SphereAvatar: A Situated Display to Represent a Remote Collaborator

Oyewole Oyekoya, William Steptoe, Anthony Steed
Virtual Environments and Computer Graphics
Department of Computer Science, University College London, UK
w.oyekoya, w.steptoe, a.steed@cs.ucl.ac.uk

ABSTRACT
An emerging form of telecollaboration utilizes situated or mobile displays at a physical destination to virtually represent remote visitors. An example is a personal telepresence robot, which acts as a physical proxy for a remote visitor, and uses cameras and microphones to capture its surroundings, which are transmitted back to the visitor. We propose the use of spherical displays to represent telepresent visitors at a destination. We suggest that the use of such 360° displays in a telepresence system has two key advantages: it is possible to understand the identity of the visitor from any viewpoint; and with suitable graphical representation, it is possible to tell where the visitor is looking from any viewpoint. In this paper, we investigate how to optimally represent a visitor as an avatar on a spherical display by evaluating how varying representations are able to accurately convey head gaze.

Author Keywords
Spherical displays, remote collaboration, telepresence, telerobotics, avatars, mixed reality

ACM Classification Keywords
H.4.3 Communications Applications: Computer conferencing, teleconferencing, and videoconferencing; H.5.1 Information Interfaces and Presentation: Multimedia Information Systems—*Artificial, augmented, and virtual realities*

General Terms
Design, Experimentation, Human Factors.

INTRODUCTION
Technologically asymmetric telecollaboration systems are an emerging form of telecommunication. In these systems, one or more remote participants use computer supported means to collaborate with others who are physically located at a specific destination space. These systems are based on the paradigm that the actions and behaviors of remote participants (*visitors*) are acquired at their physical location, and transmitted and represented at the *destination* site. Acquisition technology at the visitor sites include cameras, motion capture systems, and microphones, which capture digital representations of the visitor ready for real-time transmission to the destination site. Display technology at the destination site aims to represent visitors with physical presence, and may include humanoid robots or situated displays.

Technologically asymmetric telecollaboration can be contrasted with conventional video conferencing or collaborative virtual worlds, in which each participant's experience is mediated by approximately identical technology, and thus presents similar social affordances. Video conferencing is able to faithfully represent participants' appearance across a distance, but typically employs flat displays, which compress the representation of each participant's local 3D space. Participants have access to a range of verbal and non-verbal communicational cues that are well-supported by the medium: movement of the eyes and head, turn-taking and facial expressions [12]. However, the 2D nature of standard video and the typically narrow camera field-of-view limits the spatial nature of non-verbal communication [11]. In collaborative virtual worlds, visitors are represented by avatars. Avatars represent the presence and activities of a participant and can be visualized using standard displays or projection surfaces at a destination. Avatars are capable of eliciting appropriate responses from observers. For instance, it has been shown that the attractiveness of a participant's avatar influenced how intimate participants were willing to be with a stranger [38]. The ability of a visitor to choose different avatar representations may also be helpful in distributed meetings that include a social component [29].

This paper presents a novel method of visualizing a visitor using a spherical display as part of a technologically asymmetric telecommunication platform. The sphere we have used is 16" diameter which is conveniently somewhat bigger than the human head and thus we will depict the visitor's head on all or some of the display. A key observation is that flat displays are only visible from the front, and lack the 360° view offered by spherical displays. Exploiting the unique characteristics of such displays may bestow a participant represented on the sphere with a greater degree of social presence, as observed by those at the destination. We believe that this is for two reasons. First the display is visible from all sides so the identity of the person depicted on the display can be determined by any viewer. Second, from no position is the display seen obliquely,

CHI 2012, May 5–10, 2012, Austin, TX, USA.

so it should be possible to support the display of head gaze direction more accurately.

We have chosen to focus on the display of a head because when we interact with others we pay the most attention to the face. Faces are interesting because they convey eye gaze, expressions and gestures and are used as a central channel of communication. Detecting the gaze direction of a person is important for human-computer interaction applications, such as meeting or shared collaborative workspaces. Head gaze, or the orientation of the head during communication of attention, is crucial in interaction. Humans are very good at estimating focus of attention in meeting scenarios [32]. It has been shown that the perceived direction of eye gaze can be influenced by the angle of rotation of the head [37], affirming the importance of the head as a cue to attention direction. Consequently, any new display method needs to evaluate the ability of users to discern the focus of attention based on head gaze direction.

Therefore, in this paper, we present an implementation of a novel display of a visitor's head called *SphereAvatar*. We demonstrate three different modes of presenting the visitor which use different modes of rendering and configuration: *inflated*, *normal* and *surface* (see Figure 1). SphereAvatar is part of a larger experimental distributed collaboration system discussed in [30], which provides a heterogeneous and multimodal platform for telecollaboration. In the current paper, we investigate the use of a novel spherical display technology to embody a remote visitor. We perform an experiment to measure how accurately observers are able to determine the SphereAvatar's direction of head gaze across three varying modes of representation.

In the following sections, we review related work and present the software and hardware components needed to implement SphereAvatar. This is followed by an experimental evaluation of SphereAvatar and results. Finally, we present discussions of the results, implications for future designs, conclusions and future work.

RELATED WORK

Spherical Displays

Spherical displays have been used as a multi-touch sensitive interactive surface. Benko et al. designed a multi-user, multi-touch sensitive spherical display that facilitates collaborative interactions around the sphere [4]. An infrared camera is used for touch sensing and shares the same optical path with the projector used for the display. They placed their spherical installation in three high traffic locations and observed collaborative activities from several people interacting with the sphere. They reported that the sphere's unusual shape, large size and visibility from all directions attracted large crowds (also noted in [2]). Most described the experience as "like interacting with a crystal ball". The authors [4, 5] highlight two unique characteristics of the display: borderless but finite display and visible content changes with position and height.

