EXPLORING THE USE OF 360 DEGREE CURVILINEAR DISPLAYS FOR THE PRESENTATION OF 3D INFORMATION

by

JOHN ANDREW BOLTON

A thesis submitted to the School of Computing

in conformity with the requirements for

the degree of Master of Science

Queen's University

Kingston, Ontario, Canada

(January 2013)

Abstract

In this thesis we examine 360 degree curvilinear displays and their potential for the display of 3D information. We present two systems: a spherical display prototype and a cylindrical display. Our spherical display prototype used the lack of set user position and natural visibility constraints of a spherical display to create a 3D display system that treated the sphere as a volumetric display through the use of 360 degree motion parallax and stereoscopy. We evaluated these properties by examining how our use of stereoscopy and 360 degree motion parallax, might aid in the preservation of basic body orientation cues and in pose estimation tasks in a telepresence application with our final cylindrical display. Results suggest the combined presence of 360 degree motion parallax and stereoscopic cues within our cylindrical display significantly improved the accuracy with which participants were able to assess gaze and hand pointing cues, and to instruct others on 3D body poses. The inclusion of 360 degree motion parallax and stereoscopic cues also led to significant increases in the sense of social presence and telepresence reported by participants.

Acknowledgments

I would like to acknowledge my thesis advisor and supervisor Dr. Roel Vertegaal for his guidance and support. Without his insight and none of this work would have been possible. I would also like to thank all of the members of the Human Media Lab, specifically Kibum Kim for his knowledge in setting up these experiments, as well as David, Doug, Marty, Jesse, and Peng. Finally, I would like to thank everyone at home. My family and friends for their constant support, Mike for all the brainstorms, Fil for always being the first pilot study participant, and Hannah for her support and constant understanding.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

With the advent of new thin-film display technologies, such as Organic LEDs and E-Ink displays, it has become conceivable that one day, any surface could be a display. This means form factors will likely not be limited to flat surfaces. Spherical and cylindrical form factors, for example, have the advantage of allowing users to access 3D data from any viewpoint. They have been prevalent in Earth globes for this reason. However, these displays are currently limited to the display of spherical information, such as maps.

These displays provide an inherent 360 degree field of view where users are free to move around the display to view new information. Additionally, compared to traditional flat screens these display types contain a bounded physical volume. These two properties, when combined with motion parallax [48] and stereoscopy can be used to create the sensation that 3D data is enclosed within the display. This allows us to present data in a fashion similar to that of a volumetric display [1] and provides unique opportunities for interaction with information that is not well suited to traditional flat displays.

1.2 Motivation

Our motivation for this research is to provide users with a way of viewing and interacting with specific information that is currently inadequately represented on traditional flat displays. We make use of the natural affordances of cylindrical and spherical displays to mimic the way we would view and interact with content as if it were physically present and contained within the volume of the display surface. This is in contrast to flat displays, which in general, only provide a

fixed window into a 3D scene. By examining telepresence as a specific scenario we hope to show that our approach provides significant advantages over traditional display setups.

1.3 Objective

In this thesis, we examine 360-degree curvilinear displays focusing on how their natural affordances provide advantages and disadvantages when compared to traditional flat panel displays. Specifically, we examine visibility constraints, the lack of set user position, and the use of 360-degree curvilinear displays as a display volume rather than display surface. We present two systems: a spherical display prototype and a cylindrical display as well as two experiments. Our experiments examined how our final cylindrical display prototype, using stereoscopy and 360 degree motion parallax, might aid in the preservation of basic body orientation cues and in pose estimation tasks in a telepresence application.

Our spherical display prototype examined the feasibility of displaying 3D information as if it were nested inside the spherical display surface. We implemented a form of 360 degree motion parallax where users could walk around the display surface to view a 3D model from all sides, making use of the sphere's lack of set user position as observed previously. Additionally, this made use of the sphere as a display volume rather than a display surface. When combined with stereoscopy this created the sensation that the 3D model was nested inside the display surface.

Based on our previous 3D display prototype, we constructed a large cylindrical to examine the feasibility of using our display setup for 3D life-size telepresence with a remote user displayed as if they were standing inside the display. Telepresence was chosen due to the prevalence of subtle nonverbal cues in face-to-face communication. If the remote participant appears to the user as if they are standing inside the display surface, many of these cues can be preserved. The cylindrical form factor was used as it can be constructed to closely match the proportions of the human body. With this system, we conducted two experiments examining two nonverbal cues used in face-to-

face communication. The first analyzed a user's ability to determine the remote user's gaze and the second, the ability to determine body posture. We compared a 2D display setting, a motion parallax display setting, and motion parallax with the addition of stereoscopy. Results suggest the combined presence of motion parallax and stereoscopic cues significantly improve the accuracy with which participants are able to assess gaze and hand pointing cues, and to instruct others on 3D body poses. The inclusion of motion parallax and stereoscopic cues also led to significant increases in the sense of social presence and telepresence reported by participants.

1.4 Contributions

This thesis contributes to the field of Human-Computer Interaction in several ways. First, we provide the design and implementation of two display prototypes, a spherical display, and a cylindrical display. Both are low-cost and easily assembled using readily available parts. Our spherical display prototype uses a novel system for viewing and interacting with 3D information as if it is nested inside the display. Previous work has used projection based spherical displays to display spherical data while our display system allows for the display of 3D information to correctly represent a 3D scene. Finally, we provide the design and implementation of a cylindrical telepresence pod based on the same hardware techniques used in our spherical display prototypes. This system provides 360 degree motion parallax with stereoscopic life-sized 3D images of users.

We provide two experiments evaluating how our cylindrical display prototype, specifically its inclusion of stereoscopy and 360 degree motion parallax, might aid in the preservation of basic body orientation cues used in deixis [53] and in pose estimation tasks. The first experiment focused on how well the system preserves gaze directional and hand pointing cues. The second experiment evaluated how well the system conveys 3D body postural cues. The presence of both 360 degree motion parallax and stereoscopic cues significantly improved the accuracy with which participants were able to assess gaze and hand pointing cues, and instruct others on 3D body

posture. These cues also led to significant increases in the sense of social presence and telepresence reported by participants.

1.5 Outline of Thesis

This thesis is presented in six chapters. The first chapter provides an introduction to the thesis topic as well as the objective and motivation behind the work presented. The second chapter provides an overview of related work in the field, focusing on curved display form factors and telepresence systems, gaze direction, and 3D perspective.

Chapter three provides an overview of our display prototypes. We provide the design and implementation of each of the systems outlining both the hardware and software used.

Chapter four examines the use of a spherical display to present 3D information as if it was nested inside the display. We discuss our sample model viewer that uses in-air gestures to manipulate 3D content displayed within the sphere.

Chapter five provides the design, implementation and results of two experiments that examine how well our cylindrical display preserves specific non-verbal cues used in face-to-face communication. The first experiment examines how well our cylindrical display system preserves gaze direction and hand pointing cues. The second, evaluates how well the system conveys 3D body postural cues.

Chapter 2

Related Work

2.1 Background

In this chapter, we will first review previous work on display form factors, focusing first on hemispherical and then fully spherical displays. Following this, we will examine volumetric and perspective correct displays. Finally, to provide background for our telepresence application we will review work from virtual telepresence systems, awareness in video conference systems, and the use of 3D perspective in telepresence.

2.2 Display Form Factor

2.2.1 Hemispherical Displays

Companje et al. [13] developed Globe4D, an interactive globe. The system consisted of a hemispherical display that could be freely rotated along all axes. The earth application on this display allowed users to rotate a globe while shifting time, allowing them to see the movements of continental drift. Globe4D could be physically rotated, with the projected image kept in sync. However, it did not feature any other touch interaction techniques. This system is an early example of an interactive hemispherical display. It made use of its unique form factor to allow interaction with information that is not well suited for traditional flat displays.

2.2.2 Spherical Displays

Holman & Vertegaal [19] examined how a spherical display could be used to edit 3D NURBS (Non-uniform Rational Basis Spline) objects. Their display was projected upon using two external projectors. Multi-touch gestures were used to create curvilinear objects. A Vicon motion capture system was used for tracking both the user's fingers and the display itself. This system was based on the metaphor of a pottery table, where the concept of spinning the display while using gestures such as poking or flattening allowed the manipulation of a 3D mesh. The non-

dominant hand was used for context, while the dominant hand was used for manipulation and deformation. A set of four interaction techniques was developed for use with the sphere. These techniques were spinning, poking, pulling, and mashing. They were used in combinations to produce specific effects. For example, a pull while spinning the display resulted in an extrusion along the entire circumference of the Blobject. This was an early example of the implementation of multi-touch gestures on a fully spherical display.

