while vergence was not a significant factor.

Participants’ self-performance assessments of the POR identification task were significantly higher when judging socket-deformation conditions despite an overall performance reduction during the task. This suggests that the increased representational realism gave participants more confidence in our avatar when judging authenticity during this directly related task. This contributes to evidence that as we move forward to using higher fidelity avatars, that there will be a trade-off between supporting representational realism and supporting the actual communicative task. Representational properties of avatars may be optimised depending on the type of application in order to enhance social presence or maximise practicality for collaborative work.