Spherical displays have also been explored by Kettner et al. [17] for interacting with data projected on a spherical surface. Using external projectors, they were able to physically rotate the ball in place (like a large trackball). Grossman et al. [10] used a spherical volumetric display to allow gestural interaction and visualization with the 3D data within the display, using the 360° viewing volume. The focus of these systems have been more on how to use the spherical form factor to improve interaction. The exception is Jones et al. [16], which uses an autostereoscopic light field display to present interactive 3D graphics to multiple simultaneous viewers around the display. The spinning display surface was optimized in [15] for the display of a life-sized human face for real time teleconferencing, but the setup omits views of the back of the head. The borderless display and 360° visibility of spherical displays makes for an interesting display mode for an avatar head, as observers are potentially able to understand head direction from any perspective.

Collaborative Mixed Realities

The virtuality continuum [22] describes virtual reality (VR) related display technologies in terms of their relative extents of presenting real and virtual stimuli. The spectrum ranges from the display of a real environment (for instance video) at one end to a purely synthetic virtual environment (VE) at the other. Mixed reality (MR) occupies the range of the continuum between these extrema, merging both real and virtual objects together. MR was originally considered on a per *display* basis, and later, broadened to consider the joining together of distributed physical locations to form MR *environments* [3]. When discussing MR displays it is sufficient to do so with regard to Milgram and Kishino's evolving taxonomy [22] that ranges from hand-held devices through to immersive projection technologies. However, MR environments as outlined by Benford et al. [3] bring together a range of technologies, including situated and mobile displays and capture devices, with the aim of supporting high-quality spatial telecommunication.

Several means of representing people virtually in such MR environments have been investigated. This includes the use of situated displays [36], mobile personal telepresence robots [35], and optical or video see-through Augmented Reality (AR) displays [27]. Regarding AR displays, in order to obtain an enhanced view of the real environment, an AR display would have to be worn by each user at the destination. Live capture of 3D content and simultaneous presentation, using fiducial markers and video-see-through AR, has been demonstrated by Prince et al. [25]. Video see-through displays capture the real world with video cameras mounted on the head gear, and virtual world views are electronically combined with the video stream from the camera. Disadvantages of these displays includes narrow field-of-view and low contrast. Optical see-through displays fare much better, as the physical world is seen through semi transparent mirrors placed in front of the user's eyes. These mirrors reflect the virtual world views into the user's eyes thus optically combining of real and virtual world views. There is no limitation to the wearer's field-of-view. However the obstruction of the face, although somewhat limited with the optical display, may still hamper the ability of the visitor at the remote site to maintain good face-to-face communication with users at the destination.

Telepresence Robots

In robotics, a range of systems exist to support remote telepresence. Two complementary approaches can be found: mobile robots that represent the movement of users and situated humanoids that mimic the appearance of the head or body of users. For example, telepresence robots [7, 20] can offer video conferencing with mobile capabilities. These devices tend to have a built-in flat screen to display the video stream. Lee et al. [20] conducted a field study over 2–18 months in the use of this mobile remote presence system and commented that most people could not identify the remote pilot without walking around to check the front of the side of the screen. Robots such as Robovie [13] focus more on the interactional capabilities and do not show a video representation of the controlling user.

In the field of humanoid robotics, examples such as Geminoid [28] are very human-like in both appearance and movement. They can potentially be used to represent specific visitors at a destination but they are limited in terms of their flexibility in representing other teleoperators. Research on the dynamic characteristics of humanoid robots has been limited. Research on TeleHead [33], an acoustical telepresence robot, showed that dynamic cues from head movement play important roles in auditory localization. Head movement is an important factor in telepresence, from turn-taking during conversations to both showing and observing attention.

Earlier work by Naimark demonstrated a talking head: a technique where an image is projected onto a screen whose shape physically matches the image itself [24], to enhance telepresence. Animatronic Shader Lamps Avatars (SLA) [21] takes this technique further by using cameras and projectors to capture and map both the dynamic motion and the appearance of the embodied user onto a humanoid animatronic model. The SLA's use of front projection to texture the 3D facial geometry makes it less practical than one with internal (rear) projection, as commented by the authors. The feature-points expression face robot WD-1(Waseda-Docomo face robot No.1)[14] uses rear projection to texture a real user's face onto the robot's surface but will need further miniaturization for wider use. REFLCT [19] personalizes the experience by using head-worn projective displays to overlay multiple user-specific, customized views of an avatar on a common abstract mannequin head in a shared environment, enabling projection of animated facial expressions, similar to SLA.

The most obvious question that one can ask about a humanoid robot is whether it is actually treated as if it were a real-person. A remote-controlled android system called Geminoid HI-1 [28] was judged to be more human-like and natural than a video conferencing system but it was described by users as very *uncanny*. Uncanny is a term often used to described robots and computer generated avatars; the term was used in Mori's description of the uncanny valley [23] where it was advocated that the degree of intimacy increases with a robot's humanlike appearance and behavior but that at a certain level of humanlike appearance and behavior, however, the degree of intimacy drops sharply. A spherical display could easily be integrated into a robotic platform. Compared to current telepresence robots, the visibility of the display would be a distinct advantage. The sphere display offers flexibility compared to humanoid robotics as it isn't constrained to a single head size or shape.