Benko et al. [5] developed a multi-touch spherical display that did not require external motion tracking, based on a 24" diameter Magic Planet system. To evaluate the system, they implemented a set of interaction techniques and reported on user behavior within four applications; a simple photo and video browser, an omni-directional data viewer, a painting application, and a Pong game. Interaction techniques consisted of dragging, local rotation, and scaling. Additional techniques such as object auto-rotation, flicking with inertia, and a command to send objects to the opposite side of the sphere were designed specifically for cooperative purposes. During their observations, they found that data spanning the entire display was problematic with multiple users. Individual users could not see others touching the display and users often became confused when seeing the results of other users' interactions. As a consequence, users would compete for control of the display. They concluded that limiting the consequences of a user's action to a specific area might be necessary. Additionally, they found that users had no master position around the sphere, and instead picked a location at random. However, they found that a single user would not move extensively from their initial position during use. Their observations seem to indicate that a spherical display presenting, at most, one hemisphere to a user is an important design consideration. Users still chose a specific work area when using a spherical display. This indicates that implementing methods for viewing information on the non-visible hemisphere is an important consideration for applications on spherical displays.

2.2.3 Cylindrical Displays

Beyer et al. [7] examined cylindrical displays as a form factor for public displays and developed a prototype display in order to compare user behaviour between a cylindrical display and a flat display. The display consisted of 8 projectors, 4 mirrors and a cylindrical rear projection screen. Several interactions were developed for the cylindrical display including a painting interaction that drew a trail of bubbles as the user moved around the cylinder. They found that viewers moved freely around the display, examining the content from different angles. They suggest that cylindrical displays should be designed to support interactions that encourage movement around the display and that content should adapt to the location of the viewer around the display.

Additionally, cylindrical displays have recently been used in digital signage [27]. Recent work by SONY [45] has examined cylindrical form factors as volumetric displays, presenting information as if it is nested inside the display surface.

2.2.4 Volumetric & Perspective Corrected Displays

In the past, the editing of 3D objects by direct gestural interaction has been studied for regular planar interfaces [35] and there has been work with motion- captured gestures to edit 3D objects displayed on a separated flat surface [44] or on a 3D volumetric Actuality's Perspecta display [1]. Sheng et al. [44] examined the use of a physical object, such as a sponge, as a proxy for the user to hold and deform while gestures were detected via motion-capture. Here, the 3D object being edited was displayed on a separate screen.

Grossman et al. [17] developed a 3D geometric model building application to demonstrate multi-finger gestural interaction on a hemispherical volumetric display. The user's fingers were tracked using a Vicon motion tracking system. Interaction techniques were designed to make use of the unique features of volumetric displays, specifically, the 360 degree nature of the viewing volume. Their set of interaction techniques consisted of SurfaceBrowser for file management, model transformations, and techniques to combine models to create scenes. SurfaceBrowser could be

rotated by scrubbing the non-dominant hand's index finger along the display surface, bringing objects into view that were originally hidden. Alternatively, the user could walk around the display to view this information. Model transformations consisted of rotation, translation, and scaling. They made the observation that the 360-degree visibility of information when using a volumetric display makes it useful for exploring collaborative multi-user interaction. Due to the unique properties of a 360 degree display, considerations of how users share space must be examined. Additionally, it was found that it was useful for the SurfaceBrowser to show information on the inner surface of the display, in effect, reducing it from a 3D volumetric display to a 2D hemispherical display.

Stavness et al. [46] developed pCubee, a perspective handheld display. Five small LCD panels were arranged to form the sides of a cube. Head-coupled perspective rendering was used to simulate the feel of real objects located inside the cube. Additionally, a real-time physics engine was implemented allowing for manipulation of the displayed objects. To evaluate their system, users were required to complete a 3D tree- tracing task. They found that bimanual interaction techniques were preferred to other interaction conditions.

2.3 Telepresence, Gaze & 3D Perspective

2.3.1 Telepresence Systems

Research initiatives in electronic transmission of human telepresence trace back to as early as the late 1940s with Rosenthal's work on half-silvered mirrors to transmit eye contact during video broadcasts [42]. In the 1970s, Negroponte developed the Talking Heads project [34]. Driven by the US government's emergency procedures prohibiting the co-location of its highest-ranking five members, Talking Heads proposed a five-site system where each site was composed of one real person and four plastic heads mounted on gimbals that replicated user head orientation. Properly registered video was projected inside a life-size translucent mask in the exact shape of the face, making the physical mask appear animated with live images. However, the system was a

mockup that, in practice, would have required head mounted cameras for appropriate registration of faces.

The BiReality system [22] consisted of a display cube at a user's location and a surrogate in a remote location. Both the remote participant and the user appeared life size to each other. The display cube provided a complete 360 degree surround view of the remote location and the surrogate's head displayed a live video of the user's head from four sides. By providing a 360 degree surround environment for both locations, the user could perform all rotations locally by rotating his or her body. This preserved gaze and eye contact at the remote location. Although this system presented a life size tele-operated robotic surrogate, only the remote user's head image was rendered realistically. As implemented, the BiReality display was not responsive to viewer position, and thus, did not support motion parallax. Rendering was used to simulate the feel of real objects located inside the cube. Additionally, a real-time physics engine was implemented allowing for manipulation of the displayed objects. To evaluate their system, users were required to complete a 3D tree-tracing task. They found that bimanual interaction techniques were preferred to other interaction conditions.

2.3.2 Gaze Direction

A lightweight approach to preserving gaze directional cues was provided by Hydra [43]. Hydra used multiple cameras, monitors, and speakers to support multiparty videoconferencing. It simulated a four-way round-table meeting by placing a camera, monitor, and speaker at the position of each remote participant, preserving both head orientation and eye contact cues. Although initial prototypes suffered from vertical parallax due to the spatial separation of the camera below the monitor, subsequent designs reduced this considerably by placing the camera directly above the display. Another limitation of Hydra was the use of small screens, which limited the size of remote participants. The size of the rendered interlocutor may indeed affect the sense of the social presence [9]. The MAJIC [38] and Videowhiteboard systems [47] projected

life size images on semi-transparent surfaces by placing cameras behind the screen. However, these systems did not support 3D stereoscopic cues or motion parallax. The GAZE [51,53] groupware system provided integral support for conveying eye gaze cues using still images. Instead of using multiple video streams, GAZE measured where each participant looked by means of a desk-mounted eye- tracking system. This technique presented a user with the unique view of each remote participant, emanating from a distinct location in space. Each persona rotated around its x and y axes in 3D space, thus simulating head movements. Later, motion video was added via the use of half-silvered mirrors in GAZE-2 [52].

2.3.3 3D Motion Parallax and Stereoscopy

A variety of technical solutions have been devised to explore the preservation of 3D depth cues and motion parallax. Harrison and Hudson presented a method for producing a simple pseudo-3D experience by providing motion parallax cues via head position tracking [18]. Their system required only a single traditional webcam at each end for both scene capture and the creation of head-coupled pseudo-3D views. This system utilized a 2D display that did not provide stereoscopic vision [55]. Some CAVE-like environments provide an immersive VR experience, providing motion parallax for a single user. They typically also require the use of shutter glasses, thus precluding the possibility of eye contact transmission. For example, Blue- C, an immersive projection and communication system [16,33], combines real-time 3D video capture and rendering from multiple cameras. Developing a novel combination of projection and acquisition hardware, it created photorealistic 3D video inlays of the user in real time [33]. The use of auto-stereoscopic display technologies [26,28,36] provides similar capabilities, but without the need for special eyewear and often, adding the ability to support multiple users simultaneously, each with their own perspective-correct view. However, these are restricted to specific optimal viewing zones and typically result in significantly reduced resolution.

We should note that the above examples all rely on planar screens, limiting the ability of users to walk around the display of a remote interlocutor as is, e.g., possible with LiteFast displays [27]. Another technology, swept-surface volumetric display [21], supports 3D display with motion parallax in a form factor often more suitable for this purpose, but recent examples have been too small to render a full human body at life size.

Although the benefits of including motion parallax and stereoscopy in the presentation of graphic interfaces have been demonstrated [48], systematic evaluation of the impact of these factors in the context of task performance during video communication, specifically, in assessing pointing or poses of a remote interlocutor, is sparse. Böcker, Rundel and Mühlbach [11] compared videoconferencing systems that provide motion parallax and stereoscopic displays. While their results suggested some evidence for increased spatial presence and greater exploration of the scene, the studies did not evaluate effects on task performance. Subsequently, the provision of motion parallax was shown to generate larger head movements in users of video conferencing systems, suggesting that users do utilize such cues [9].

Chapter 3

Display Prototypes

3.1 Overview

In this chapter we provide details on the implementation of our display prototypes. These consist of a spherical display prototype and a cylindrical display.

3.2 SnowGlobe

Our spherical display prototype, referred to as SnowGlobe, is a 3D object viewer that uses gesture input and user tracking to mimic volumetric projection, where a 3D object appears to be sitting inside the display volume. Gesture input is tracked using Microsoft Kinect depth sensitive cameras [29], which provide full skeleton information for the user within view. This allows our system to preserve motion parallax and present the correct 3D perspective of 3D objects when walking around the display. A DepthQ 3D projector is used to provide stereoscopic projection and allows objects to appear inside the display rather than on the surface.

Figure 3.1. SnowGlobe System

3.2.1 Spherical Display Surface

Our spherical display surface consists of a 90 cm diameter hollow sphere made of acrylic, with a 48 cm hole cut in the bottom (see Figure 3.1). The acrylic was sandblasted to create a diffuser on the inside, allowing it to act as a projection surface. The spherical display surface sits on a custom wooden base designed to hold the projector. The projector is aimed upwards in order to project off a 46 cm, hemispherical mirror mounted inside the top part of the sphere. This mirror projects light to the rest of the surface of the display.

