Int. J. Human-Computer Studies 86 (2016) 138–148

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Effects of 3D perspective on head gaze estimation with a multiview autostereoscopic display ☆

Ye Pan *, Anthony Steed

Virtual Environments & Computer Graphics Group, Department of Computer Science, University College London, UK

ARTICLE INFO

Article history:
Received 21 January 2015
Received in revised form
17 July 2015
Accepted 21 October 2015
Communicated by B. Lok
Available online 30 October 2015

Keywords:
Gaze
Telecommunication
Autostereoscopic display
Random hole display
Avatars

ABSTRACT

Head gaze, or the orientation of the head, is a very important attentional cue in face to face conversation. Some subtleties of the gaze can be lost in common teleconferencing systems, because a single perspective warps spatial characteristics. A recent *random hole display* is a potentially interesting display for group conversation, as it allows multiple stereo viewers in arbitrary locations, without the restriction of conventional autostereoscopic displays on viewing positions. We represented a remote person as an avatar on a random hole display. We evaluated this system by measuring the ability of multiple observers with different horizontal and vertical viewing angles to accurately and simultaneously judge which targets the avatar is gazing at. We compared three perspective conditions: a conventional 2D view, a monoscopic perspective-correct view, and a stereoscopic perspective-correct views. In the latter two conditions, the random hole display shows three and six views simultaneously. Although the random hole display does not provide high quality view, because it has to distribute display pixels among multiple viewers, the different views are easily distinguished. Results suggest the combined presence of perspective-correct and stereoscopic cues significantly improved the effectiveness with which observers were able to assess the avatar's head gaze direction. This motivates the need for stereo in future multiview displays.

1. Introduction

Gaze has several roles in group communication, including facilitating turn-taking, conveying cognitive activity, and expressing involvement etc (Argyle and Cook). However, standard telepresence systems often distort or destroy gaze cues (see e.g., Nguyen et al., 2005; Vertegaal et al., 2003; Schreere et al., 2005), because the single perspective view of the camera does not preserve the spatial characteristics of the face to face situation. In particular, in group conferencing, when a participant looks into the camera, everyone feels that the participant is looking toward them; when the participant looks away from the camera (for example, toward other participants in the meeting), no one sees the participant looking at them (see e.g., Roberts et al., 2013).

A variety of systems have been developed to support gaze awareness, though the majority use a multiple view 2D video system (see e.g., Nguyen et al., 2005) or a single user virtual reality system (see e.g., Roberts et al., 2009). In particular, the use of autostereoscopic display technologies could support multiple users simultaneously each with their own perspective-correct view without the need for special eyewear. However, these are usually restricted to specific optimal viewing zones. Our telepresence system uses the random hole display design (Ye et al., 2010; Nashel and Fuchs, 2009) which has a dense pattern of tiny, pseudo-randomly placed holes as an optical barrier mounted in front of a flat panel display. This allows observers anywhere in front of the display to see a different subset of the display's native pixels through the random-hole barrier. Additionally, it is technically quite simple to build and can be constructed very cheaply in comparison to holographic displays and volumetric displays.

Recently, avatar-mediated communication, where a remote person is represented by a graphical humanoid, has increased in prevalence and popularity as an emerging form of visual remote interaction. The avatar represents the presence and activities of a remote user and can be visualized using standard displays or projection surfaces in the local room with perspective-correct graphical rendering via head tracking of the local user (Roberts et al., 2009). We developed a view-dependent ray traced rendering method to represent a remote person as an avatar on the random hole display. The method allows multiple observers in arbitrary

☆This paper has recommended for acceptance by B. Lok.

* Corresponding author. Tel.: +44 07403023986.
E-mail address: y.pan@cs.ucl.ac.uk (Y. Pan).

http://dx.doi.org/10.1016/j.ijhcs.2015.10.004
1071-5819/

locations to perceive stereo images simultaneously. We investigated using the random hole display to represent a remote person for group teleconferencing. A study explores the effectiveness with which observers can discriminate the avatar's head orientation when the avatar's eyes are centered in the head, because head gaze is a good indicator of focus of attention in human computer interaction applications (Stiefelhagen and Zhu, 2002; Oyekoya et al., 2012). We compared three different conditions: *conventional 2D*, *perspective-correct*, and *perspective-correct & stereoscopy* across nine varying viewing angles. Results show that the presence of both perspective-correct and stereoscopic cues significantly improved the accuracy with which participants were able to assess the avatar's head gaze in both horizontal and vertical directions. This demonstration motivates the further study of novel display configurations and suggest parameters for the design of teleconferencing systems.

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

2. Related work

2.1. Autostereoscopic displays for teleconferencing applications

Depth perception, or 3D perception, can add a lot to the feeling of immersiveness in many applications such as 3D teleconferencing. However, a conventional stereo display hardware would require the use of 3D glasses, which are cumbersome and make it difficult to support eye contact perception in two way teleconferencing. Autostereoscopic displays, which present a 3D image to a viewer without the need for glasses or other encumbering viewing aids, can be used to improve the teleconferencing experience.

In particular, parallax displays based on barriers or lenticular lens sheets provide a relatively simple and inexpensive solution for autostereoscopy. A parallax barrier is a flat film composed of transparent and opaque regions, while a lenticular screen is a sheet of cylindrical lenses. Parallax barrier displays occlude certain parts of the screen from one eye while allowing another eye to see them. Systems such as Perlin et al.'s autostereoscopic display (Perlin et al., 2000), Varrier (Sandin et al., 2005), and Dynallax (Peterka et al., 2008) demonstrate this concept. Lenticular displays include Kooima et al.'s work (Kooima et al., 2010) and the MERL display (Matusik and Pfister). Additionally, Kim et al. proposed another approach enabling concurrent dual views on twisted-nematic LCD screens, by exploiting a technical limitation of these LCD screens (Kim et al., 2012).

However, neither autostereoscopic displays nor conventional stereo displays support both vertical motion parallax and multiple arbitrary views. Firstly, most conventional autostereo displays do not offer multiuser motion parallax (multiple distinct views) along the vertical direction. Integral imaging displays using a 2D array of lenslets could generate fullparallax autostereo images, but these have a limited viewing angle and low resolution. Therefore, it would be difficult to provide perspective correct views for observers with different heights. With regular multi-user autostereoscopic displays, untracked viewers must remain in certain viewing areas or they will see incorrect imagery or the same imagery as other viewers. In autostereoscopic display systems with user tracking, multiple viewers are usually not supported because individual display pixels will be seen from multiple views.

Recently, an interesting approach to build multi-view displays is based on viewing the screen through a hole-mask that is placed at a certain distance from the data to serve as a barrier that mediates the view for different users. Kitamura et al.'s Illusion Hole uses a display mask which has a hole in its center (Kitamura et al., 2001). Naschel et al.'s random hole display prototype extends their approach by using a randomized hole distribution parallax barrier (Nashel and Fuchs, 2009). The random hole display design eliminates the repeating zones found in regular barrier and lenticular autostereoscopic displays, enabling multiple simultaneous viewers in arbitrary locations (Nashel and Fuchs, 2009). Ye et al. demonstrate a full multi-user multi-view system using this concept with their Tabletop Autostereoscopic Display (Ye et al., 2010). Instead of using a static hole-mask, Karnik et al.'s MUSTARD uses a dynamic random hole mask allowing coverage of the entire screen by constantly changing the hole-mask from frame to frame (Karnik et al., 2012).

While autostereoscopic and multiple arbitrary views capabilities of a random hole display are novel, the effectiveness of using the random hole display for telepresence is not yet clear. We run an experiment to demonstrate that the random hole display can convey head gaze relatively accurately, particularly for group conferencing.

2.2. Gaze in telepresence systems

Gaze, attention, and eye contact are important aspects of face to face conversation. They help create social cues for turn taking, establish a sense of engagement, and indicate the focus and meaning of conversation (Argyle and Cook). However, perceiving gaze direction is difficult in most teleconferencing systems and hence limits their effectiveness (Nguyen et al., 2005). Chen reported that the perception of eye contact decreases if the horizontal contact angle is greater than 1° or the vertical contact angle is greater than 5° (Chen, 2002).