Perception of Head and Eye Gaze Direction

Whilst the 360° and multi-view capabilities of a spherical projection are novel, it is not clear whether observers can interpret gaze direction on closely-spaced target objects. The direction of a person's gaze is one feature that is relevant in judging objects of interest in an environment. Early work indicates that gaze direction may be perceived by both the direction in which the head is oriented and the eye's position relative to the head [8]. Previous research has focused on studies in which the eyes and the head were counter-rotated to varying degrees while maintaining fixation on the subject [8, 1]. These studies consistently showed an interaction between eye and head position in the perception of gaze direction. Gibson et al. [8] examined three head gaze conditions: head to front, left and right. In each condition, an observer at a distance of 2m gazed at seven positions in a prearranged random order, each 0.1m apart on a wall behind participants. Participants made yes or no judgment of whether or not they felt they were being looked at. The frequency distributions of 'yes' judgments showed a *head-turn* effect such that when the target's head was rotated in one direction, participants' judgments tended to perceive gaze to be rotated in the opposite direction. In addition to the three head gaze conditions, Anstis et al. [1] investigated three orientations of a TV screen. They found: the same *head-turn* effect; *TV-screen-turn* effect, apparent displacement of the perceived direction in the same direction as the turn of the screen; and an overestimation of the deviation of looker's gaze from the straight ahead. They suggested that the convex curvature of the screen probably caused the *TV-screen-turn* effect. Overestimation was found to increase with the complexity of the viewing condition. Overall, these studies suggest that observers may be constructing a mental line based on the head orientation before judging the eye direction relative to the head.

Despite the importance of the head as an attentional cue, there has been relatively little research on the perception of its orientation. Troje and Siebeck [34] quantified accuracy of head orientation discrimination under varying illumination conditions with the eyes pointing directly forwards. Discrimination was shown to be most accurate within the $\pm 15^\circ$ range of forward gaze directions but was markedly poorer at 30° head rotation. This was also observed by Wilson et al., who also found that changes in head orientation could be perceived even when the internal features of the head or the outline head contour is removed, suggesting that the deviation of nose angle may be a likely cue [37].

Troje and Siebeck [34] also concluded in their third experiment that subjects were not influenced by potential changes in eye gaze direction but in fact judged the head orientations as instructed. Perception of an avatar's gaze direction has also been studied in virtual environments [26, 31]. In an object-focused multiparty immersive collaborative virtual environment scenario, tracked eye gaze has been shown not to provide statistically significant advantage over just tracked head gaze

[31]. With both tracked eye and head gaze, avatars' eyes and heads were controlled by head-mounted mobile eye trackers and head tracking worn by participants, while head tracked avatars featured static centered eyes with no gaze control, so visual attention must be inferred from head orientation only. Therefore, in this initial study into the use of spherical displays for representing remote participants, we employ the static gaze condition in evaluating SphereAvatar, although the underlying system supports full eye gaze as well as facial expressions.

SYSTEM DESIGN

The SphereAvatar system is a part of a larger distributed system [30] supporting both avatar- and video-mediated communication. The aim of the overall system is to allow a visitor to have multiple presences in a real environment through fixed and mobile display units. The visitor is tracked using motion capture and video systems. In this study, we focus specifically on one aspect of the system: the forms of representation of the visitor that are possible on a spherical display. The spherical display is chosen because it is small enough to situate almost anywhere in a room, but it is visible from all directions. For example, it could be positioned in a seat around a table and all others at the table would see a wide aspect of the display unlike they would with a flat panel display where even if the display could rotate, some participants would see face-on views whereas others would see oblique views.

The SphereAvatar system has the ability to show video captured content and computer generated content. However for the purposes of the experiment described in this paper, we have chosen to focus on a photorealistic computer generated head so as to provide reproducibility of cues between the conditions. In the discussion section, we describe how we can support novel real-time video-based rendering using multi-view video or could support 3D reconstruction: however these are both research topics and the visual quality achievable is not fairly comparable. Using a rendered avatar allows us to consider the ideal representations for a human head on such a spherical display without potential confounds because the detail of the representations is significantly different. The computer generated head is also precisely controllable whereas video content would have to be captured from a person performing the actions.

Hardware

SphereAvatar uses the commercially available Magic Planet display by Global Imagination®. The Magic Planet is a projection display device with a 16" sphere-shaped surface. The spherical surface is an empty plastic ball coated with a diffuse material that serves as a passive curved projector screen. It features a standard 1024×768 projector at 60Hz, coupled with a fisheye lens to project imagery. The projected light travels through the bottom of the sphere. Hence the sphere is completely illuminated except for the area immediately around the lens itself. The unique aspect of such displays are their 360° horizontal visibility, allowing the displayed avatar head to face any direction in the destination.

Software

In order to project an avatar's head onto the sphere there are two main stages: first the scene is rendered in to an environment map, and then a 3D sphere is drawn using the environment map as its sole texture. The second stage of the process is the same for each display mode. The first stage is different and the various methods are explained in the following sections.

An environment map is an image that represents the complete scene around a point. Greene proposed the idea of storing environment maps as cube maps [9] where six subimages representing the six different faces of a cube. Figure 2(a) depicts a cube map where every cube map face is a different solid color. This corresponds to scene which comprises a cube where each wall was a different color, left yellow, top blue, etc. The output of the second stage is to create an image from this environment map that can be projected through the fisheye lens. With the given environment the image to project should be as shown in Figure 2(b). The most common way of performing the relevant distortion is to use environment mapping, and to render a sphere with the environment map as its texture. Environment mapping was first proposed by Blinn and Newell [6]. It simulates the reflectance of a surface, by using the reflected eye vector as a lookup in to the texture rather than a simple texture coordinate. OpenGL and Direct3D have both supported environment mapping for many years, and it is built in to most recent graphics processing units on graphics cards. We have implemented this process in OpenGL and GLUT. We render the scene in to an environment map using six cameras, but see the following sections for how these cameras and the objects are configured. The environment map is rendered in to a frame buffer object (FBO) which means the rendering process can be real-time. This two-stage rendering approach is a standard process and, for example, a similar implementation can be found in the open source OpenSceneGraph software. Figures 2(c) and 2(d) show example of the results of renderings of the environment mapping stage. Once projected on through the fisheye lens and onto the spherical surface these would appear similar to the images in Figure 1.