3.2.2 Projection Distortion

The projected image for SnowGlobe is rendered using Microsoft's XNA 4.0 framework. A custom distortion class was developed, creating a two-dimensional semi-circular object. The distortion object is made up of 4,000 total vertices split into 10 rows of 400 vertices each. The

radius of the distortion object can be changed to account for different distances from the projector to the mirror. The *uv* texture coordinates of the distortion object are modified according to a polar transformation to account for the distortions introduced by the hemispherical mirror and the spherical display surface. At each row the *v* component of the texture coordinate is determined by a basic quadratic function, accounting for the curvature of both the mirror and spherical projection surface. A virtual camera view of a 3D model is used to texture this distortion object; this results in an undistorted projection of the 3D model on the sphere. When the user moves around the display, the virtual camera view changes to match the user's perspective and the distortion object rotates ensuring that the 3D model remains at the center of the user's field of view

3.2.3 Sensing

Microsoft Kinect Cameras are used to determine user position around the display and detect gestures. Two Kinect cameras are mounted around the sphere, allowing 360 degree tracking of a user as they move around the display. These cameras provide the user's position in real world coordinates relative to the sensor. This allows us to determine the angle at which a user is viewing the information displayed inside the sphere using the user's *z* and *x* coordinates relative to the sensors. To see 3D content on the sphere, users wear a pair of shutter glasses. The user's position coordinates are sent over the network using Open Sound Control (OSC) [49]. Changes in the angles between the user and the center of the sphere are used for motion parallax compensation, and rotate the current model and the projected image as the user moves around the display. This allows the model to appear to be at a static position inside the display.

Figure 3.2. The TeleHuman system: local user (left) interacting with remote interlocutor (right) in 3D.

3.3 TeleHuman

TeleHuman is a 3D video-based conferencing system that provides the capabilities of 3D capture, transmission, and display in a lightweight, low-cost, low-bandwidth configuration (see Figure 3.2). TeleHuman's display system uses the same underlying projection and sensing concepts as SnowGlobe, with changes to the distortion to accommodate the cylindrical form factor. The system relies on Microsoft Kinect depth sensitive cameras for capturing 360º 3D video models of users. The information to construct the model is efficiently broadcast over the network by adding a grayscale depth map frame to each frame of video. These video frames are then synthesized locally to construct the 3D representation of a user. The 3D video models are rendered with perspective correction and stereoscopy on a life-sized cylindrical display, using an off-the-shelf 3D projector (see Figure 3.3).

3.3.1 Design Rationale

Our main consideration in the design of our capture and display system was to support 3D cues. These aid in the preservation of information related to head orientation pose, gaze, and overall body posture of a human interlocutor. In this context, we identified a number of relevant design attributes:

3.3.1.1 3D Cues

TeleHuman supports 3D both through optional use of stereoscopic shutter glasses and motion parallax. The latter results in a change of view and relative shifts of objects in the visual field due to changes in the observer's tracked position, allowing users to walk around and observe a virtually projected interlocutor from any angle.

3.3.1.2 Form Factor

Providing full 360º motion parallax required the use of a cylindrical form factor display [27] proportionate to the human body. Since this offers an unobstructed 360º field of view, it enables a user to explore different perspectives by natural physical movement.

3.3.1.3 Directional Cues

Being able to determine where users are looking or pointing has been shown to be an important cue in videoconferencing [51]. These cues can help regulate conversation flow, provide feedback for understanding, and improve deixis [23, 31]. The use of 3D video models, as opposed to the direct display of a single 2D video camera output, facilitates preservation of eye contact. However, stereoscopy through shutter glasses inhibits estimation of eye orientation in bi-directional scenarios. We believed that motion parallax alone may suffice for estimation of gaze or pointing direction, as users are free to move to the location in which gaze and arm orientations align to point at the user [9].

3.3.1.4 Size

Prior work, such as Ultra-Videoconferencing [14] and that of Böcker et al. [8], suggests that to avoid misperceptions of social distance [3] and to aid in a sense of realism, preservation of body size is important [37]. This motivated the conveyance of life-size images in our design.

Figure 3.3. TeleHuman hardware: a cylindrical display surface with 6 Kinects and a 3D projector inside its base.

3.3.2 TeleHuman Implementation

Our implementation of TeleHuman revolved around the design of a cylindrical display coupled with 3D tracking and imaging. We first discuss the imaging hardware, after which we discuss software algorithms for capturing, relaying, and displaying live 3D video images.

3.3.2.1 TeleHuman Cylindrical 3D Display

Figure 3.3 shows the cylindrical display deployed in TeleHuman. The display consists of a 170 cm tall hollow cylinder with a diameter of 75 cm made of 6.3 mm thick acrylic. The cylinder was sandblasted inside and out to create a diffuse projection surface. The cylinder is mounted on top of a wooden base that holds the projector, giving the entire system a height of approximately 200 cm. These dimensions were chosen to allow for a range in size of remote participants. A DepthQ stereoscopic projector [24] is mounted at the bottom of the display, pointed upwards to reflect off a 46 cm hemispherical convex acrylic mirror. This allows projections of images across the entire surface of the cylinder. The DepthQ projector has a resolution of 1280 *x* 720 pixels. However, since only a circular portion of this image can be displayed on the surface of the cylinder, the effective resolution is described by a 720 pixel diameter circle, or 407,150 pixels.

An Nvidia 3D Vision Kit [32] is used with the projector to create an active stereoscopic display. This kit provides an IR emitter that connects to a 3-pin sync port on our system's graphics card. Compatible shutter glasses are synced with the IR emitter and projected image, refreshing at 120 Hz. As a result, when viewing the display, a distinct image is shown to each eye, and disparity between these two images creates stereoscopy. By combining depth cues with perspective corrected motion parallax [48] the remote interlocutor appears to be standing inside the cylinder.

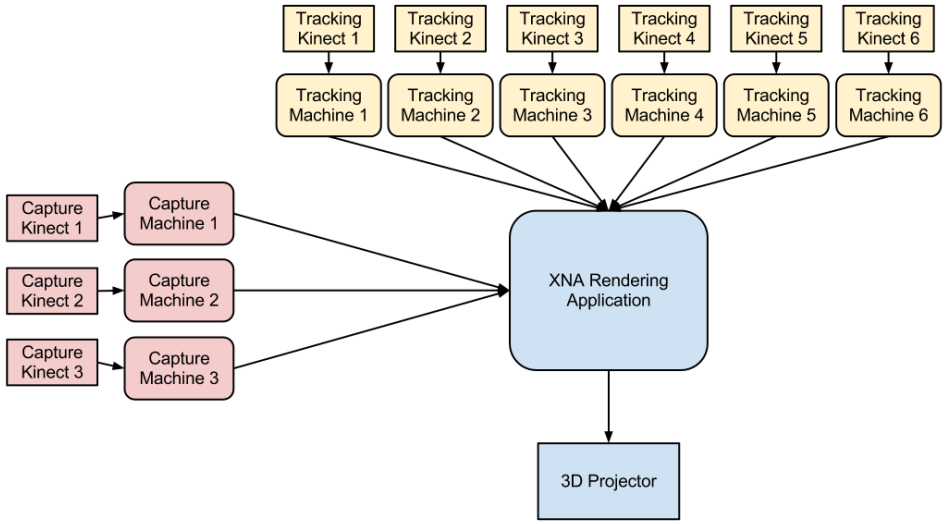

Figure 3.4. TeleHuman system diagram. The capture system is represented in red, the tracking system in yellow, and the projection system in blue

3.3.2.2 User Tracking

We used Microsoft Kinect depth-sensitive cameras [29] to determine the location of users around the cylinder. Six Kinects are mounted on the top of the cylinder, pointed downwards (see Figure 3.3). User position is represented in polar coordinates in order to easily determine the angle at which the user is viewing the information displayed by the cylinder. In the one-way setup used for our experiment the sensors on the top of the display track the location of the user 360 degrees around the display. Each Kinect is connected to a PC, which sends the user's position (both angle and distance from the sensor) via Open Sound Control [49] to a Microsoft XNA application that controls the projection (Figure 3.4). Each of these PCs calculate the user's angle relative to the Kinect sensor and provides the distance from the user to this sensor, not to the true center of the cylinder. The main application controlling the rendering of the remote interlocutor is responsible for calculating the global angle of the user based on the angular transformation between the sensors and the sensor's position relative to the centre of the cylinder. Both of these are

determined through the calibration process. The final global angle represents the user's angular position around the cylinder in the range 0 to 360 degrees.