2.2.1. Telepresence systems

Over the years a number of solutions have been developed to convey gaze direction during multiparty video conferencing, including MAJIC (Okada et al., 1994), Hydra (Sellen et al., 1992), GAZE-2 (Vertegaal et al., 2003), MultiView (Nguyen et al., 2005), cylindrical multiview system (Pan and Steed, 2014), 3D facial display (Nagano et al., 2013), animatronic shader lamps avatars (Lincoln et al.) and One-to-Many System (Jones et al., 2009). Also, a variety of solutions have been devised to explore the preservation of 3D depth cues and motion parallax via a single user head position tracking and the use of shutter glasses, such as, Kim et al. (2012), SphereAvatar (Oyekoya et al., 2012; Pan et al., 2014), PCubee (Stavness et al., 2010), Spheree (Ferreira et al.), 3-d live (Prince et al., 2002) and some CAVE-like environments (Roberts et al., 2009; Gross et al., 2003). However, these systems are currently developed for a single observer.

Our system allows multiple observers to see correct stereo images from arbitrary locations in front of the display.

2.2.2. Perception of head and eye gaze direction

The direction of a person's gaze is one feature that is relevant in judging objects of interest in an environment. Gibson et al. established that gaze direction may be perceived by both the direction in which the head is oriented and the eyes' position relative to the head (Gibson and Pick, 1963). Anstis et al. investigated gaze estimation influenced by three orientations of a TV screen. They found a *TV screen turn* effect such that apparent displacement of the perceived direction in the same direction as the turn of the screen and suggested that the convex curvature of the screen probably caused the TV screen turn effect (Anstis et al.). They also reported an *overestimation* effect such that when gaze was to one side of the participant, the participant judges it to be further to that side than it actually was. They suggested that this overestimation became greater as the complexity of the viewing situation increased. 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 (Oyekoya et al., 2012).

In spite of the importance of the head gaze as an attentional cue (Stiefelhagen and Zhu, 2002), there is relatively little research on the perception of its orientation. Troje and Siebeck have provided evidence for the use of a head asymmetry cue to gaze direction estimation (Troje and Siebeck, 1998). Wilson et al. reported that head orientation discrimination is based upon both cues: deviation of head shape from bilateral symmetry, and deviation of nose orientation from vertical (Wilson et al., 2000). Perception of an avatars gaze direction has also been studied in virtual environments. 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 (Steptoe et al., 2009).

In this initial study, we have followed the previous work and used static centered eyes (Oyekoya et al., 2012), so visual attention must be inferred from the direction of the head only, although the underlying system supports full eye gaze as well as facial expressions.

2.2.3. Motion parallax and stereoscopy

Although the benefits of including motion parallax and stereoscopy in the presentation of graphic interfaces have been demonstrated, systematic evaluation of the impact of these factors in the context of task performance during avatar mediated interaction, specifically in assessing head gaze, is sparse. Arthur et al. (1993) experiments tested user performance under two conditions in fish-tank virtual reality: whether or not stereoscopic display was used, and whether or not the perspective display was coupled dynamically to the positions of a users eyes. Subjects using a comparison protocol consistently preferred head coupling without stereo over stereo without head coupling. Error rates in a tree-tracing task similar to one used by Sollenberger and Milgram (1991) showed an order of magnitude improvement for head-coupled stereo over a static (nonhead-coupled) display, and the benefits gained by head coupling were more significant than those gained from stereo alone. Böcker et al. compared videoconferencing systems that provide motion parallax and stereoscopic displays and found this increased spatial presence and greater exploration of the scene (Böcker et al., 1995). Böcker et al. subsequently found that the provision of motion parallax generated larger head movements in users of video conferencing systems (Böcker et al., 1996). Kim et al. used three gaze targets (front, left, right) and found the combined presence of motion parallax and stereoscopic cues significantly improved the accuracy with which participants were able to assess gaze (Kim et al., 2012).

We further investigated the effects of reproducing perspective-correct and stereoscopic cues in telepresence in both horizontal and vertical directions.

3. System design

3.1. Hardware

Our hardware is based on the design of Ye et al.'s Tabletop Autostereoscopic Display (Ye et al., 2010). The display uses three layers to create its viewing zones. A diagram of the layers is shown in Fig. 1. The back-most layer is a single LCD display panel. The HP ZR30w 30-inch S-IPS LCD Monitor was used for two reasons. Firstly, as a parallax barrier reduces the effective resolution of the display, we selected a high-resolution (2560 × 1600) and reasonably priced LCD. Secondly, we used the S-IPS type display, because it has very large horizontal and vertical viewing angles. In contrast, the twisted-nematic (TN) panels, which are widely used for low-cost consumer-grade LCD displays, have a limited vertical viewing angle and exhibit colour inversion when viewed from below. The next layer is a Lexan™ polycarbonate sheet, which forms the separating layer. The thickness of the sheet is 6 mm (cost approximately $40). The Lexan™ polycarbonate sheet's refractive index is slightly above 1.5 and similar to the index of the LCD panel's built-in transparent cover. The last layer is the random hole mask that was printed on a thin polyester film at 1200 dpi (cost approximately $20).

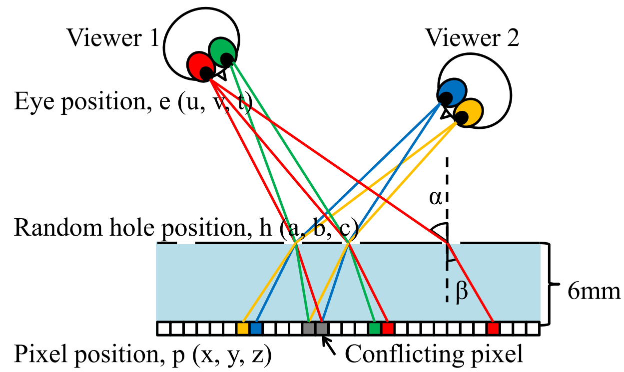

Fig. 1. A top down diagram of the random hole display showing two viewing positions.

3.2. Software

We developed a view-dependent ray traced rendering method to represent a remote person as a commercially available Rocketbox® avatar on the display. We used the NVIDIA® OptiX ray tracing engine. Instead of tracing a ray from a viewpoint through each pixel in a virtual screen, we trace a ray from each eye through each hole in the mask (see Fig. 1). We consider the refractive effects when the light passes through the barrier film, the Lexan™ polycarbonate sheet separating layer, and the LCD panel's own protective cover. If multiple eyes see the same pixels behind the barrier, then a conflict occurs. We choose one of the conflict views randomly. We then calculate the colour of the object visible on a certain area of the screen through each hole for each eye. This rendering algorithm can support perspective correct images for multiple observers. It also could be extended to other display systems that have a three dimensional display surface (Pan et al., 2014) or a screen with different refractive indices.

Also, by using the pseudo-random Poisson distribution of the hole pattern (Dunbar and Humphreys, 2006), visual conflicts between views are distributed across the viewing area as high frequency noise. The high frequency noise is typical of these displays; however, users can clearly identify images and objects. Note that although our software can update the viewer locations in realtime, for the purposes of our controlled experiment, we used static viewing positions.

Fig. 2 shows the source image that combines the six views actually displayed on the LCD panel. It allows three observers in front of the display to see perspective-correct stereo images on subset of the display's native pixels through the random-hole screen. Fig. 3 shows photographs from six viewing positions, corresponding to the three stereo views of the three observers.

Fig. 2. Source image of six simultaneous views.