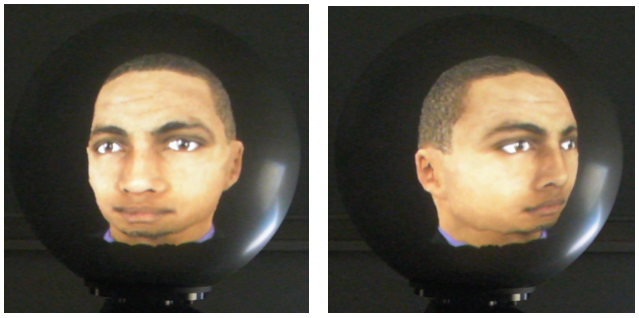
Display Modes

We implement three methods for displaying the animation of an avatar's head. As described above, in this first stage a cube map must be generated by rendering six views of the avatar head. The key difference between the three display modes is the positioning and orientation of the camera to capture the views of the six cube faces.

Inflated

With this display mode, the camera center is positioned inside the avatar's head at the center and six views are rendered facing out to each cube face. However, in this state, the camera would render the inside of the head. In order to get the correct rendering, we reverse the OpenGL depth ordering and polygon direction, so the head is rendered from the inside, but also inside out. This creates a cube map, which when environment mapped in the second stage always covers the whole sphere, see Figure 2(c). Due to the inherent differences between the shape of the human head and the sphere, this mode is

(a) Inflated Mode 1 (b) Inflated Mode 2



(c) Normal Mode 1 (d) Normal Mode 2

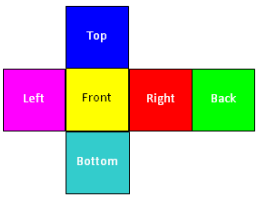
(e) Surface Mode 1 (f) Surface Mode 2

Figure 1. Examples of inflated, normal and surface display modes. In the left column the avatar head is looking at the same angle approximately $10°$ to the right of the viewer. In the right column the avatar head is at the same angle approximately $45°$ to the left of the viewer.

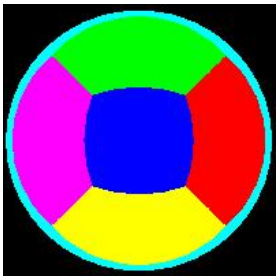
somewhat distorted on the spherical display (see Figures 1(a) and 1(b)) and the head looks like it has been blown up to a sphere. Hence the name of the display mode. Despite the distortion, all the features are locally consistent and the shape of the features in the center of the face are nearly correct. It is obviously a head and it is easy to determine the direction in which it is looking.

Normal

With this display mode (Figures 1(c) and 1(d)), the camera is positioned outside the cube at the position of the observer's head. A cube map is rendered using the now non-symmetric view volumes. The resulting cube map looks as if the head is outside looking in, but once reflected in the environment mapping, it gives the illusion that the head is situated within the spherical display.



(a) Cube map



(b) Distorted cube map for projection

(c) Inflated display mode as generated for projection

(d) Normal display mode as generated for projection

Figure 2. Illustrating stages of the rendering pipeline.

If the head were situated inside the spherical display, as an observer moved around the view of the head would change. In our implementation we can take this into account by tracking the position of the user's head using Microsoft Kinect. This mode can thus only support a single viewer as the image on the sphere can only be adjusted based on one person's position. It does however mean that the head appears solid: if the user walks around to the back of the sphere, the image will be adjusted accordingly (i.e. the back of avatar's head would be displayed at the back of the sphere). To ensure accuracy and stability, in the experiment described in the next section, head tracking was disabled and the participant sat in fixed positions. An operator keyed in the correct position.

Surface

With this display mode (Figures 1(e) and 1(f)), the avatar's head is rendered always looking straight ahead on the spherical display but rotation of the head is depicted by moving the head around the sphere. This is done by setting the cameras inside the cube as for the inflated mode, but then positioning the head outside the cube looking directly toward the center of the cube. This rendering type lends itself to rendering of face images and video as well as 3D models: the image or video could be placed on a planar billboard facing the cube center.

EXPERIMENT

In this experiment, we explore the accuracy with which participants can discriminate the SphereAvatar's head orientation for all three display modes. Specifically we measure the ability of participants to identify which of a set of targets the avatar head appears to be gazing toward. Given the three display modes, we expect that participants viewing the normal mode will be able to identify more correct targets compared to those

viewing the surface mode. We expect the inflated mode to lie between these two in performance.

Method

Apparatus and Materials

We captured a visitor's head motion in a CAVE-like system with Intersense IS-900 head tracking when looking at virtual balls in a prearranged random order (Table 1). The balls (stimuli) were placed in a 7×7 grid, excluding the central position, hence producing 48 target positions. The balls were 5.5cm in diameter, and were spaced 0.3m ($9.46°$) apart both horizontally and vertically. The most extreme head orientation to the outer-most ball horizontally or vertically was $26.56°$, while the diagonal extreme was $35.26°$. As the visitor moved his head during the recording process, he would call out a target number currently being looked at (i.e. "Target 1", "Target 2", ... , "Target 48").

Table 1. Target Order

23	38	3	9	44	5	20
11	17	24	32	27	47	28
29	42	31	40	14	25	4
35	2	37	-	34	46	12
18	48	33	6	19	16	41
7	15	26	43	30	1	22
39	45	10	21	8	36	13

A real environment was set up to accurately match the virtual environment (Figures 3 and 4). Alignment between the two frames of reference as seen at the visitor site and the destination site ensures consistency between the visitor's head movements as he perceives the virtual destination and the where the embodied avatar head at the physical destination is objectively facing. At the physical destination site, the spherical display is used to visualize the avatar head. 48 real balls (with the same radius as their virtual counterparts) were hung from the ceiling with thin thread in the same locations as the match the virtual balls. To improve discriminability, the balls were color-coded by horizontal rows in the following order: green, blue, red, yellow, red, blue, green. The target balls were situated centrally between the spherical display and the participants' seated position at a distance of 1.8m to each. There were three seating positions for the participants with a horizontal spacing of 0.9m, enabling the investigation of viewing effects from different angles. The viewing distance from participant to avatar for the center position was 3.6m. We ensured that participants' eye level were roughly the same as the avatar's eye position by seating them on a chair.