3.3.2.2.1 Kinect Calibration

The six Kinects used to track the user around the display are calibrated using a custom calibration application in order to determine the angular differences necessary each to convert local angles at each sensor to a global angle in the range 0 to 360 degrees. To calibrate the sensors, a user stands where the field of view of two sensors overlaps. The application compares the local angles between the sensors in order to calculate the angular difference between them. This calculation uses the midpoint between the two shoulders of the skeleton data as the user's position. The z coordinate of this joint and the measured distance from the sensor to the center of the cylinder are used to determine the user's distance from the center of the cylinder. This is used as the distance component of the representation of the user's position in polar coordinates. This value is used in conjunction with the x coordinate to determine the user's local angle at the sensor through a basic arctangent operation. At each calibration point between the sensors, the application calculates the calibration 200 times to compensate for instability in the skeleton data and ensure a reliable calibration. When the angular difference is calculated at the first sensor, the user moves to the next intersection and this process is repeated. When the angular differences between each sensor have been determined, they can be used to easily transform the user's angle relative to any sensor to the user's global angle 360 degrees around the display.

3.3.2.3 Image Generation

Images from the Kinects are accessed using OpenNI [41] drivers. Each camera provides a 640 x 480 pixel stream at 30 fps with both RGB and depth images. Background subtraction is performed on both the RGB and depth images by removing pixels past a certain depth. Using the depth and RGB streams, the system calculates a four-channel image via OpenCV [40]. This image contains RGB information in the first three channels and depth information in the fourth

channel. Images are then sent via a TCP connection to the XNA projection application running on a separate machine. Three Kinect sensors are used for image capture and are arranged around the remote interlocutor to prevent overlap and the associated Kinect interference. One sensor directly faces the remote interlocutor with another at either the side of the interlocutor. Each sensor is connected to a single machine running identical software. Currently, our system sends images over a gigabit LAN connection, relying on the associated high network speeds to provide multiple live streams with low latency.

3.3.2.4 Live 3D Model Generation

In order to create a 3D representation of a user, the depth values from the TCP image stream are used to position vertices in a 3D XNA application.

Using the depth map, the XNA display application creates vertices corresponding to each pixel of the user. The depth value is used to determine the vertex locations along the *z* axis. First, the application iterates through the image stream to create a list of potential vertex locations in 3D space based on the information from the stream. Initially, the image stream is resampled to a 320 by 240 representation. This compensates for any noise in the depth data and results in a better final mesh while speeding up the indexing process used to triangulate the vertices in preparation for rendering. Once the stream has been resampled, the position of each pixel in real world coordinates, relative to the capturing Kinect sensor, are calculated. The horizontal and vertical fields of view of the sensors are used in conjunction with each pixel's *x, y,* and depth *z* coordinate to calculate these values. The resulting coordinates are assigned to an array of three dimensional vectors which hold the vertex positions. Additionally, the texture coordinates for each vertex is calculated based on its original *x* and *y* locations in the image stream.

Following this, these vertex positions are indexed in order to identify the triangles needed to construct the mesh. The vertex position array is iterated through to identify vertices within the

appropriate depth range. For each vertex in the correct range, the application looks at the vertices to its right in the array and at the same position in the row below. If these vertices are also within the appropriate depth range their indices within the vertex array are added to the index array.

Following this, vertices are placed in a vertex buffer while the previously created array of indices are added to an index buffer. The content of these buffers are read and used to render each triangle of the mesh, as indicated by the index array. The result is a mesh primitive that can be displayed by the XNA application. As the information to create the mesh is provided by three sensors placed around the remote interlocutor, three mesh primitives are created, corresponding to the 3D view of the remote interlocutor from each sensor. These meshes are aligned by manually adjusting the position until the edges correctly overlap. As the sensors are in a fixed location, the alignment procedure only needs to be done once. Based on the distance of the viewer from the cylindrical display, the model is rendered such that the center of mass of the TeleHuman appears to be in the middle of the cylinder, which we treat as the origin. The RGB values from the input image are used to texture the resulting mesh model according to the texture coordinates determined during the creation of the vertex array.

Figure 3.5. Textured 3D model with hemispherical distortion. When reflected off the convex mirror onto the cylinder, this produces a 3D model with proper proportions.

3.3.2.5 Motion Parallax and Projection Distortion

The view of a user on the cylinder is rendered from the perspective of a virtual camera targeted at his or her 3D model. The angular position of the user viewing the cylinder controls the angle with which this virtual camera looks at the 3D model of the interlocutor. Although multiple sensors can track the user at the same time, user position is calculated relative to the sensor that the user is currently closest to. This prevents slight inaccuracies in the calibration from creating a shaking effect in the image when two user positions with slightly different angles are reported. As a user's position changes, the position of the camera changes accordingly, allowing him or her to view a motion parallax corrected perspective of the 3D video model of the other user. This camera view is rendered and stored as a texture. 3D information is preserved during this process allowing the texture to be viewed with stereoscopy.

The projected image is rendered using Microsoft's XNA 4.0 framework. The same custom distortion object developed for SnowGlobe is used to render the remote participant. Different parameters are used for the distortion object, including a smaller radius to account for the greater

distance between the projector and the hemispherical mirror and the use of a cubic function to account for the change from the spherical display surface to the cylindrical. The texture coordinates of this object are modified according to a polar distortion to account for the distortions introduced by the hemispherical mirror and the cylindrical display surface. The distortion model is textured using the previously rendered camera view (Figure 3.5). To account for the different shape, compared to the spherical display, the virtual camera is rendered with an aspect ratio of 0.85 as the cylindrical screen is significantly taller than it is wide. When reflected off the hemispherical convex mirror, this creates an undistorted projection of the remote participant on the surface of the cylinder. When the user moves around the display, the distortion model ensures that the remote participant remains at the center of the user's field of view. As this projection changes based on user position, it creates a cylindrical Fish Tank VR view that preserves motion parallax [48]. Note that our approach does have the side effects of causing both resolution and brightness to drop off at lower elevations of the cylinder.

Chapter 4

SnowGlobe: A Spherical Fish-Tank VR Display

4.1 Introduction

In this chapter, we discuss the design of a 3D object viewer that uses gesture input and user tracking to mimic volumetric projection, where a 3D object appears to be sitting inside the display volume. With this project, we seek to examine the feasibility of using motion-parallax 3D projection on a sphere to display 3D objects as if they were volumetric. According to past thought experiments [4], this is categorized as a "nonplanar 2D to 3D mapping". The user can visualize the object they are viewing as being positioned at the center of the sphere, with interaction mapped to the object's rotational axes. This provides a more natural feel than traditional planar alternatives. Additionally, we allowed for scaling through interaction with the display, allowing users to focus on specific portions of the model. The vocabulary of gestural interactions was designed to be very simple, mapping directly to how a spherical surface containing a physical object would behave in the real world. We expect that this minimal interface utilizing the unique characters of a spherical display creates a compelling alternative to traditional methods of viewing 3D objects on flat displays, while providing a much more affordable and high resolution alternative to a true volumetric display.

4.2 Interactions

In air gestures are determined by tracking user hand positions and the resulting gestures are subsequently mapped to actions on the 3D object inside the sphere. We designed three basic gestural interactions for use with the sphere: Rotation (Figure 4.1), Scaling (Figure 4.2), and Scrubbing (Figure 4.3).

Figure 4.1. Rotation gesture. To rotate a model, a user raises a hand and then moves the hand horizontally or vertically. The axis of rotation is determined by whether the user moves their hand horizontally or vertically. This analogous to spinning the entire volume.

4.2.1 Rotation

Rotation of the model is limited to the x and y axes. Raising a hand and moving it along an imaginary line of latitude will rotate the model along the y axis. Raising a hand and moving along a line of longitude will rotate the model rotated along its x axis. This gesture is based on the concept of the user spinning the sphere as if it was a free-spinning globe.

Figure 4.2. Scaling gesture. To scale a model, a user raises both hands. As the distance between the hands increases, the models is enlarged, as the distance decreases, the model is shrunk.

4.2.2 Scaling

Scaling is performed as a bimanual gesture. By raising both hands the user can uniformly scale the model, using a pinching gesture. If the two hands move closer together the model is shrunk, while movement of the hands away from each other will enlarge the model. This gesture is common in multi-touch interactions and should be easily recognizable to most users.

Figure 4.3. Model switching gesture. To change models, the user preforms a waving gesture. This is designed to represent the act of scrubbing the model off the display.

4.2.3 Scrubbing

A user can change the current model by waving at the display. This gesture was designed to mimic the idea of scrubbing the model off the display.

4.3 Discussion

We believe our 3D viewer may help with the visualization of 3D objects, which can be difficult on flat panel screens. Additionally, we believe that we can duplicate the advantages of volumetric displays in a more high resolution and much less expensive ($4,000) system. The utility of our system could be improved by developing additional applications. Our model viewer does not support the deformation or manipulation of objects. We are considering additional use cases, such as virtual clay modeling.

Currently, our system only supports one user but this limitation is easily overcome by adding additional head tracking. By allowing two users to use the display concurrently, we can present two completely different views of an object making use of a spherical display's characteristic of

only providing a single viewable hemisphere to a user at a time. This would allow for private collaboration on a single display surface while maintaining a sense of two users working on a single task, which is currently not possible on a volumetric display.

Chapter 5

Effects of 3D Perspective on Gaze and Pose Estimation with a Life-size Cylindrical Telepresence Pod

5.1 Introduction

To evaluate our 3D display system, notably our implementation of 360 degree motion parallax and stereoscopy, we constructed a cylindrical display prototype based on our SnowGlobe system. The cylindrical display was constructed specifically for a telepresence application in order to present a remote participant as if they are standing inside the display surface, preserving nonverbal cues often lost in traditional videoconferencing.