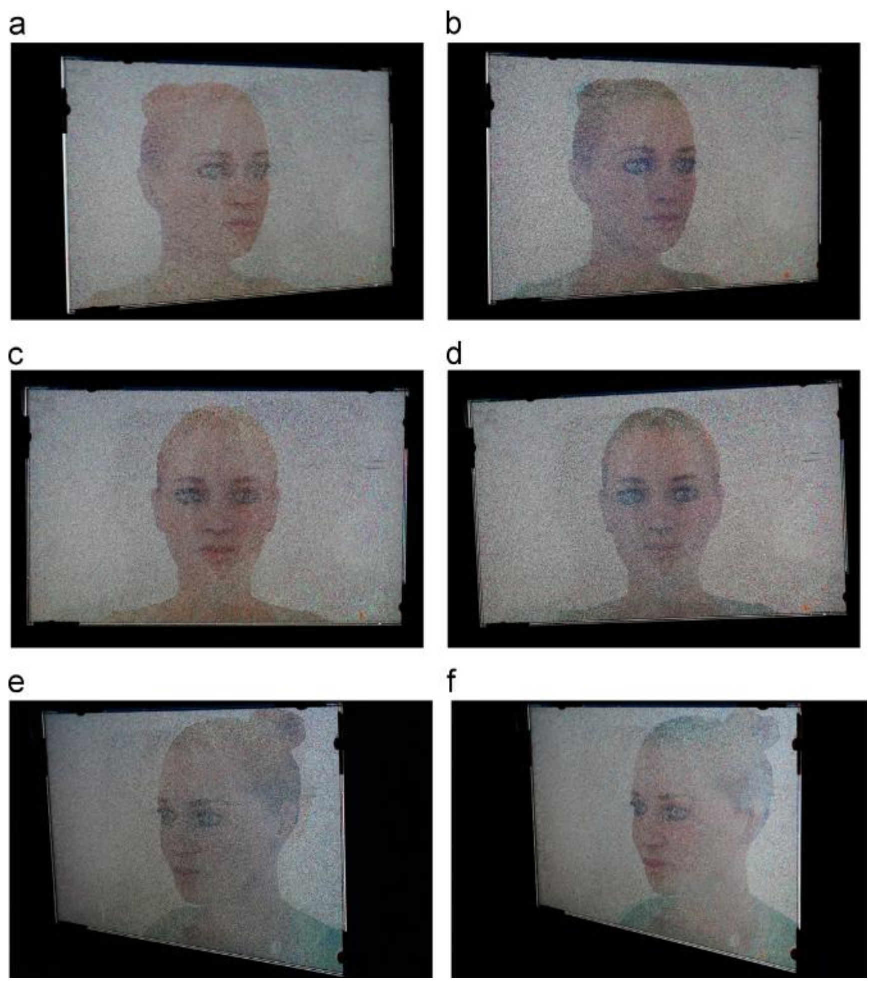

Fig. 3. Photos of six simultaneous views of the random hole display at 170 cm from the display. (a) Left eye (b) Right eye (c) Left eye (d) Right eye (e) Left eye (f) Right eye.

4. Experiment

The purpose of the study was to investigate whether our system can better represent the remote person's head gaze. The experiment was designed for three participants simultaneously as a very practical demonstration of the gaze preserving capability for group teleconferencing. We measured the effectiveness of the display by measuring the ability of multiple observers to accurately judge which target the avatar was gazing at.

We compared 3 perspective conditions. For the conventional *2D condition*, the conventional display was shown from the perspective of a front facing camera, centered on the avatar's head (see Fig. 5(b)). This condition mimicked the commonly found Mona Lisa gaze effect. For the *perspective-correct condition*, the random hole display was displayed with perspective correct monoscopic view based on the location of the observer relative to the display. For the *perspective-correct & stereoscopy condition*, the random hole display was displayed with correct perspective for each of observers' eyes, that provided them with a fully stereoscopic image, giving the impression that the avatar's head was inside the display. The apparent size of avatar remained the same in all conditions.

We explored 9 observers' viewing angles, including three horizontal viewing angles $(-30^\circ, 0^\circ \ \& \ +45^\circ)$ and three vertical viewing angles $(-10^\circ, 0^\circ \ \& \ +20^\circ)$. Previous research (e.g. Pan and Steed, 2014; Oyekoya et al., 2012; Nguyen et al., 2005) has demonstrated that the ability to assess gaze is symmetric with respect to viewing angle horizontally, that is, there is no bias in assessing angles to the left or right. In this experiment, we used asymmetric viewing angles to investigate the influence of various viewing angles for three different perspective conditions. The two extreme vertical viewing positions are where the observer sat right on the floor (-10°) and the observer stood up straight (20°).

4.1. Hypotheses

4.1.1. Hypothesis 1a

Horizontally, we expect that the participants can better identify correct targets in the perspective-correct & stereoscopy condition, followed by the perspective-correct condition and then the conventional 2D condition.

4.1.2. Hypothesis 1b

Vertically, we expect that the participants can better identify correct targets in the perspective-correct & stereoscopy condition, followed by the perspective-correct condition and then the conventional 2D condition.

4.1.3. Hypothesis 2a

Horizontally, we expect the ability of observer to perceive targets at all horizontal viewing angles will remain stable in both the perspective-correct & stereoscopy condition and the perspective-correct condition. However, for the conventional 2D condition, the ability of observer to perceive targets will decrease as the viewing angle diverges horizontally from the central viewing angle.

4.1.4. Hypothesis 2b

Vertically, we expect the ability of observer to perceive targets at all vertical viewing angles will remain stable in both the perspective-correct & stereoscopy condition and the perspective-correct condition. However, for the conventional 2D condition, the ability of observer to perceive targets will decrease as the viewing angle diverges vertically from the central viewing angle.

4.2. Method

4.2.1. Participants

27 participants, students and staff at University College London, were recruited to take part as observers in our user study. All participants had normal or corrected to normal eye sight.

4.2.2. Design

The experiment had a 3 perspective conditions $\times$ 3 horizontal viewing angle $\times$ 3 vertical viewing angle $\times$ 35 target positions mixed design, with a within-subjects design for target positions but a between-subject design regarding perspective conditions, horizontal viewing angles and vertical viewing angles.

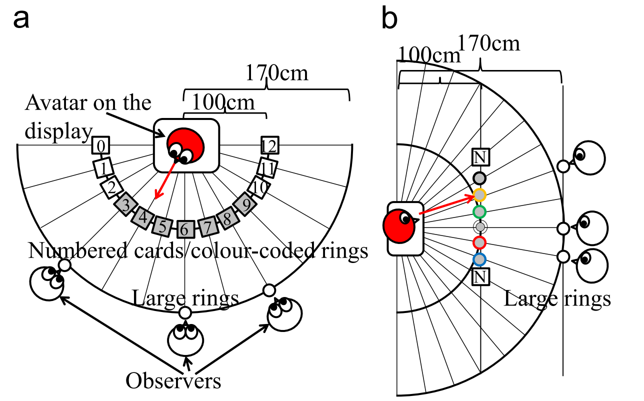

Fig. 4. Schematic layout of experiment setup. Note that the gray area covered actual target positions. (a) Horizontal view (b) Vertical view.

4.2.3. Apparatus and materials

Figs. 4 and 5 show the layout of the experiment room. We arranged small rings as potential target positions. The rings were 1.5 cm in diameter, and were placed in a 13×8 grid. On the top and bottom rows were 13 numbered cards (0–12) in a semicircle of radius 100 cm at every 15°. Each column consists of two cards and 6 rings hung from the ceiling with thin thread 10° apart from one another. To improve discriminability, the rings were colour-coded in the following order: black, yellow, green, white, red, and blue. We further arranged 9 large rings to control participants' eye position for 9 viewing angles by asking them to view avatar though one of large rings. The viewing distance from participant to avatar position was approximately 170 cm.

In the experiment, we created 35 visual stimuli by rotating avatar's head to look at 7×5 target positions out of 13×8 potential target positions in a prearranged random order (Table 1). Note that the grid of potential target positions was larger than the area of actual target positions, enabling the quantitative investigation of bias in observer perceived target positions. A new target position was given every 10 s. Each target position was gazed at only once, amounting to 35 visual stimuli. The most extreme visual stimuli to the outer-most target positions horizontally and vertically were 45° and 20°, respectively. We ensured the avatar's visual stimulus lined up exactly with the centre of corresponding rings.