Design

Each participant judged only one of three display modes (between subjects design), but a within-subjects design was employed regarding the two factors of seating position (left, center or right) and the 48 target positions. The target positions were split into three groups: 1–16, 17–32, 33–48. Using a counterbalanced measures design, we mixed the three seating positions in order to reduce any confounding influence of the orderings such as learning effects or fatigue.

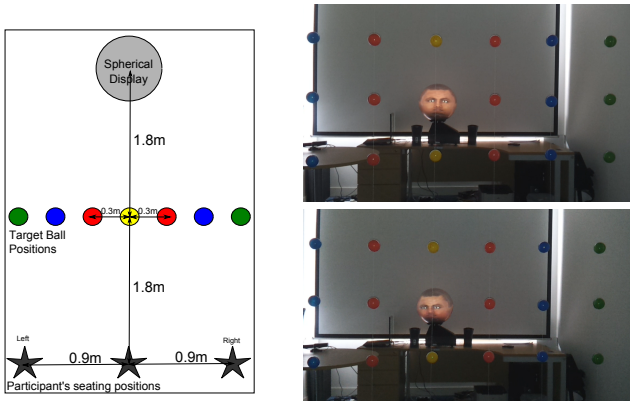
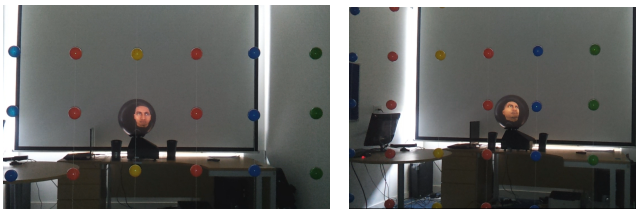


Figure 3. Experiment Setup: (a) Schematic Layout, (b) Inflated Views.



(a) Rotated Surface Display Mode (b) Rotated Normal Display Mode (Right seating position)

Figure 4. Picture of experiment room taken from different perspective of the participant

Procedure

Thirty-six unpaid participants made up of students and staff were recruited to take part in the study. Each participant only judged one of three display modes. Participants were initially seated in one of the three positions depending on the ordering that applied. Participants were given a sheet of paper with an empty 7×7 grid. Each time the avatar's head reoriented to face a new target and uttered a target number, the participant was instructed to judge which target was being gazed at, and write the uttered target number in the corresponding position in their grid sheet. They were advised that they could write multiple targets in the same grid location if that is where they perceived the avatar to be facing. Following each target gaze iteration, the avatar head returns to the central gaze position and pauses for around 10 seconds before proceeding to gaze at the next target. This provided time for participants write their judgment in the grid. At each 16^{th} iteration, the session was paused to allow the participants to change to the according seating position. Additionally for the normal mode, the viewing frustum was manually adjusted by $\pm 14°$ depending on the left or right seating position in order to correct the rendering perspective as discussed earlier.

Results

Accuracy

The accuracy of participants' judgments was tabulated. The dependent variable data (accuracy) were entered into a mixed design Analysis of Variance (ANOVA) with the three factors of display mode, seating position, and target position. There was

a significant main effect of display mode, ($F_{(2,9)} = 9.692, p < 0.01$) with higher accuracies for the inflated (*Mean, M = 0.464*) and normal modes (*M = 0.465*) than the surface mode (*M = 0.248*). Post-hoc Tukey tests revealed significant mean differences between surface and inflated modes ($p = 0.011$)) and between surface and normal modes ($p = 0.010$)). There was no difference between normal and inflated modes ($p = 1.000$)). Both normal and inflated outperformed the surface mode. We employed Mauchly's test of sphericity to validate our repeated measures factor ANOVAs, thus ensuring that variances for each set of difference scores are equal. Mauchly's test indicated that the assumption of sphericity had not been violated ($\chi^2 = 2.652, p > 0.05$). The main effect of seating position was not significant ($F_{(2,8)} = 1.541, p => 0.05$), with similar levels of accuracy for left (*M = 0.405*), center (*M = 0.410*) and right (*M = 0.363*) positions. The main effect of target position was significant ($F_{(47,423)} = 2.575, p < 0.0001$). While this absolute accuracy is a good basic measure, it reduces measurements to the binary rating scale of correct or incorrect. In order to get a more detailed view of accuracy, horizontal and vertical errors must be taken into account.

Horizontal Errors

Horizontal error was determined by the difference between the correct target and the misjudged target position. The dependent variable data (horizontal error) were entered into a mixed design ANOVA with the three factors of display mode, seating position, and target position. There was a significant main effect of display mode, ($F_{(2,9)} = 6.945, p < 0.05$) with lower horizontal errors for the inflated (*M = 0.299*) and normal modes (*M = 0.344*) than the surface mode (*M = 0.785*). Post-hoc Tukey tests revealed significant mean differences between surface and inflated modes ($p = 0.020$)) and between surface and normal modes ($p = 0.033$)). There was no difference between normal and inflated modes ($p = 0.948$)). Both normal and inflated had significantly lower horizontal errors than the surface mode. Mauchly's test indicated that the assumption of sphericity had not been violated ($\chi^2 = 1.025, p > 0.05$). The main effect of seating position was not significant ($F_{(2,8)} = 2.217, p => 0.05$) with similar horizontal errors for left (*M = 0.434*), center (*M = 0.446*) and right (*M = 0.547*) positions. As expected, there was a significant main effect of target positions ($F_{(47,423)} = 2.839, p < 0.0001$).