Current videoconferencing systems range from the popular, low-end, small displays of Skype and FaceTime to expensive, large-screen business systems such as Cisco TelePresence and Polycom RealPresence, the latter of which can support life-size display. However, all of these systems suffer limitations in their ability to support important nonverbal communication cues such as eye contact, 3D spatial reasoning, and movement of interlocutors. The effect of these cues on remote communication may be difficult to measure, and may not affect typical parameters, such as task performance [50]. However, we believe that differences in user experience of telecommunication versus face-to-face communication may be attributed to subtle violations of such nonverbal communication [43].

Since the Talking Heads system [31], researchers have worked on preserving cues in telecommunication to enhance human telepresence [8]. However, very few systems approach the richness of direct face-to-face communication. Most only preserve a partial set of visual cues or suffer from costly and complex implementations [16]. One approach has been the use of animated 3D avatars of users [15] and head-mounted 3D virtual reality systems [51]. In such systems, a 3D

model of the user is produced once, then animated in real time by *measuring* the user's behavior. Since only animation parameters are transmitted in real time, these systems typically require little bandwidth. However, they do so at a cost in realism that results in an Uncanny Valley effect [30].

While recent advances in 3D avatar systems offer highly realistic renditions [2], we believe there are significant advantages to using 3D video instead. Video-based systems differ from avatar systems in that they capture a realistic 3D video model of the user every frame, which is then broadcast and rendered in real time across the network [16]. This results in a highly realistic replication of behavioral cues, but at a cost of network bandwidth. The capturing and transmission of 3D video has, to date, required many special considerations in terms of camera placement and projection environment [16]. The associated requirements of such environments are prohibitive for the typical workplace.

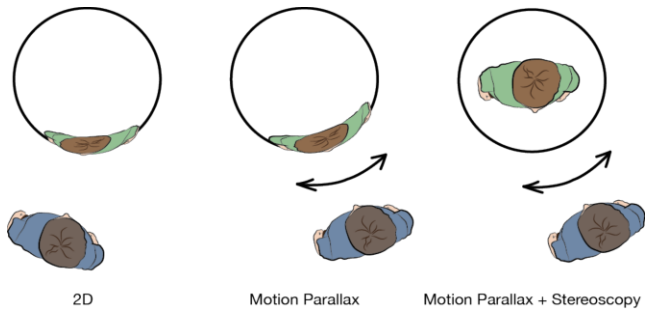

Figure 5.1. Top-view drawing of perspective conditions: conventional 2D (left), motion parallax (middle), motion parallax + stereoscopy (right). In the case of motion parallax, the display would show the remote individual from a slightly side perspective. The black circle represents the cylinder, the person with a green shirt is the perception of the remote participant. The local user is wearing a blue shirt.

5.1.1 Empirical Evaluation

We designed two experiments to evaluate effects of stereoscopy and 360º motion parallax on the preservation of nonverbal cues in our TeleHuman system. Our first experiment focused on how stereoscopy and motion parallax might aid in the preservation of basic body orientational cues. The second experiment focused on how stereoscopy and 360º motion parallax around the display might aid in conveying body postural cues.

5.1.2 Experiment 1: Effects of 3D Perspective on Gaze and Pointing Direction Estimates

The first experiment was designed to gauge the effects of motion parallax and stereoscopy on judgment of eye gaze and hand pointing by a TeleHuman 3D video model.

5.1.2.1 Task

Participants were asked to indicate where a TeleHuman model was looking or pointing. To ensure equal conditions for all participants, we used a static prerecorded TeleHuman 3D video model in all conditions. We used a simplified, asymmetrical setup in which only one TeleHuman pod was used. At each position, participants were first asked if the TeleHuman was pointing or looking directly at them. If they answered negatively, they were asked to indicate where the TeleHuman was pointing or looking, with reference to a tape measure mounted on a wall behind them. Next, participants were asked to move parallel to the wall until they were satisfied that the remote participant was looking or pointing straight at them, at which point we recorded their position.

5.1.2.2 Experiment Design

We used a within-subjects design in which we evaluated the effect of two fully factorial independent variables: *perspective* and *pointing cue.* To allow for a more realistic scenario, and a richer set of cues, we also varied the participant's location in front of the display: left, center, and right, and the TeleHuman's pointing angle: left, center and right, between conditions.

5.1.2.2.1 Perspective

The perspective factor consisted of three levels: *conventional 2D, motion parallax, motion parallax + stereoscopy* (see Figure 5.1). For the conventional condition, the TeleHuman was shown from the perspective of a front-facing camera, centered on the human. In the *motion parallax* condition, the TeleHuman was displayed with continuous perspective correction based on the location of the participant relative to the display. In the *motion parallax + stereoscopy* condition, participants additionally wore shutter glasses that provided them with a fully stereoscopic image of the TeleHuman, giving the impression that the human was *inside* the cylinder.

5.1.2.2.2 Pointing Cue

The pointing cue factor had three levels: *gaze, hand,* and *gaze + hand*. In the *gaze* condition, the TeleHuman indicated the pointing direction by both eye gaze and head orientation directed towards the same location on the wall. In the *hand* condition, the TeleHuman pointed at the target with their arm, hand and index finger. In this condition, the gaze of the TeleHuman was fixated directly to the center, unless the actual target was the center, in which case, gaze was oriented randomly to the left or right of the target. In the *gaze + hand* condition, the TeleHuman's arm, hand and index finger all pointed in the same direction as the eyes and head.

5.1.2.3 Setup and Procedure

The TeleHuman display was placed 2 m from a wall behind the participant. This wall showed a tape measure with markings at 5 cm intervals from left to right. To ensure presentation of consistent stimuli to all participants, we used a recorded still 3D image to constitute the pointing cues factor. These were rendered according to the perspective factor, as shown in Figure 5.1. For each condition, participants were asked to stand in between the display and a wall behind them, approximately 190 cm from the display and 10 cm from the wall. Participants experienced the perspective and pointing cue conditions from three locations, distributed between-conditions:

directly in front of the cylindrical display, 45 cm to its left, and 45 cm to its right. In addition, in each condition, the TeleHuman pointed in a different angle, selected from left, center, or right. Note that while pointing targets were not visible within our display setup, targets could be projected in the environment in a real videoconferencing scenario.

5.1.2.3.1 Trials

Each participant carried out a total of 9 trials, by factorial combination of 3 perspectives (*2D*, *motion parallax*, *motion parallax + stereoscopy*) with 3 pointing cues (*gaze, hand, gaze+hand*). To allow for a richer set of cues, we also varied the locations of the participant (3 locations) and the directions of pointing between conditions (3 directions). We did not perform a fully factorial presentation as it would have led to 81 trials per participant. The order of presentation of conditions was counterbalanced using a Latin square. All participants were presented with the same set of stimuli, in different orders. The experimental session lasted one hour.

5.1.2.3.2 Participants

We recruited 14 participants (mean of 21 years old, 7 male), who were paid $15 for their participation. Three of the participants wore corrective glasses.

5.1.2.3.3 Measures

We determined the mean accuracy of pointing location through two measures: 1) *visual assessment*, where participants judged where the TeleHuman was pointing without moving from their initial location; and 2) *visual alignment*, where participants moved to the location at which the TeleHuman appeared to be pointing right at them. Visual assessment allowed us to determine any effects of a more stationary perspective on the accuracy of pointing direction estimates. We expected visual alignment to provide the most accurate method for determining where the TeleHuman pointed or looked, as it allowed users to align themselves such that the TeleHuman appeared to be looking or pointing directly at them. Each measure was calculated as the angular difference between reported viewing direction and the actual TeleHuman pointing direction.

5.1.2.3.4 Questionnaire

To evaluate the degree of telepresence and social presence experienced, participants completed a seven-point Likert scale questionnaire after each perspective condition. Telepresence was defined as the feeling of "being there", while social presence was defined as the perceived ability to connect with people through the medium. In the questionnaire, a 1 corresponded to strongly agree and 7 to strongly disagree.

Perspective	2D	Motion Parallax	Motion Parallax + Stereoscopy
Visual Assessment	15.3° (1.6)	11.5° (1.5)	8.4° (1.2)
Visual Alignment	21.6° (1.9)	5.2° (.89)	3.9° (.43)

Table 5.1. Angular mean difference between actual and reported target locations and standard error (s.e.) in degrees.

5.1.2.4 Results

All results were analyzed using a within-subjects analysis of variance (ANOVA), evaluated at an alpha level of .05.

5.1.2.4.1 Pointing Location Estimation

Table 5.1 shows the accuracy of *pointing location* estimates for our two measures: visual assessment and visual alignment.

5.1.2.4.1.1 Visual Assessment

Results for visual assessment of pointing direction show a significant main effect of *perspective* on accuracy ($F(2,26)=6.35$, $p=0.006$), but no significant effect for *pointing cues* ($F(2,26)=1.92$, $p=0.17$). Bonferroni post-hoc tests showed that mean accuracy of visual assessment was

significantly greater in the *motion parallax + stereoscopy* condition than in the *conventional 2D* condition (*p=0.009*). However, there were no significant differences between other conditions.