4.2.4. Procedure

Nine groups of participants were used for testing, and each group had three participants. Each group experienced one of three different perspective conditions with one of three vertical viewing angles. Each observer sat at one of the three horizontal viewing angles (see Fig. 5). Each observer was given a sheet of paper with an empty grid of 35 squares. The avatar reoriented to a new target every 10 s. At the same time an audio prompt to the observers instructed them that this was a new target position. Then, observers would judge which target the avatar was gazing at and then write this in the relevant grid square. There is no discussion allowed during the task, and the participant cannot see others judgments. The experiment took about 6 min. Participants received chocolates as compensation.

4.3. Result

4.3.1. Horizontal error

The primary measurement in our results was the horizontal error in perceiving targets. Any given stimulus i can be defined by a horizontal position (i_h) and a vertical position (i_v). We defined horizontal error of each target (ϵ_{i_h}) to be the absolute value of a difference between the horizontal position of observer perceived

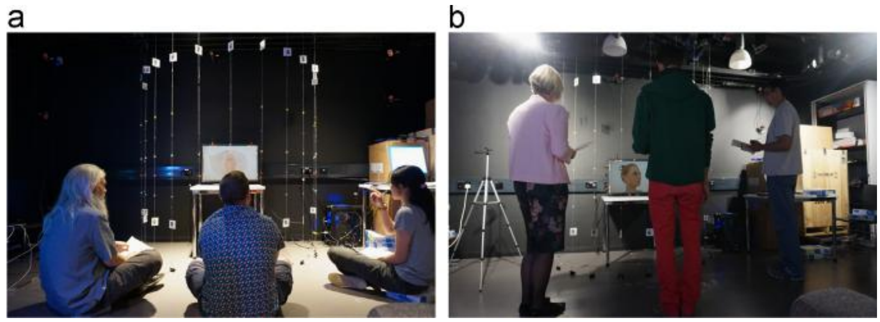

Fig. 5. Pictures of the experiment room were taken from different display conditions and vertical viewing angles. (a) Perspective-correct & stereoscopy with vertical viewing angle -10° (b) Conventional 2D with vertical viewing angle 20°.

Table 1
The target order generated for the experiment.

	-45°	-30°	-15°	0°	$+15^\circ$	$+30^\circ$	$+45^\circ$
$+20^\circ$	18	21	13	22	19	32	33
$+10^\circ$	1	16	14	25	6	17	20
0°	15	3	7	4	35	23	27
-10°	9	26	34	29	31	10	12
-20°	2	11	28	30	24	5	8

target (t_{oi_h}) and the horizontal position of the actual target (t_{ai_h}), converted to degrees, based on horizontal targets being 15° apart from each other:

$$\epsilon_{i_h} = |\, t_{oi_h} - t_{ai_h} | \times 15^\circ$$

Fig. 6 shows the mean horizontal error over all target positions at the three horizontal viewing angles for each three display conditions. Overall, the means of the perspective-correct & stereoscopy condition show that it achieved the lowest mean horizontal error. For both the perspective-correct & stereoscopy condition and the perspective-correct condition, the errors were similar across the three viewing angles, indicating that the viewing angle had little impact. However, for the conventional 2D condition, the errors increased as the viewing angle diverged from the central. Fig. 8(a) shows the mean horizontal error over all observer's viewing angles for each target positions and display conditions. For the target positions in the perspective-correct & stereoscopy condition and the perspective-correct condition, the mean horizontal errors are less than 15° (one target error). Interestingly, the errors in the perspective-correct & stereoscopy condition were more evenly distributed than the perspective-correct condition. The perspective-correct condition resulted in higher errors when viewing the horizontal edges of the target position grid than the more central locations.

A 3 display conditions $\times$ 3 horizontal viewing angles $\times$ 3 vertical viewing angles $\times$ 7 horizontal target positions mixed design ANOVA was conducted on the horizontal error, with display condition, horizontal viewing angles and vertical viewing angles as between-subjects factors and horizontal target positions as a within-subjects factor. Firstly, the main effect of display conditions was significant, $F(2, 108) = 341.029, p < .001$. Bonferroni post hoc tests revealed significant mean horizontal error differences between each of the display conditions. The perspective-correct & stereoscopy ($M = 5.095$, 95% *CI* [4.219, 5.971]) gave significantly lower mean horizontal error than the perspective-correct condition ($M = 9.857$, 95% *CI* [8.981, 10.733]), $p < .001$, and the conventional 2D condition ($M = 21$, 95% *CI* [20.124, 21.876]), $p < .001$. This supports the hypothesis 1a. Secondly, results revealed a significant main effect of horizontal viewing angles, $F(2, 108) = 108.166, p < .001$. Bonferroni post hoc tests revealed significant mean differences between each of the horizontal viewing angles. The mean at viewing angle 0°($M = 7.048$, 95% *CI* [6.171, 7.924]) is significantly lower than viewing angle -30°($M = 12.762$, 95% *CI* [11.886, 13.638]), $p < .001$ and viewing angle 45°($M = 16.143$, 95% *CI*[15.267, 17.019]), $p < .001$. The display conditions $\times$ horizontal viewing angle interaction was significant, $F(4, 108) = 146.865, p < .001$, indicating that the error due to viewing angles were different in three display conditions. This supports the hypothesis 2a. Thirdly, 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 been violated ($\chi^2(20) = 70.799, p < .001$), therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .804$). The mean horizontal error differed significantly across horizontal target positions, $F(4.826, 521.216) = 5.148, p < .001$. The display conditions $\times$ horizontal target positions interaction was also significant, $F(9.652, 521.216) = 6.198, p < .001$, indicating that the error due to horizontal target positions was different in three display conditions.

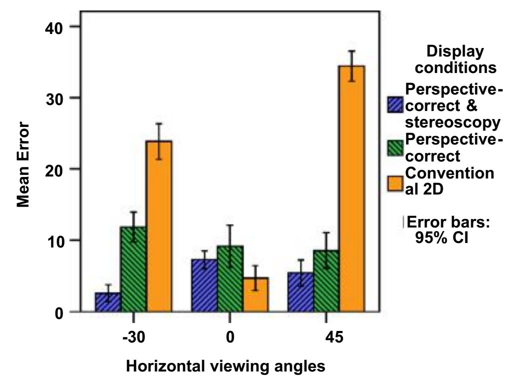

Fig. 6. The mean horizontal error for each display conditions and horizontal viewing angles.

4.3.2. Vertical error

We then define vertical error of each target (ϵ_{i_v}) to be the absolute value of difference between the vertical position of observer perceived target (t_{oi_v}) and the vertical position of actual target (t_{ai_v}) converted to degrees, based on attention targets being 10° apart from each other:

$$\epsilon_{i_v} = |\, t_{oi_v} - t_{ai_v} | \times 10^\circ$$

Fig. 7 shows the mean vertical error over all target positions at the three vertical viewing angles in three display conditions. The interpretations of the results in Fig. 7 were similar to those in Fig. 6. Fig. 9(a) shows the mean vertical error over all observer's viewing angles for each target positions and display conditions. The heat maps show that the perspective-correct & stereoscopy condition has lower mean horizontal error than the perspective-correct condition, particularly when viewing the top edge of the target position grid.