Vertical Errors

Vertical error was determined by how many balls away from the the correct target ball to the misjudged target ball position vertically. The dependent variable data (vertical error) were also entered into a mixed design ANOVA with three factors of display mode, seating position, and target position. There was a significant main effect of display mode, ($F_{(2,9)} = 11.365, p < 0.05$) with lower vertical errors for the inflated (*M = 0.431*) and normal modes (*M = 0.378*) than the surface mode (*M = 0.856*). Post-hoc Tukey tests revealed significant mean differences between surface and inflated modes ($p = 0.010$)) and between surface and normal modes ($p = 0.005$)). There was no difference between normal and inflated modes ($p = 0.885$)). Both normal and inflated had significantly lower vertical errors than the surface mode. Mauchly's test indicated that the assumption of sphericity had not been violated ($\chi^2 = 0.687, p > 0.05$). The main effect of the seating positions was not significant ($F_{(2,8)} = 3.393, p > 0.05$) with similar vertical errors for left (*M = 0.519*), center (*M = 0.550*) and right (*M = 0.595*). Regardless of seating positions, the level of vertical errors did not significantly differ. There was a significant main effect of target positions ($F_{(47,423)} = 6.042, p < 0.0001$).

DISCUSSION

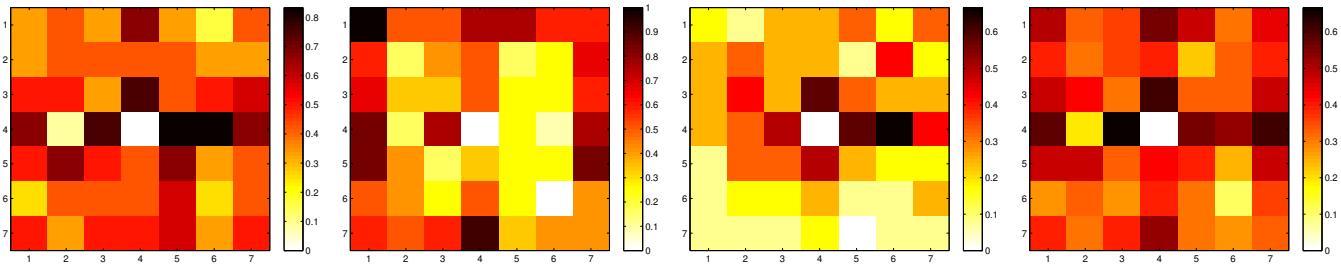
The main findings of the study are:

- Participants' could interpret the direction of the normal and inflated modes more accurately than the surface mode. This was also backed up by results from the vertical and horizontal errors.
- The differences in accuracy and errors between the normal and inflated modes were not statistically significant.
- The differences in accuracy and errors between the left, center and right seating positions were not statistically significant.

In the experiment, participants judged the direction of three different display modes of an animated avatar's head rendered a spherical display, which fixated on 48 target points arranged $9.46°$ apart with a maximum horizontal or vertical range of $26.56°$ and maximum diagonal range of $35.26°$. This allowed us to explore how subjects perceived a range of horizontal and vertical head orientations. We also computed an estimate of the mean angular error from the vertical and horizontal errors of each display mode. The mean angular error for the inflated mode was $5.00°$, normal mode was $4.87°$, and surface mode was $10.96°$. The average deviation of participants' inaccurate judgments of the surface mode was twice as large as the other two modes.

An analysis of the heat maps in Figure 5 shows that subjects' accuracy when viewing the inflated mode was more evenly distributed (Figure 5(a)) than when viewing the normal mode, which resulted in higher accuracy when viewing the edges of the grid than the more central locations (Figure 5(b)). The surface mode was biased subjects' judgments toward the central targets (Figure 5(c)).

Our results should be discussed in relation to the finding by Wilson et al. that internal features or outline head contour can be used to perceive changes in head orientation [37]. The third experiment reported by Wilson et al. investigated what cues subjects use to discriminate head orientation. The example stimuli presented in the paper shows a face with the head contour alone and another with the isolated internal features (eyebrows, eyes, nose and lips). The authors found no statistically significant difference between the full face, head and features conditions. They concluded that both the head contour and internal features provide cues of equivalent strength. Likewise in our findings, the inflated mode does not have a defined outline head contour, as the display hardware is always spherical regardless of the rendered head orientation. However, the avatar does have well-defined internal features that participants successfully used to make good judgments. The normal mode has defined internal features and outline head contour



(a) Accuracy - Inflated (b) Accuracy - Normal (c) Accuracy - Surface (d) Accuracy - Combined

Figure 5. Heat maps showing the level of accuracy for each display mode and target position.

that changes with head orientations which helped participants when making judgments. The lack of statistical significance in difference we found between these two display modes is similar to Wilson et al.'s findings. We conclude that the internal features of the face may provide sufficient information for discerning direction of head gaze. Unlike the normal display mode, the internal features and the outline head contour of the surface mode always remains the same regardless of orientation, which is likely to have reduced its effectiveness in discerning targets.

System Design Implications

Head Representation

Perhaps the most surprising aspect of our study is that the inflated avatar head is as accurate as the normal head display mode. We expected the normal mode to be the most accurate because, to the observer, it has the same dimensions as an average human head and has a similar appearance. The inflated mode was originally added as an interim test of the rendering pipeline, and although casual visitors to the lab who have seen this display have described it as being a bit unusual, they have not been averse to interacting with it. One potential advantage of the inflated mode is that it is larger than life and thus features on the display are perhaps easier to see than the normal or surface modes, thus it might be better for use in a larger space.

The poor performance of the surface mode was expected because the head is always rendered exactly face-on, and as it moves to the side of the display (i.e. the head turns) the number of visible pixels from the experimental seating positions reduces. Although greater angles such as 90° were not considered, in such a case, the head would only be partially visible from the front, as it would be positioned on the side of the display.