5.1.2.4.1.2 Visual Alignment

Results for visual alignment show a significant main effect for *perspective* ($F(2,26)=66.51$, *p<0.001),* but not for *pointing cues* ($F(2,26)=0.88$, *p=0.425*). Post-hoc pairwise Bonferroni corrected comparisons of the *perspective* conditions show that mean accuracy was significantly greater in the *motion parallax* condition (*p<0.001*) and in the *motion parallax + stereoscopy* condition (*p<0.001*), compared to the *conventional 2D* condition. There was no significant difference between the *motion parallax* and *motion parallax + stereoscopy* conditions *(p=0.71)*.

Statements	2D	Motion Parallax	Motion Parallax + Stereoscopy
It was as if I was facing the partner in the same room. (S1)	4.21 (2.0)	3.21 (1.8)	3.14 (2.0)
My partner seemed a real person. (S2)	4.43 (2.3)	3.86 (2.0)	3.36 (2.2)
I felt immersed in the environment. (T1)	4.07 (1.9)	3.14 (2.1)	2.64 (1.8)
I felt surrounded by the environment. (T2)	4.00 (2.1)	3.21 (1.9)	2.50 (1.4)

Table 5.2. Means and standard errors (s.e.) for social presence (S) and telepresence (T) scores. Lower scores indicate stronger agreement.

5.1.2.4.2 Questionnaire

Table 5.2 summarizes the answers to each question for each of the three perspective conditions presented. A Friedman test indicated that there were significant differences between perspective conditions in S1 “It was as if I was facing the partner in the same room” ($\chi2(2)=6.69$, p=0.035), S2 “My partner seemed a real person” ($\chi2(2)=9.05$, p=0.011), T2 “I felt immersed in the

environment" ($\chi 2(2)=15.37$, $p<0.001$) and T3 "I felt surrounded by the environment" ($\chi 2(2)=16.06$, $p<0.001$).

Wilcoxon Signed-Rank post-hoc analysis for social presence showed significant differences in rankings between the motion parallax and conventional 2D perspectives ($Z=-2.22$, $p=0.026$ for S1, $Z=-1.99$, $p=0.046$ for S2) and between the motion parallax + stereoscopy and conventional 2D perspectives ($Z=-2.70$, $p=0.007$ for S1, $Z=-2.41$, $p=0.016$ for S2). However, we found no significant differences between the motion parallax and the motion parallax + stereoscopy conditions.

For the degree of telepresence, there was a significant difference between the motion parallax and conventional 2D perspectives ($Z=-2.32$, $p=0.020$ for T1, $Z=-2.37$, $p=0.018$ for T2), and between the motion parallax + stereoscopy condition ($Z=-2.65$, $p=0.008$ for T1, $Z=-2.99$, $p=0.003$ for T2) and the conventional 2D condition. However, there were no significant differences between motion parallax and motion parallax + stereoscopy conditions.

5.1.3 Experiment 2: Effects of Perspective Cues on Communication of 3D Body Postural Cues

In the second experiment, we examined whether support for a 360° life-size stereoscopic view with motion parallax improved the ability to convey the body pose of a remote person on the TeleHuman.

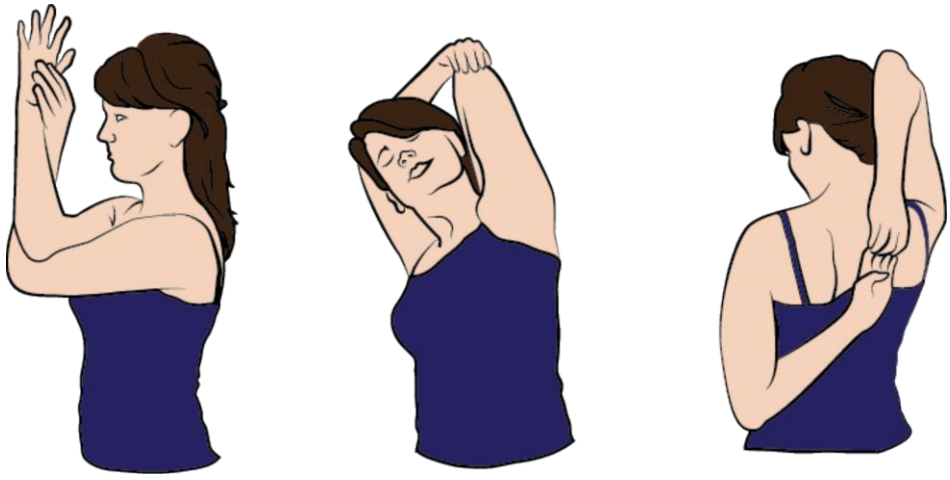

Figure 5.2. Sample Yoga stances used in Experiment 2.

5.1.3.1 Task

A remote instructor, displayed on the TeleHuman, first positioned herself in one of three predetermined yoga poses (see Figure 5.2), one per condition. At that point, the main participant ("coach") instructed a co-located partner ("poser") to reproduce the pose as accurately as possible, within a 3 minute time limit. The reason for using a poser, rather than having the coach assume the pose him or herself is that this allowed the coach to walk freely around the display, as well as around the poser. Participants were asked to walk around the TeleHuman to examine the pose, and around the poser to examine the result, in all conditions. Note that while participants were allowed to ask the instructor to rotate in order to view her back during the conventional 2D conditions, none did, as this would have interfered with her ability to perform the pose.

5.1.3.2 Experiment Design

We used a within-subject experiment design to evaluate the effects of the perspective factor only, as per the first experiment (see Figure 5.1).

5.1.3.3 Setup and Procedure

The coach and the poser were co-located in the same room as the TeleHuman system; but only the coach could see the TeleHuman system. The instructor was in a separate room, and displayed using a live 3D 360 degree video model on the TeleHuman system. We used an asymmetrical version of the system that allowed for full 360 degree motion parallax, in which the coach could see and hear the instructor as represented by the TeleHuman, but the instructor could not see the coach. The instructor was not allowed to interfere with the directions of the coach to the poser. Once the coach was satisfied with the poser's posture, the instructor would go to the poser's room to evaluate the poser's stance, while the coach filled out a questionnaire.

We used pairs of participants, unfamiliar with yoga, alternating as coach and poser. To alleviate learning effects, a different yoga pose was used for every condition between pairs of participants, for a total of six yoga poses. All yoga poses, preselected by the yoga instructor, were of an intermediate level of difficulty, and focused on upper body positioning (Figure 5.2). All poses had limb elements positioned on the back, front and sides of the instructor. The choice of yoga pose was randomly assigned to each coach and condition, and no feedback was provided by the instructor to the poser about the quality of any poses. The three visual perspective conditions were counter-balanced for each coach. The poser was never instructed on the perspective level at hand.

5.1.3.4 Participants

Eleven of the fourteen participants from the first experiment took part in the second experiment. They were paid a further $15 for their participation.

5.1.3.5 Measures

The instructor evaluated the similarity between her pose and that of the poser on a scale from 0 to 10 (10 meaning perfectly identical). In this process, she took into account limb angles and orientations, as well as overall posture. After each condition, coaches completed the same questionnaire administered in the first experiment, which evaluated the degree of telepresence and social presence experienced.

Perspective	2D	Motion Parallax	Motion Parallax + Stereoscopy
Similarity Score	4.5 (0.71)	5.5 (0.79)	7.1 (0.55)

Table 5.3. Mean pose similarity score and standard error (s.e.) on a scale from 0 to 10 by yoga instructor, per condition.

5.1.3.6 Results

We used a within-subjects ANOVA to evaluate differences between conditions, at an alpha level of .05.

5.1.3.7 Posture Similarity Scores

Table 5.3 shows the mean pose similarity score and standard error for each perspective condition. Results show that posture similarity scores were significantly different between perspective conditions ($F(2,20)=4.224$, $p=0.03$). Post-hoc tests using Bonferroni correction show that scores in the motion parallax + stereoscopy condition were significantly different from scores in the conventional 2D condition ($p=0.04$).

Statements	2D	Motion Parallax	Motion Parallax + Stereoscopy
It was as if I was facing the partner in the same room. (S1)	4.82 (1.1)	2.91 (1.1)	3.00 (1.3)
My partner seemed a real person. (S2)	4.36 (1.5)	2.82 (0.9)	2.82 (1.0)
I could get to know someone that I only met through this system. (S3)	4.55 (1.4)	3.18 (1.2)	3.45 (1.0)
I felt immersed in the environment. (T1)	4.45 (1.8)	2.82 (1.6)	3.09 (1.4)
I felt surrounded by the environment .(T2)	5.18 (1.5)	3.55 (1.6)	3.45 (1.4)
The experience was involving. (T3)	3.64 (1.4)	2.00 (0.6)	2.27 (0.8)

Table 5.4. Mean agreement and standard errors (s.e.) with social presence and telepresence statements. Lower scores indicate stronger agreement.