A 3 display conditions × 3 horizontal viewing angles × 3 vertical viewing angles × 5 vertical target positions mixed design ANOVA was conducted on the vertical error, with display condition, horizontal viewing angles and vertical viewing angles as between-subjects factors and vertical target positions as a within-subjects factor. Firstly, the main effect of display conditions was significant, $F(2,162)=45.483, p<.001$. Bonferroni post hoc tests revealed significant mean vertical error differences between each of the display conditions. The perspective-correct & stereoscopy ($M=3.016$, 95% CI [2.417, 3.614]) gave significantly lower mean vertical error than the perspective-correct condition ($M=5.429$, 95% CI [4.83, 6.027]), $p<.001$, and the conventional 2D condition ($M=7.079$, 95% CI [6.481, 7.678]), $p<.001$. This supports the hypothesis 1b. Secondly, there is a significant main effect of vertical viewing angles, $F(2,162)=26.967, p<.001$. Bonferroni post hoc comparisons shown the mean vertical error at vertical viewing angle 20° ($M=6.984$, 95% CI [6.386, 7.583]) is significantly higher than the mean vertical error at vertical viewing angle −10° ($M=4.413$, 95% CI[3.814, 5.011]), $p<.001$, and vertical viewing angle 0°($M=4.127$, 95 % CI [3.529, 4.725]), $p<.001$. However, the mean vertical error at vertical viewing angle 0° did not significantly differ from vertical viewing angle −10°($p>.05$). The display conditions × vertical viewing angle interaction was significant, $F(4,162)=29.25, p<.001$, indicating that the error due to viewing angles were different in three display conditions. This supports the hypothesis 2b. Thirdly, Mauchly's test indicated that the assumption of sphericity had not been violated ($\chi^2(9)=8.97, p>.05$). The mean vertical error differed significantly across vertical target positions, $F(4,648)=7.189, p<.001$. The display conditions × vertical target positions interaction was also significant, $F(8,648)=2.801, p=.005$, indicating that the error due to vertical target positions was different in three display conditions.

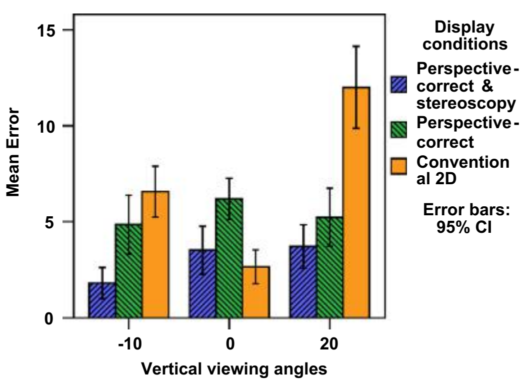

Fig. 7. The mean vertical error for each display conditions and vertical viewing angles.

4.3.3. *Horizontal bias*

We further investigated whether there was leftward bias or rightward bias in perceiving targets in different display conditions. We defined the horizontal bias of each target (β_{i_h}) to be the difference between the horizontal position of observer's perceived target (t_{oi_h}) and the horizontal position of the actual target (t_{ai_h}) converted to degrees:

$$\beta_{i_h}=(t_{oi_h}-t_{ai_h})\times 15^\circ$$

Fig. 10 shows the horizontal bias at three viewing angles in three display conditions. Positive values indicated leftward biases whereas negative values indicated rightward bias. For both the perspective-correct & stereoscopy and the perspective-correct conditions, the mean target bias did not change substantially across different viewpoints. By contrast, for the conventional 2D condition, the biases depended on the observers' horizontal viewing angles. When we consider the target positions in Fig. 12, we see the bias does not vary with target positions for the perspective-correct & stereoscopy condition, however, it increases as the target position gets further away from the observer for perspective-correct condition. Considering the horizontal observer at viewing angle −30°, we see that for the target position −30°, both the perspective-correct & stereoscopy condition and the perspective-correct condition has similar bias around 0°; however, for the target position 45° the bias increases to 20° in perspective-correct condition. This overestimation pattern is repeated for all horizontal viewing angles.

A 3 display conditions × 3 horizontal viewing angles × 3 vertical viewing angles × 7 horizontal target positions mixed design ANOVA was conducted on the horizontal bias, with display condition, horizontal viewing angles and vertical viewing angles as between-subjects factors and horizontal target positions as a within-subjects factor. Firstly, the main effect of display conditions was significant, $F(2,108)=15.068, p<.001$. However, Bonferroni post hoc tests revealed that the mean horizontal bias in the perspective-correct & stereoscopy did not significantly differ from the perspective-correct condition, $p>.05$. Secondly, results revealed a significant main effect of horizontal viewing angles, $F(2,108)=388.936, p<.001$. Bonferroni post hoc tests revealed significant mean differences between each of horizontal viewing angles. Thirdly, Mauchly's test indicated that the assumption of

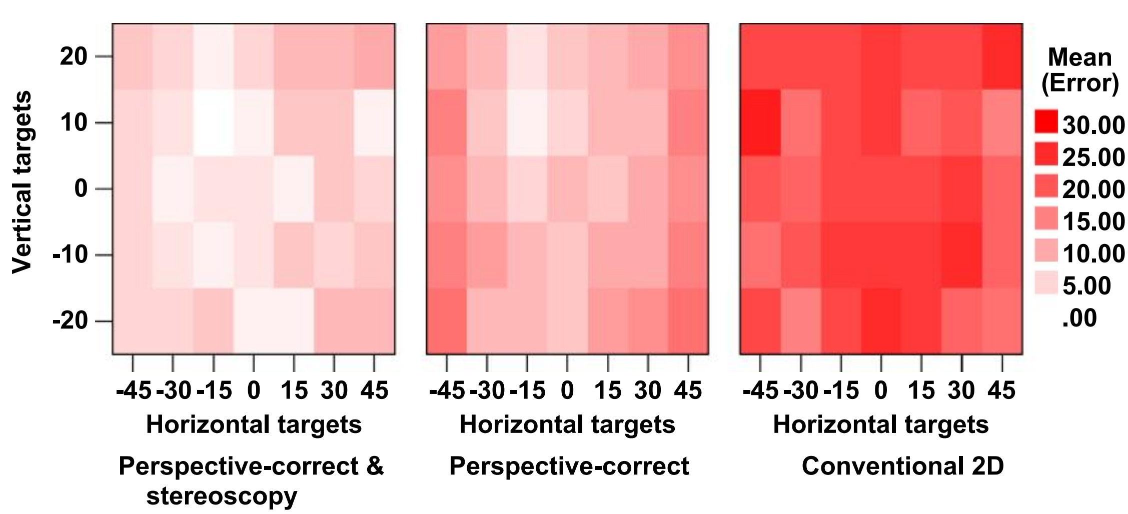

Fig. 8. Heat maps showing the mean horizontal error for each display condition and target position.

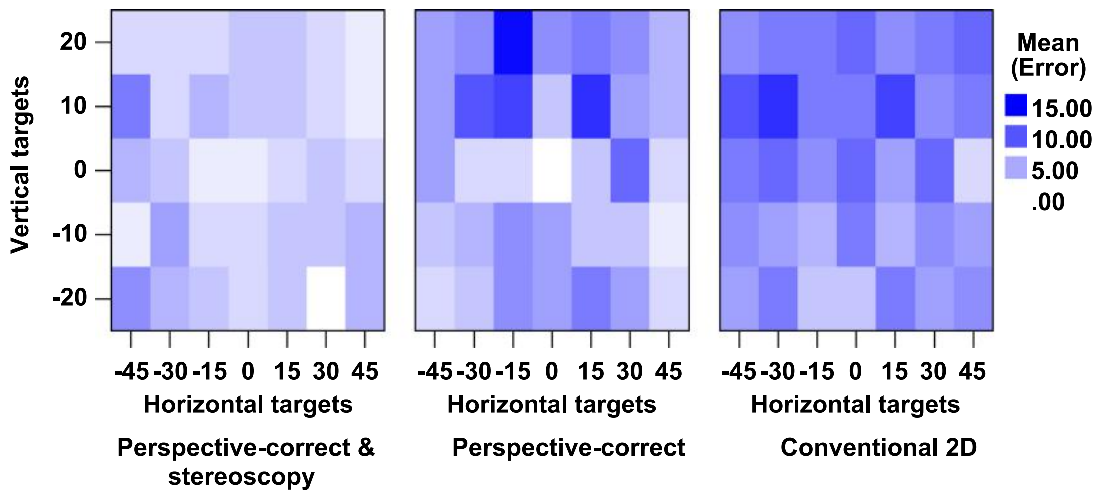

Fig. 9. Heat maps showing the mean vertical error for each display condition and target position.