A related issue is the performance with multiple viewers. The inflated display mode has the distinct advantage that the head is always visible from all angles. Thus any number of observers can watch the sphere and can determine the direction the head is looking. This immediately solves a problem identified by Lee et al. [20] in their study of a mobile remote telepresence system where users could not identify the remote visitor without seeing the front of the telepresence robot. In contrast, the normal mode can only be correct for one viewer. This is because the position of the observer is needed in order to render the head correctly for that perspective. The surface display mode also has visibility problems with multiple viewers, though they are slightly different from those encountered with the normal display mode. Specifically, as the surface mode turns through a complete horizontal revolution, it is sometimes visible and invisible to any viewer, whereas as the normal mode turns, it is always visible to the one selected viewer and it may or may not be visible to the other viewers.

View Generation

With our current demonstration we are using computer generated renderings of avatar heads. Although the animation we have used in the experiment is simple, the software system provides a fully animated head with mouth movement, eye movement and facial expression. It can also render an avatar body, but since this would not be visible on the inflated display, we did not consider this in the current study. All three display modes are easily generated from 3D models.

An interesting question is the potential support for live video streaming. Telepresence robots have, so far, generally used flat screens, with a webcam view of the remote participant. This webcam view could be rendered on to a spherical display and oriented, independent of the robot base, to face in any direction. This would support more rapid head movement than turning the base itself. This could help in social situations where attention needs to be directed quickly. The direction of this surface video view could be driven in multiple ways, including following the eye or head direction of the visitor.

Given that determining gaze direction from the surface display mode is difficult, this suggests that image-based rendering or computer vision reconstruction techniques might be appropriate in order to construct normal or inflated video-based views. An equivalent normal display mode could be supported using multiple video cameras that surround the visitor. The correct video could then be selected to be shown on the sphere, by choosing the camera whose viewing angle was closest to the user viewing the SphereAvatar. The construction of a video equivalent to the inflated mode is not so simple and would require reconstruction of some proxy geometry for the head (perhaps a spheroid) that could then be re-rendered. This seems eminently tractable given the current state of the art in computer vision for reconstruction of objects from video

(e.g. [18] focuses on full body avatars, but similar approaches would work for the head or head and shoulders).

CONCLUSIONS AND FUTURE WORK

We have presented a novel display system for technologically asymmetric telecollaboration called the SphereAvatar. It comprises a spherical display on to which we project an avatar representation of a remote visitor. The SphereAvatar can represent the identity and presence of the remote visitor and we have also shown that it can successfully convey that person's direction of attention. We demonstrated three potential ways of rendering the remote visitor's head, and have discussed how these afford support for different numbers of people viewing the display. We have also discussed how these modes might be created directly from video. The key advantage of using a sphere is that it can be seen from all directions, and for no observer is the display at an oblique angle. Perhaps the most surprising finding was that the inflated display mode was very successful at conveying direction of head gaze. This is despite the inflated mode showing a head that was overly large and distorted to fit a sphere. It is thus a promising line of development. We suspect that the normal display mode which we demonstrate might still be preferable for a single user viewing the display, but it can't support multiple viewers.

The repetitive nature of the experimental task, combined with the minimal avatar animation, which did not feature eye movement or facial expression, were critical to the core aims of the experiment, which sought to determine how accurately people could identify direction of head gaze. This scenario is clearly an unusual one to perform over a telepresence system built to support normal remote interaction, and hence, we decided not to collect subjective opinions of the SphereAvatar from experimental subjects. Future studies will be designed to enable collection of such subjective data. We will also concentrate on analyzing the head-eye coordination of this display medium. Another limitation of this study is the lack of investigation along the depth dimension, as targets were arranged in a 2D plane.

SphereAvatar is technically quite simple to build and can be constructed very cheaply in comparison to volumetric displays, robotics and animatronic shader lamp avatars. An alternative approach would have been to project onto an ellipsoid that is more "head-shaped" than a sphere, however this would have worked for head rotations around the vertical axis (heading/yaw) while the projection would be severely distorted for rotations around other axes.

SphereAvatar can be statically situated or it could be mounted on a robot. In future work we want to investigate the use of a mobile SphereAvatar and how this might alleviate the need to have the robot continually turn to face the target of interest in order to properly convey the remote visitor's interest. We will also work on integration of video-based representations. There is also more work to be done on altering the rendering modes or systematically changing cameras so that the poor performance of the surface display mode is raised. Subtle shading or exaggerating the angle of the head might better represent head gaze in the surface display mode, but this might restrict it to being a single-view only display. Finally, in the larger context of heterogeneous multimodal mixed reality telecommunication [30], we will study how we might have a static or mobile SphereAvatar represent the dynamic free movements of a remote person. Thus upcoming experiments will focus on case studies of more complex telecollaboration scenarios.

REFERENCES

1. Anstis, S., Mayhew, J., and Morley, T. The perception of where a face or television'portrait'is looking. *The American Journal of Psychology 82*, 4 (1969), 474–489.
2. Barthel, C., and Rowe, S. Visitor interactions with 3-d visualizations on a spherical display at a science museum. In *OCEANS 2008* (sept. 2008), 1.
3. Benford, S., Greenhalgh, C., Reynard, G., Brown, C., and Koleva, B. Understanding and constructing shared spaces with mixed-reality boundaries. *ACM Transactions on computer-human interaction (TOCHI) 5*, 3 (1998), 185–223.
4. Benko, H., Wilson, A., and Balakrishnan, R. Sphere: multi-touch interactions on a spherical display. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, ACM (2008), 77–86.
5. Benko, H., and Wilson, A. D. *Design Challenges of Interactive Spherical User Interfaces*. 2009, 1–4.
6. Blinn, J., and Newell, M. Texture and reflection in computer generated images. *Communications of the ACM 19*, 10 (1976), 542–547.
7. Desai, M., Tsui, K., Yanco, H., and Uhlik, C. Essential features of telepresence robots. In *Technologies for Practical Robot Applications (TePRA), 2011 IEEE Conference on*, IEEE (2011), 15–20.
8. Gibson, J., and Pick, A. Perception of another person's looking behavior. *The American Journal of Psychology 76*, 3 (1963), 386–394.
9. Greene, N. Environment mapping and other applications of world projections. *Computer Graphics and Applications, IEEE 6*, 11 (nov. 1986), 21 –29.
10. Grossman, T., Wigdor, D., and Balakrishnan, R. Multi-finger gestural interaction with 3d volumetric displays. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, ACM (2004), 61–70.
11. Hauber, J., Regenbrecht, H., Billinghurst, M., and Cockburn, A. Spatiality in videoconferencing: trade-offs between efficiency and social presence. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, ACM (2006), 413–422.
12. Isaacs, E., and Tang, J. What video can and cannot do for collaboration: a case study. *Multimedia Systems 2*, 2 (1994), 63–73.