5.1.3.7.1 Questionnaire

Table 5.4 summarizes the mean scores for each question, per perspective condition. A Friedman test indicated that there were significant differences between perspective conditions for all social presence ratings (S1, same room $\chi2(2)=16.06$, $p=0.001$), (S2, real person $\chi2(2)=12.87$, $p=0.002$), and (S3 acquaintance $\chi2(2)=11.29$, $p=0.004$). Differences between perspective conditions were also significant for all telepresence ratings (T1, immersion $\chi2(2)=8.63$, $p=0.013$), (T2 surrounding $\chi2(2)=12.65$, $p=0.002$), and (T3, involvement $\chi2(2)=14.4$, $p=0.001$).

Wilcoxon Signed-Rank post-hoc analysis for social presence and telepresence ratings showed significant differences in rankings between the motion parallax and conventional 2D conditions ($Z=-2.83$, $p=0.005$ for S1, $Z=-2.54$, $p=0.011$ for S2, $Z=-2.55$, $p=0.011$ for S3, $Z=-2.85$, $p=0.004$ for T1, $Z=-2.54$, $p=0.011$ for T2, $Z=-2.55$, $p=0.011$ for T3) and between the motion parallax + stereoscopy and conventional 2D conditions ($Z=-2.69$, $p=0.007$ for S1, $Z=-2.55$, $p=0.011$ for S2, $Z=-2.36$, $p=0.018$ for S3, $Z=-2.54$, $p=0.011$ for T1, $Z=-2.06$, $p=0.040$ for T2, $Z=-2.56$, $p=0.011$

for T3). However, there were no significant differences between motion parallax and the motion parallax + stereoscopy conditions.

5.2 Discussion

We now present a discussion of results from our two experiments.

5.2.1 Effects of 3D Perspective on Pointing Cue Assessment

Results from our first experiment confirmed a strong effect of perspective on the accuracy of assessment of remote pointing cues. Motion parallax + stereoscopy significantly increased the accuracy of angular judgment over traditional 2D conditions in cases where participants were stationary. As expected, motion parallax alone, in this situation, was limited, and thus, the addition of stereoscopy was important. When participants were allowed to move, motion parallax was shown to provide the dominant effect, with participants achieving higher accuracy in angular judgment of remote pointing cues as compared to 2D conditions. In this case, stereoscopy appeared to provide little additional benefit. Note that the type of pointing cue: *gaze, hand only,* or *gaze + hand*, had no significant effect on accuracy measures.

Qualitative measures support the above analysis. Social presence rankings were significantly higher in conditions where motion parallax cues were supported, with no significant additional effect for motion parallax augmented by stereoscopy. As for the degree of telepresence or immersion, the combined effect of motion parallax and stereoscopy was critical for obtaining significant differences from 2D conditions.

Stereoscopy therefore appears to be beneficial for judgment of pointing angle when motion parallax cannot be exploited. However, this comes at the cost of preventing reciprocal gaze awareness if shutter glasses are deployed. Motion parallax, even in the absence of a stereoscopic display, may, however, suffice for preservation of social presence or pointing cues.

5.2.2 Effects of 3D Perspective on Body Pose Assessment

Results for our second experiment, in which we evaluated the effects of perspective cues on preservation of postural cues, were in line with those from Experiment 1. The presence of *motion parallax + stereoscopy* cues significantly increased the accuracy of pose scores over conventional 2D conditions. These results suggest that both motion parallax and stereoscopy needed to be present in order to judge and convey poses accurately. Surprisingly, the presence of motion parallax cues alone only marginally improved scores. This was likely due to the fact that while motion parallax allowed users to see the sides and back of poses, stereoscopy helped improve their judgment of the relative angles of the limbs.

Qualitative measures indicate little additional effect of the presence of stereoscopic cues. Social presence rankings were significantly higher in conditions where *motion parallax* or *motion parallax + stereoscopy* were supported. As for the degree of telepresence, rankings were significantly higher in cases where *motion parallax* or *motion parallax + stereoscopy* were supported. However, there appeared to be little additional effect of the presence of stereoscopic cues over *motion parallax* only. While the presence of stereoscopy did not significantly affect qualitative measures, we can conclude that in this task both motion parallax and stereoscopy were required.

5.3 Limitations

Our first study was limited by the fact that the TeleHuman was a static 3D image, and communication was not reciprocal. Although this permitted us to evaluate the effect of stereoscopy on pointing cue assessment, it necessitated an artificial communication condition in which the shutter glasses had no detrimental effect on perception of eye contact. There is an obvious tradeoff between supporting eye contact between interlocutors and presentation of a stereoscopic display requiring the use of shutter glasses. However, other display technologies, such as autostereoscopic and volumetric displays do support glasses-free stereo viewing. We hope

to conduct future experiments to evaluate the added benefit that such technologies might offer in terms of eye contact perception with TeleHuman. Note that participants in our study did not ask the instructor to rotate in the 2D condition. There may be cases in which such rotation would provide adequate information to complete a 3D pose task. To avoid introducing confounding factors, we did not specifically compare results with traditional 2D flat display conditions. However, we believe that the results of our 2D conditions would generalize to such conditions.

Chapter 6

Conclusions & Future Work

6.1 Conclusions

In this thesis we examined 360 degree curvilinear displays and how they can be used to present 3D information. We presented two systems, a spherical display prototype and a cylindrical display. We conducted two experiments to examine how our final cylindrical display prototype, using stereoscopy and 360 degree motion parallax, might aid in the preservation of basic body orientation cues and in pose estimation tasks in a telepresence application.

Our spherical display prototype, referred to as SnowGlobe, allowed users to view and manipulate 3D models as if they were nested inside the spherical display. This prototype made use of the natural affordance of walking around the display in order to view occluded information to allow for 360 degree motion parallax around the display. Additionally, we made use of a spherical display's bounded physical volume which allowed us to treat it as a 3D display volume rather than strictly as a 2D display surface. Based on these properties, a combination of 360 degree motion parallax and stereoscopy was used to create the sensation that 3D information was contained within the display surface. Additionally, our gesture set allowed users to rotate and scale objects using in air gestures. Initial experiences suggest this display gives the illusion that the object on the display is suspended within the 3D space of the globe.

Using the hardware and software concepts from our SnowGlobe system, we developed the TeleHuman system, a cylindrical display portal for life-size 3D human telepresence. The system transmits telepresence by conveying 3D video images of remote interlocutors in a way that preserves 360º motion parallax around the display, as well as stereoscopy. We empirically evaluated the effect of perspective on the user's accuracy in judging gaze, pointing direction, and body pose of a remote partner using an asymmetrical version of the system. Results for pointing

directional cues suggest that the presence of stereoscopy is important in cases where the user remains relatively stationary. However, when users move their perspective significantly, motion parallax provides a dominant effect in improving the accuracy with which users were able to estimate the angle of pointing cues. As for pose estimation, the presence of both 360º motion parallax cues and stereoscopic cues appeared necessary to significantly increase accuracy. Both motion parallax and stereoscopy appear important in providing users with a sense of social presence and telepresence. We conclude that we recommend inclusion of both motion parallax and stereoscopic cues in video conferencing systems that support the kind of tasks used in our evaluation, with the caveat that tools such as shutter glasses, which obstruct views of the remote participants eyes, are most likely not recommendable for bi-directional communication systems.

6.2 Future Work

Future work is focused on the TeleHuman system, which has potential applications in a number of areas where 2D displays may limit a user's viewpoints. One example is in remote sports instruction. As the body pose experiment demonstrates, examination of the mechanics of limb movement may benefit from the ability to review movement and posture from any angle. For example, this may be helpful in teaching golfers to improve their swing. Applications also exist in telemedicine and remote medical instruction, for which the benefits of arbitrary view control were demonstrated previously in the context of surgical training [39]. TeleHuman could similarly offer doctors the ability to examine remote patients from any angle, but at full scale. This may be particularly beneficial for orthopedic or postural conditions, where the patient cannot reorient herself for a side view. Additionally, we have examined the feasibility of using the TeleHuman system to explore detailed 3D models of human anatomy. This has potential educational applications as a user can view a detailed life-size human anatomy and using gestures and speech focus on specific anatomical systems or body parts.

6.2.1 Support for Multiparty Videoconferencing

In the near future, we hope to leverage TeleHuman for *multiparty* teleconferencing scenarios. To support such experimentation, we will be replacing the current TCP communication layer with a UDP-based alternative, suitable for low-latency interaction over larger distances. Support of a teleconference with *n* users requires *n2-n* setups and, barring multicast support, a similar number of data streams. This entails significant bandwidth requirements for transmission of 3D video models. However, our design allows for such scaling without modifications to the TeleHuman hardware.

References

1. Actuality's Perspecta display, http://www.actuality- medical.com/

2. Alexander, O., Lambeth, W., and Debevec, P. Creating a Photoreal Digital Actor : The Digital Emily Project Previous Efforts at Photoreal Digital Humans. Proc. SIGGRAPH, (2009), 1-15.

3. Argyle, M. and Ingham, R. Gaze, mutual gaze and proximity. *Semiotica 6*, 1 (1972), 32-49.

4. Balakrishnan, R., Fitzmaurice, G., & Kurtenbach, G. User interfaces for Volumetric Displays. *IEEE Computer 34,* 3 (2001), 37-45.

5. Benko, H., Wilson, A. D., and Balakrishnan, R. Sphere: multi-touch interactions on a spherical display. In Proc. UIST, (2008), 77-86.