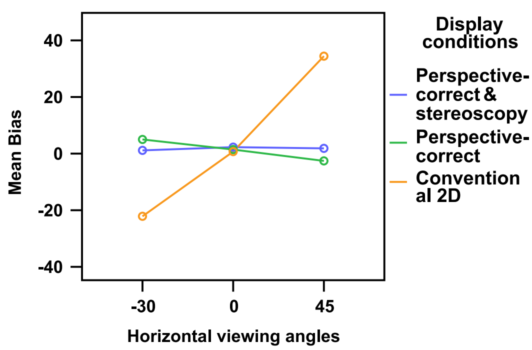

Fig. 10. The mean horizontal bias for each display conditions and horizontal viewing angles.

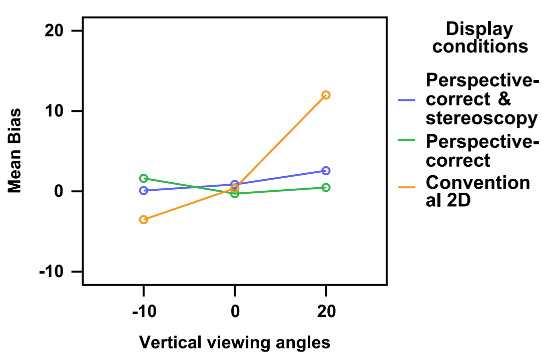

Fig. 11. The mean vertical bias for each display conditions and vertical viewing angles.

sphericity had been violated ($\chi^2(20) = 68.76, p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .819$). The mean horizontal bias differed significantly across horizontal target positions, $F(4.914, 530.689) = 125.396, p < .001$. The display conditions × horizontal target positions interaction was significant, $F(9.828, 530.689) = 11.389, p < .001$. The horizontal viewing angle × horizontal target positions interaction was significant, $F(9.828, 530.689) = 1.832, p < .001$. The display conditions × horizontal viewing angle × horizontal target positions interaction was also significant, $F(19.655, 530.689) = 6.515, p < .001$, indicating that the bias due to horizontal target positions was present differently in three horizontal viewing angles and three display conditions.

4.3.4. Vertical bias

Next, we defined the vertical bias of each target (β_{i_v}) to be the difference between the vertical position of observer's perceived target (t_{oi_v}) and the vertical position of actual target (t_{ai_v}) converted to degrees:

$$\beta_{i_v} = (t_{oi_v} - t_{ai_v}) \times 10^\circ$$

Fig. 11 shows the vertical bias at three viewing angles in three display conditions. Fig. 13 shows the vertical bias for each display conditions, vertical viewing angles and horizontal target position. Positive values indicated upward biases whereas negative values indicated downward bias. The interpretation of the vertical bias were similar to those of horizontal bias, but with less effect.

A 3 display conditions × 3 horizontal viewing angles × 3 vertical viewing angles × 5 vertical target positions mixed design ANOVA was conducted on the vertical bias, with display condition, horizontal viewing angles and vertical viewing angles as between-subjects factors and horizontal target positions as a within-subjects factor. Firstly, the main effect of display conditions was significant, $F(2, 162) = 13.141, p < .001$. However, Bonferroni post hoc tests revealed that the mean vertical bias in the perspective-correct & stereoscopy did not significantly differ from the perspective-correct condition, $p > .05$. Secondly, results revealed a significant main effect of vertical viewing angles, $F(2, 162) = 79.521, p < .001$. Bonferroni post hoc comparisons indicated the mean vertical bias at vertical viewing angle 20° is significantly different from vertical viewing angle $-10^\circ, p < .001$ and vertical viewing angle $0^\circ, p < .001$. However, the mean vertical error at vertical viewing angle 0° did not significantly differ from vertical viewing angle $-10^\circ, p > .05$. Thirdly, Mauchly's test indicated that the assumption of sphericity had not been violated ($\chi^2(9) = 13.571, p > .05$). The mean vertical bias differed significantly across vertical target positions, $F(4, 648) = 37.908, p < .001$. The display conditions × vertical target positions interaction was significant, $F(8, 648) = 9.108, p < .001$. However, the display conditions × vertical viewing angle × vertical target positions interaction was not significant, $F(16, 648) = 1.562, p > .05$.

5. Discussion

5.1. Effects of 3D perspective on head gaze cue assessment

Results from this experiment confirmed our hypotheses. We found that participants performed with the lowest error when interpreting the avatar's head gaze direction in the perspective-correct & stereoscopy condition, followed by the perspective-correct condition, and

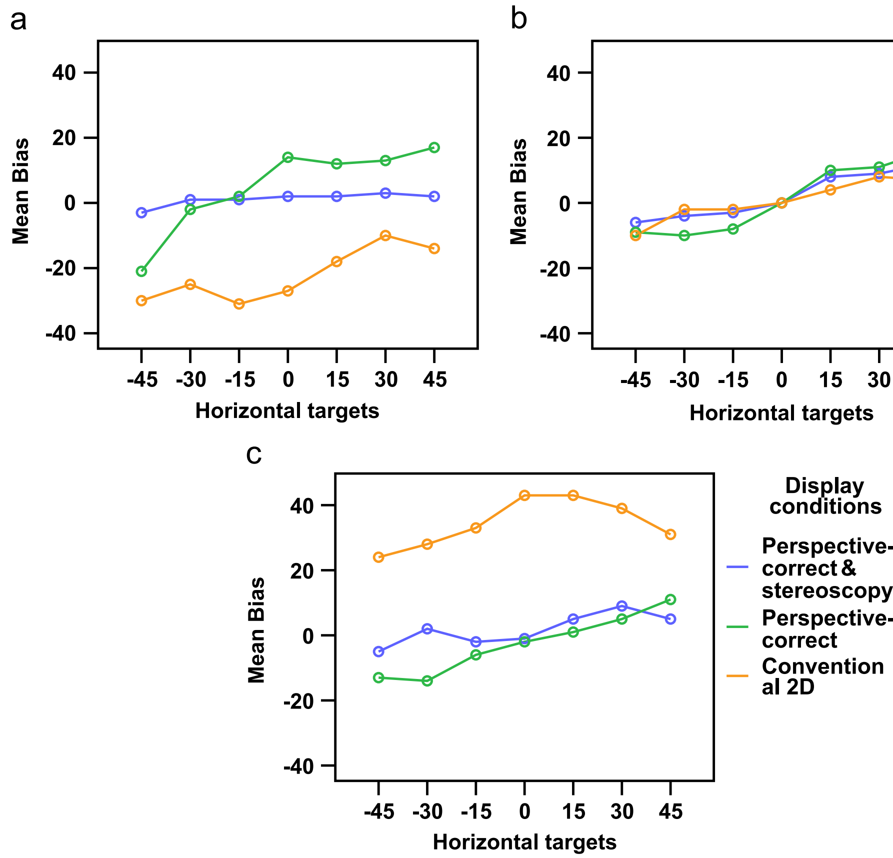

Fig. 12. The mean horizontal bias for each display conditions, horizontal viewing angles and horizontal target position. (a) Horizontal viewing angle -30° (b) Horizontal viewing angle 0° (c) Horizontal viewing angle 45°.

then the traditional 2D condition. This is consistent with Kim et al.'s previous findings in 3D video communication (Kim et al., 2012).

The poor performance of the traditional 2D condition was expected because the head is always rendered from the front perspective. The only position with the correct perspective would be the observer at centre where the front perspective correlates to that observer's perspective. For the other viewing positions, the biases depended on the observers' viewing angles. The observers would be experiencing the Mona Lisa gaze effect. They would perceive the head gaze direction as if they were standing straight in front of the display. Thus, they would see the head gaze in a relative rather than an absolute manner. As expected, Figs. 12 and 13 show that the curves of the traditional 2D condition maintain a similar shape, but are shifted depending on observer's perspective. This parallels the previous findings (Al Moubayed et al., 2012; Nguyen et al., 2005) for 2D video conditions.

For the comparison between the perspective-correct condition and the perspective corrected & stereoscopy condition, we found the differences in vertical and horizontal errors were statistically significant. However, the differences in vertical and horizontal bias were not statistically significant. This suggested that perspective-correct alone could reduce the shifting bias discussed above.