13. Ishiguro, H., Ono, T., Imai, M., Maeda, T., Kanda, T., and Nakatsu, R. Robovie: an interactive humanoid robot. *Industrial robot: An international journal 28*, 6 (2001), 498–504.

14. Itoh, K., Miwa, H., Onishi, Y., Imanishi, K., Hayashi, K., and Takanishi, A. Development of face robot to express the individual face by optimizing the facial features. In *Humanoid Robots, 2005 5th IEEE-RAS International Conference on*, IEEE (2005), 412–417.

15. Jones, A., Lang, M., Fyffe, G., Yu, X., Busch, J., McDowall, I., Bolas, M., and Debevec, P. Achieving eye contact in a one-to-many 3d video teleconferencing system. In *ACM Transactions on Graphics (TOG)*, vol. 28, ACM (2009), 64.

16. Jones, A., McDowall, I., Yamada, H., Bolas, M., and Debevec, P. Rendering for an interactive 360 light field display. *ACM Transactions on Graphics (TOG) 26*, 3 (2007), 40.

17. Kettner, S., Madden, C., and Ziegler, R. Direct rotational interaction with a spherical projection. In *In Interaction: Systems, Practice and Theory Proceedings* (2004).

18. Knoblauch, D., Font, P. M., and Kuester, F. Virtualizeme: Real-time avatar creation for tele-immersion environments. In *IEEE Virtual Reality 2010 Proceedings*, IEEE (2010), 279–280.

19. Krum, D., Suma, E., and Bolas, M. Augmented reality using personal projection and retroreflection. *Personal and Ubiquitous Computing*, 1–10. 10.1007/s00779-011-0374-4.

20. Lee, M. K., and Takayama, L. "now, i have a body": uses and social norms for mobile remote presence in the workplace. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, ACM (New York, NY, USA, 2011), 33–42.

21. Lincoln, P., Welch, G., Nashel, A., State, A., Ilie, A., and Fuchs, H. Animatronic shader lamps avatars. *Virtual Reality* (2009), 1–14.

22. Milgram, P., and Kishino, F. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems E series D 77* (1994), 1321–1321.

23. Mori, M. The uncanny valley. *Energy 7*, 4 (1970), 33–35.

24. Naimark, M. Elements of realspace imaging: A proposed taxonomy. In *Proc. SPIE*, vol. 1457 (1991).

25. Prince, S., Cheok, A., Farbiz, F., Williamson, T., Johnson, N., Billinghurst, M., and Kato, H. 3d live: Real time captured content for mixed reality. In *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on*, IEEE (2002), 7–317.

26. Roberts, D., Wolff, R., Rae, J., Steed, A., Aspin, R., McIntyre, M., Pena, A., Oyekoya, O., and Steptoe, W. Communicating eye-gaze across a distance: Comparing an eye-gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, IEEE (2009), 135–142.

27. Rolland, J., and Fuchs, H. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators & Virtual Environments 9*, 3 (2000), 287–309.

28. Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H., and Hagita, N. Android as a telecommunication medium with a human-like presence. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, ACM (2007), 193–200.

29. Shami, N., Cheng, L., Rohall, S., Sempere, A., and Patterson, J. Avatars meet meetings: Design issues in integrating avatars in distributed corporate meetings. In *Proceedings of the 16th ACM international conference on Supporting group work*, ACM (2010), 35–44.

30. Steptoe, W., Normand, J. Oyekoya, O., Pece, F., Giannopoulos, E., Tecchia, F., Steed, A., and Slater, M. Acting in collaborative multimodal mixed reality environments. *Presence: Teleoperators & Virtual Environments* (In Press).

31. Steptoe, W., Oyekoya, O., Murgia, A., Wolff, R., Rae, J., Guimaraes, E., Roberts, D., and Steed, A. Eye tracking for avatar eye gaze control during object-focused multiparty interaction in immersive collaborative virtual environments. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, IEEE (2009), 83–90.

32. Stiefelhagen, R., and Zhu, J. Head orientation and gaze direction in meetings. In *CHI'02 extended abstracts on Human factors in computing systems*, ACM (2002), 858–859.

33. Toshima, I., and Aoki, S. The effect of head movement on sound localization in an acoustical telepresence robot: Telehead. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, IEEE (2006), 872–877.

34. Troje, N., and Siebeck, U. Illumination-induced apparent shift in orientation of human heads. *PERCEPTION-LONDON- 27* (1998), 671–680.

35. Tsui, K., Desai, M., Yanco, H., and Uhlik, C. Exploring use cases for telepresence robots. In *Proceedings of the 6th international conference on Human-robot interaction*, ACM (2011), 11–18.

36. Venolia, G., Tang, J., Cervantes, R., Bly, S., Robertson, G., Lee, B., and Inkpen, K. Embodied social proxy: mediating interpersonal connection in hub-and-satellite teams. In *Proceedings of the 28th international conference on Human factors in computing systems*, ACM (2010), 1049–1058.

37. Wilson, H., Wilkinson, F., Lin, L., and Castillo, M. Perception of head orientation. *Vision research 40*, 5 (2000), 459–472.

38. Yee, N., and Bailenson, J. The proteus effect: The effect of transformed self-representation on behavior. *Human communication research 33*, 3 (2007), 271–290.