6. Benko, H. Beyond flat surface computing: challenges of depth-aware and curved interfaces. In Proc. MM, (2009), 935-944.

7. Beyer, G., Alt, F., Müller, J., Schmidt, A., Isakovic, K., Klose, S., Schiewe, M., and Haulsen, I. Audience behavior around large interactive cylindrical screens. In Proc. CHI, (2011), 1021-1030.

8. Buxton, B. Telepresence: integrating shared task and person spaces. *Proc. GI*, (1992), 123-129.

9. Böcker, M. and Mühlbach, L. Communicative Presence in Videocommunications. *Proc. Human Factors and Ergonomics Society*, (1993), 249-253.

10. Böcker, M., Blohm, W., and Mühlbach, L. Anthropometric data on horizontal head movements in videocommunications. *Proc. CHI*, (1996), 95-96.

11. Böcker, M., Rundel, D., and Mühlbach, L. On the Reproduction of Motion Parallax in Videocommunications. *Proc. HFES*, (1995), 198-20.

12. Clark, H. Using Language. Cambridge University Press, Cambridge, MA, USA, 1996.

13. Companje, R., van Dijk, N., Hogenbirk, H., and Mast, D. Globe4D: time-traveling with an interactive four-dimensional globe. In Proc. SIGGRAPH. (2007), 26.

14. Cooperstock, J.R. Multimodal Telepresence Systems: Supporting demanding collaborative human activities. *IEEE Signal Processing Magazine, Special Issue on Immersive Communications 28*, 1 (2011), 77-86.

15. Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., and Sasse, M.A. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. *Proc. CHI*, 5 (2003), 529.

16. Gross, M., Würmlin, S., Naef, M., et al. blue-c: a spatially immersive display and 3D video portal for telepresence. *ACM Transactions on Graphics 22*, 3 (2003), 819-827.

17. Grossman, T., Wigdor, D., and Balakrishnan, R. Multi-finger gestural interaction with 3d volumetric displays. In Proc. UIST, (2004), 61-70.

18. Harrison, C. and Hudson, S. Pseudo-3D Video Conferencing with a Generic Webcam. *Proc. IEEE International Symposium on Multimedia*, (2008), 236-241.

19. Holman, D. and Vertegaal, R. Organic user interfaces: designing computers in any way, shape, or form. Communication of ACM 51, 6 (2008), 48-55.

20. Holman, D., Vertegaal, R., Altosaar, M. Paper Windows: Interaction Techniques for Digital Paper. In *Proc. CHI 2005,* ACM Press (2005), 591-599.

21. Jones, A., Lang, M., Fyffe, G., et al. Achieving eye contact in a one-to-many 3D video teleconferencing system. *ACM Transactions on Graphics 28*, 3 (2009), 64:3-64:8.

22. Jouppi, N.P., Iyer, S., Thomas, S., and Slayden, A. BiReality: mutually-immersive telepresence. *Proc. MM*, (2004), 860-867.

23. Kendon, A. Some functions of gaze direction in social interaction. *Acta Psychologica 26*, 44 (1967), 22-63.

24. Lightspeed Design. DepthQ Stereoscopic Projector. http://www.depthq.com.

25. Lin, J., Chen, Y., Ko, J., Kao, H., Chen, W., Tsai, T., Hsu, S., and Hung, Y. i-m-Tube: an interactive multi-resolution tubular display. In Proc. MM, (2009), 253-260.

26. Lincoln, P., Nashel, A., Ilie, A., Towles, H., Welch, G., and Fuchs, H. Multi-view lenticular display for group teleconferencing. *Proc IMMERSCOM*, (2009).

27. Litefast. http://www.litefast-display.com/

28. Matusik, W. and Pfister, H. 3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. *ACM Transactions on Graphics*, (2004), 814–824.

29. Microsoft Kinect. http://www.xbox.com/kinect

30. Mori, M. The Uncanny Valley. *Energy 7*, 4 (1970), 33-35.

31. Morikawa, O. and Maesako, T. HyperMirror: toward pleasant-to-use video mediated communication system. *Proc. CSCW*, (1998), 149-158.

32. NVidia. nVidia 3D Vision Kit. http://www.nvidia.com/object/3d-vision-main.html

33. Naef, M., Lamboray, E., Staadt, O., and Gross, M. The blue-c distributed scene graph. *Proc. IPT/EGVE Workshop*, (2003), 125-133.

34. Negroponte, N. *Being Digital*. Vintage Books, New York, NY, USA, 1995.

35. Nishino, H., Utsumiya, K. and Korida,K. 3D Object Modeling Using Spatial and Pictographic Gestures. In *Proc. VRST 1998,* ACM Press (1998), 51-58.

36. Nguyen, D. and Canny, J. MultiView: spatially faithful group video conferencing. *Proc. CHI*, (2005), 799–808.

37. Nowak, K.L. and Biocca, F. The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence Teleoperators and Virtual Environments 12*, 5 (2003), 481-494.

38. Okada, K.-I., Maeda, F., Ichikawaa, Y., and Matsushita, Y. Multiparty videoconferencing at virtual social distance: MAJIC design. *Proc. CSCW*, (1994), 385-393.

39. Olmos, A., Lachapelle, K., and Cooperstock, J.R. Multiple angle viewer for remote medical training. *Proc. International Workshop on Multimedia Technologies for Distance Leaning*, (2010), 19–24.

40. OpenCV. http://opencv.willowgarage.com/

41. OpenNI. http://openni.org/

42. Rosenthal, A. Two-way Television Communication Unit. (1947).

43. Sellen, A., Buxton, B., and Arnott, J. Using spatial cues to improve videoconferencing. *Proc. CHI*, (1992), 651-652.

44. Sheng, J., Balakrishnan, R. and Singh, K. An Interface for Virtual 3D Sculpting via Physical Proxy. In *Proc. Graphite 2006,* ACM Press (2006), 213-220.

45. SONY 360 Stereoscopic Display. http://www.sony.co.jp/SonyInfo/News/Press/200910/09-123/

46. Stavness, I., Lam, B. and Fels, S. pCubee: a perspective-corrected handheld cubic display. In *Proc. CHI 2010.* ACM Press (2010), 1381-1390.

47. Tang, J.C. and Minneman, S. VideoWhiteboard: video shadows to support remote collaboration. *Proc. CHI*, (1991), 315-322.

48. Ware, C., Arthur, K., and Booth, K.S. Fish tank virtual reality. *Proc. CHI*, (1993), 37-42.

49. Wright, M. and Freed, A. Open Sound Control: A New Protocol for Communicating with Sound Synthesizers. *Proc. International Computer Music Conference*, (1997), 101-104.

50. Vertegaal, R. and Ding, Y. Explaining Effects of Eye Gaze on Mediated Group Conversations : Amount or Synchronization ? *Proc. CSCW*, (2002), 278-285.

51. Vertegaal, R., Slagter, R., Der Veer, G. Van, and Nijholt, A. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. *Proc. CHI*, (2001), 301-308.

52. Vertegaal, R., Weevers, I., Sohn, C., and Cheung, C. GAZE-2: conveying eye contact in group video conferencing using eye-controlled camera direction. *Proc. CHI*, (2003), 521–528.

53. Vertegaal, R. The GAZE groupware system: mediating joint attention in multiparty communication and collaboration. *Proc. CHI*, (1999), 294-301.

54. Yamashita, N. and Ishida, T. Effects of Machine Translation on Collaborative Work. In Proc. CSCW(2006), 515–524.

55. Zhang, C., Yin, Z., and Florencio, D. Improving depth perception with motion parallax and its application in teleconferencing. *Proc. IEEE International Workshop on Multimedia Signal Processing*, (2009), 1-6.

Appendix A

Questionnaires

Effects of 3D Perspective on Gaze and Pose Estimation with a Life-size Cylindrical Telepresence Pod - Experiment 1: Effects of 3D Perspective on Gaze and Pointing Direction Estimates

1. I was able to assess my partner's reactions.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

2. This was like a face-to-face meeting.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

3. I was in the same room with my partner.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

4. My partner seemed "real."

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

5. I could use this system of interaction for a meeting in which I wanted to persuade others of something.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

6. I could get to know someone that I only met through this system.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

7. My partner was making eye contact with me.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

8. I was comfortable talking to my partner.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

9. My partner's head and body orientation were apparent.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

10. Please feel to share with us any other comments or thoughts you have about your experience with this system.

Effects of 3D Perspective on Gaze and Pose Estimation with a Life-size Cylindrical Telepresence Pod - Experiment 2: Effects of Perspective Cues on Communication of 3D Body Postural Cues

1. I was able to assess yoga instructor's reactions.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

2. I was able to instruct the partner efficiently.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

3. This was like as if I were facing the yoga instructor in the same room.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

4. The yoga instructor seemed a real person.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

5. I could get to know someone that I only met through this system.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

6. The experience was involving.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

7. I felt immersed in the environment I saw.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

8. I felt surrounded by the environment I saw.

Strongly agree [ ] Agree [ ] Agree somewhat [ ] Undecided [ ] Disagree somewhat [ ] Disagree [ ] Strongly disagree [ ]

9. Please feel to share with us any other comments or thoughts you have about your experience with this system.