Figs. 12 and 13 show that the overestimation pattern in the perspective-correct condition is interesting. They indicate the addition of stereoscopy could reduce an overestimation of the deviation of avatar's head gaze, thus further improving the observers' ability to identify more correct targets. This was also backed up by results from the vertical and horizontal errors. An analysis of the heat maps in Fig. 8(a) and Fig. 9(a) show that observers performed with higher level of error when viewing the edges of the target grid than the more central locations in the perspective-correct condition. This effect appears very reliable and this means that it may be possible to model and thus predict the distortion. We plan to further explore on this finding, with our next step being to collect more data for more viewing angles.

We also investigated judgments of vertical direction of head gaze. Figs. 12 and 13 show that the magnitude of the shifting bias in 2D condition and the overestimation pattern in the perspective-correct condition are smaller in vertical direction comparing to horizontal direction. This discrepancy in results between judgments of horizontal and of vertical head gaze reflects the asymmetric sensitivity of users when perceiving avatar's head outline. This is supported by the previous findings (Wilson et al., 2000) that the perceived direction of gaze can be influenced by deviation of the head profile from bilateral symmetry, and deviation of nose orientation from vertical.

5.2. Display characteristics

The results demonstrated that our system can better represent the remote person's head gaze for group teleconferencing (e.g., observers with different heights), by providing multiple simultaneous stereo views from arbitrary positions. The observer's maximum viewing angle depends on the LCD panel's viewing angle. We used the SIPS type display and the maximum viewing angle is at least 70° in each direction. Although the random hole type display has a limited spatial resolution (see Section 3.2), on our display the different views are easily distinguished. Fig. 3 is a set of stereo pair images, showing the autostereoscopic image quality provided to three users. Additionally, all participants in our experiment confirmed that they are able to clearly tell where the avatar's eyeball is actually looking. In the future, we expect to use brighter and higher-density LCD/LED panels or high-resolution multiple projector systems to further improve image quality.

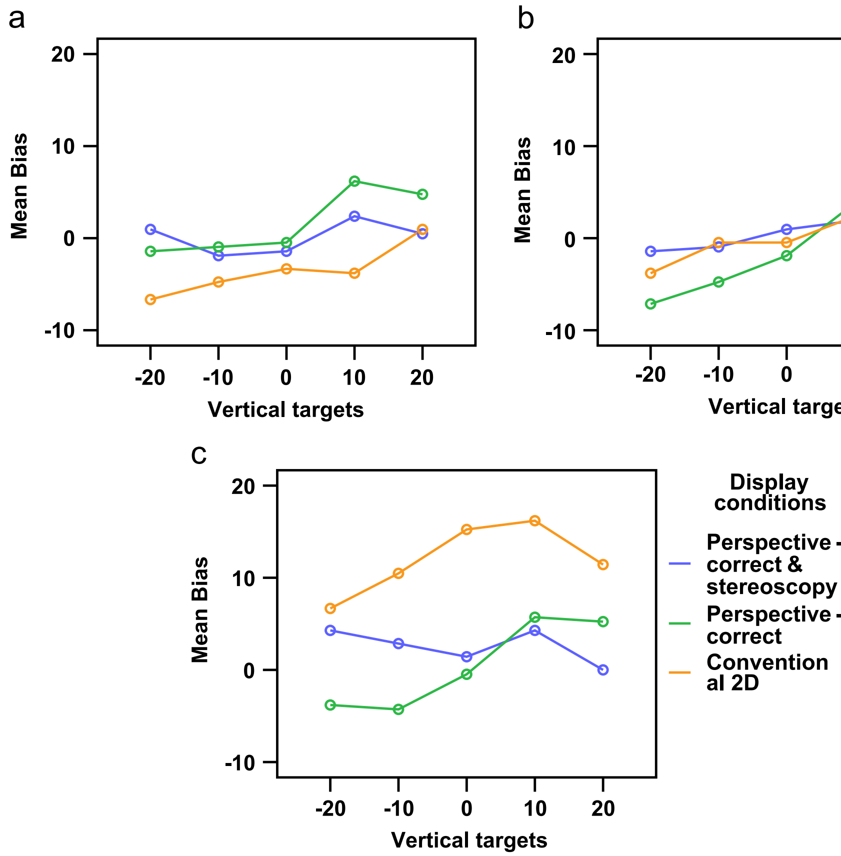

Fig. 13. The mean vertical bias for each display conditions, vertical viewing angles and horizontal target position. (a) Vertical viewing angle -10° (b) Vertical viewing angle 0° (c) Vertical viewing angle 20°.

5.3. Support for different telepresence scenarios

With our current demonstration we are using a ray-traced avatar head. Although the animation we have used in the experiment is simple, the software system supports a fully animated head with eye movement and facial expression, using Faceshift® with Microsoft Kinect™ to obtain the remote person's eye movement and facial expression in realtime.

The current experiment demonstrated head gaze preserving capability of our system for three simultaneous participants. We plan to investigate eye gaze preserving capability of our by recording the remote person's eye movement. Also, since better gaze judgment will make for better conversation, we expect that the social interaction will be genuinely improved (particularly for multi-person conversation than in dyad). Thus upcoming experiments will focus on social effect (e.g. trust).

There are several routes for development to support different conversation scenarios. Firstly, our current system can be used for asymmetric conversations. This setup could be mirrored to support symmetric conversations. Secondly, our current display allows observers to see perspective-correct stereo images from multiple viewpoints. It could also support free viewpoints by tracking observers' positions. Thirdly, we hope to leverage our system for 3-way or N-way teleconferencing scenarios. Support of a teleconference with N users requires $N \times (N-1)$ data streams. Since avatar mediated interaction does not require significant bandwidth for transmission, our design would easily allow for such scaling. Lastly, an interesting question is the potential support for live video streaming. We plan to further investigate on this topic, perhaps using a light field camera to capture the remote person or a 360° array of cameras around the remote person.

6. Conclusion

We have presented a ray-traced view-dependent rendering method to represent the remote person as a virtual avatar on a random hole display. The display offers a number of capabilities that are not found in most existing autostereoscopic displays, including display for multiple users in arbitrary viewing positions. Although the image quality of the random hole display is limited compared to other display technologies, the unique view content is easily distinguished. The low cost and ease of setup make this system an interesting platform on which to simulate scenarios that require multiple simultaneous stereo views from arbitrary positions.

We empirically evaluated the effect of perspective correct and stereo imagery on users' accuracy in judging head gaze direction. Results revealed that perspective correct imagery provides a dominant effect in improving the effectiveness with which users were able to estimate the gaze direction, with additional effect for perspective-correct augmented by stereoscopy. Results also show magnitude of the bias due to the lack of perspective-correct and stereoscopic cues is less sensitive to vertical direction than horizontal direction.

As the amount of time we spent in mediated interaction increases, these findings have significant implications for teleconferencing in general. Our system provides perspective-correct stereoscopic imagery for multiple users in arbitrary positions without the need for special glasses; and hence avoids the distortion of the gaze cues we have observed with traditional displays. Gaze, attention, eye contact is a fundamental part of human interaction and we intend to explore other important scenarios and natural interaction in future work.

Acknowledgements

The authors would like to acknowledge the FP7 EU collaborative project BEAMING(248620) and the China Scholarship Council. We thank Henry Fuchs, Andrei State and Gu Ye for their advice on building the random hole display. We thank the UCL VECG group members for their general support and the anonymous reviewers for their feedback.

References

Al Moubayed, S., Edlund, J., Beskow, J., 2012. Taming mona lisa: communicating gaze faithfully in 2D and 3D facial projections. Trans. Interact. Intell. Syst. 1 (2), 25.
Anstis, S.M., Mayhew, J.W., Morley, T. The perception of where a face or television ‘portrait’ is looking. Am. J. Psychol.
Argyle, M., Cook, M. Gaze and 1976,Cambridge U Press.
Arthur, K.W., Booth, K.S., Ware, C., 1993. Evaluating 3D task performance for fish tank virtual worlds. ACM Trans. Inf. Syst. (TOIS) 11 (3), 239–265.
Böcker, M. Runde, D., Mühlbach, L., 1995. On the reproduction of motion parallax in videocommunications. In: HFES, vol. 39, SAGE Publications.
Böcker, M., Blohm, W., Mühlbach, L., 1996. Anthropometric data on horizontal head movements in videocommunications. In: SIGCHI, ACM, Vancouver, British Columbia, Canada.
Chen, M., 2002. Leveraging the asymmetric sensitivity of eye contact for videoconference. In: SIGCHI, CHI '02, ACM, Minneapolis, Minnesota, USA, pp. 49–56.
Dunbar, D., Humphreys, G., 2006. A spatial data structure for fast poisson-disk sample generation. In: TOG, vol. 25, ACM, pp. 503–508.
Ferreira, F., Cabral, M., Belloc, O., Miller, G., Kurashima, de Deus Lopes, C.R., Stavness, I., Anacleto, J. Zuffo, M., Fels, S. In: UIST, Spheree: A 3D perspective-corrected interactive spherical scalable display, ACM, Vancouver, British Columbia, Canada, pp. 77-86.
Gibson, J.J., Pick, A.D., 1963. Perception of another person's looking behavior. Am. J. Psychol., 386–394.
Gross, M., Würmlin, S., Naef, M., Lamboray, E., Spagno, C., Kunz, A., Koller-Meier, E., Svoboda, T., Van, L., G., Lang, S., Strehlke, K., Moere, A., Staadt, O., 2003. Blue-c: a spatially immersive display and 3D video portal for telepresence. In: SIGGRAPH, ACM, San Diego, California, USA.
Jones, A., Lang, M., Fyffe, G., Yu, X., Busch, J., McDowall, I., Bolas, M., Debevec, P., 2009. Achieving eye contact in a one-to-many 3D video teleconferencing system. In: TOG, vol. 28, ACM, p. 64.
Karnik, A., Mayol-Cuevas, W., Subramanian, S., 2012. Mustard: a multi user see through ar display. In: SIGCHI, ACM, Austin, Texas, USA, pp. 2541–2550.
Kim, S., Cao, X., Zhang, H., Tan, D., 2012. Enabling concurrent dual views on common LCD screens. In: SIGCHI, ACM, Austin, Texas, USA, pp. 2175–2184.
Kim, K., Bolton, J., Girouard, A., Cooperstock, J., Vertegaal, R., 2012. Telehuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In: SIGCHI, ACM, Austin, Texas, USA, pp. 2531–2540.
Kitamura, Y., Konishi, T., Yamamoto, S., Kishino, F., 2001. Interactive stereoscopic display for three or more users. In: SIGGRAPH, ACM, Los angeles, California, USA, pp. 231–240.
Kooima, R., Prudhomme, A., Schulze, J., Sandin, D., DeFanti, T., 2010. A multi-viewer tiled autostereoscopic virtual reality display. In: Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology, ACM, Hong Kong, pp. 171–174.
Lincoln, P., Welch, G., Nashel, A., Ilie, A., Fuchs, H. Animatronic shader lamps avatars,.In: Virtual Reality, Springer, pp. 225-238.
Matusik, W., Pfister, H. 3d tv: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. In: TOG, vol. 23, ACM.
Nagano, K., Jones, A., Liu, J., Busch, J., Yu, X., Bolas, M., Debevec, P., 2013. An autostereoscopic projector array optimized for 3D facial display. In: ACM SIGGRAPH 2013 Emerging Technologies, ACM, Anaheim, California, USA, 3.
Nashel, A., Fuchs, H., 2009. Random hole display: a non-uniform barrier autostereoscopic display. In: 3DTV-CON, IEEE, Potsdam, Brandenburg, Germany, pp. 1–4.
Nguyen, D., Canny, J., 2005. Multiview: spatially faithful group video conferencing. In: SIGCHI, Portland, Oregon, USA, pp. 799–808.
Okada, K., Maeda, F., Ichikawaa, Y., Matsushita, Y., 1994. Multiparty videoconferencing at virtual social distance: Majic design. In: CSCW, Chapel Hill, North Carolina, USA, pp. 385–393.
Oyekoya, O., Steptoe, W., Steed, A., 2012. Sphereavatar: a situated display to represent a remote collaborator. In: SIGCHI, Austin, Texas, USA, pp. 2551–2560.
Pan, Y., Steed, A., 2014. A gaze-preserving situated multiview telepresence system. In: SIGCHI, ACM, Toronto, Ontario, Canada, pp. 2173–2176.
Pan, Y., Steptoe, W., Steed, A., 2014. Comparing flat and spherical displays in a trust scenario in avatar-mediated interaction. In: SIGCHI, ACM, Toronto, Ontario, Canada, pp. 1397–1406.
Perlin, K., Paxia, S., Kollin, J.S., 2000. An autostereoscopic display. In: SIGGRAPH, ACM Press/Addison-Wesley Publishing Co., pp. 319–326.
Peterka, T., Kooima, R.L., Sandin, D.J., Johnson, A., Leigh, J., DeFanti, T.A., 2008. Advances in the dynallax solid-state dynamic parallax barrier autostereoscopic visualization display system. IEEE Trans. Vis. Comput. Graph. 14 (3), 487–499.
Prince, S., Cheok, A.D., Farbiz, F., Williamson, T., Johnson, N., Billinghurst, M., Kato, H., 2002. 3-D live: real time interaction for mixed reality. In: CSCW, New Orleans, Louisiana, USA.
Roberts, D., Wolff, R., Rae, J., Steed, A., Aspin, R., McIntyre, M., Pena, A., Oyekoya, O., Steptoe, W., 2009. Communicating eye-gaze across a distance: Comparing an eye-gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together. In: VR, Lafayette, LA,USA, pp. 135–142.
Roberts, D.J., Rae, J., Duckworth, T.W., Moore, C.M., Aspin, R., 2013. Estimating the gaze of a virtuality human. IEEE Trans. Vis. Comput. Graph. 19 (4), 681–690.
Sandin, D.J., Margolis, T., Ge, J., Girado, J., Peterka, T., DeFanti, T.A., 2005. The varrier TM autostereoscopic virtual reality display. In: TOG, vol. 24, ACM.
Schreer, O., Kauff, P., Sikora, T., 2005. 3D Videocommunication. Wiley Online Library.
Sellen, A., Buxton, B., Arnott, J., 1992. Using spatial cues to improve videoconferencing. In: SIGCHI, Monterey, California, USA, pp. 651–652.
Sollenberger, R.L., Milgram, P., 1991. A comparative study of rotational and stereoscopic computer graphic depth cues. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 35, SAGE Publications, pp. 1452–1456.
Stavness, I., Lam, Fels, B.S., 2010. pcubee: a perspective-corrected handheld cubic display. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Atlanta, Georgia, USA, pp. 1381–1390.
Steptoe, W., Oyekoya, O., Murgia, A., Wolff, R., Rae J., Guimaraes, E., Roberts, D., Steed, A., 2009. 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, Lafayette, LA, USA, 2009, pp. 83–90.
Stiefelhagen, R., Zhu, J., 2002. Head orientation and gaze direction in meetings. In: CHI'02 Extended Abstracts on Human Factors in Computing Systems, ACM, Minneapolis, Minnesota, USA, pp. 858–859.
Troje, N.F., Siebeck, U., 1998. Illumination-induced apparent shift in orientation of human heads. Perception–London 27, 671–680.
Vertegaal, R., Weevers, I., Sohn, C., Cheung, C., 2003. Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction. In: SIGCHI, Ft. Lauderdale, Florida, USA, pp. 521–528.
Wilson, H.R., Wilkinson, F., Lin, L.-M., Castillo, M., 2000. Perception of head orientation. Vis. Res. 40 (5), 459–472.
Ye, G., State, A., Fuchs, H., 2010. A practical multi-viewer tabletop autostereoscopic display. In: ISMAR, IEEE, Seoul, South Korea, pp. 147–156.