

# **Telethrone Reconstructed; ongoing testing toward a more natural situated display.**

John O'Hare<sup>a</sup>,  
Allen Fairchild<sup>a</sup>,

Robin Wolff<sup>b</sup>, and  
David Roberts

<sup>a</sup> School of Computer Science and Engineering  
University Of Salford, United Kingdom  
j.ohare@salford.ac.uk

<sup>b</sup> Deutschen Zentrums für Luft- und Raumfahrt  
Linder Höhe, 51147 Köln, Germany  
Robin.Wolff@dlr.de

<sup>c</sup> School of Health Sciences  
University Of Salford, United Kingdom  
d.j.roberts@salford.ac.uk

## **Abstract**

The concept of supporting ad-hoc or dynamic membership tele-present meetings through pulling up a chair is novel. In real world business situations, people pull up a chair after catching the eye of someone already seated. Telethrone is a situated display on a chair which allows multiple correct views of a remote collaborator. The system has been expanded to support informal meetings where chairs can be moved around. This is facilitated through the novel integration of a 3D reconstructed model of a person, with live viewpoint dependent rendering onto a retro-reflective surface. This removes the need for painstaking alignment of multiple cameras and projectors each time a chair is moved. A between subjects experiment tested accuracy of reconnected mutual gaze mediated by part of the system. Subjectively easier and harder situations are compared. Specifically best and worst cases, both in terms of orientation of eyes in the reconstructed head, and angle of observer gaze onto the display. Discussion compares results to experiments that used other systems to attempt to convey eye gaze by different techniques. This research builds toward a scalable system for ad-hoc business meetings; a paradigm poorly supported by current video conferencing. It is also applicable to supporting conversations between seated people in any scenario where seats might be moved, for example in interaction between client and therapist in tele-therapy.

**Keywords:** Tele-presence; situated display; gaze; retroreflective projection technology; non-verbal cues; multi-view

## **1 Introduction**

The vision of this work, from tele-presence artist Paul Sermon, is of pulling up a chair after catching someone's eye; in teleworking just as the real world. The approach is to project various viewpoint dependent renders of the remote

participant onto a chair that reflects each render in a separate retro-projection frustrum. The original Telethrone (O'Hare et al., 2016) was a chair covered in a retro-reflective material onto which 2D video streams showing the remote person from different perspectives were projected. The retroreflective material allows viewers from each pulled up chair to have their own perspective view. With chairs evenly distributed around the Telethrone, it should in principle also be possible to support mutual awareness between a telepresent seated person and a person walking past. *withyou* (Roberts et al., 2015, Fairchild 2016) was a telepresence system that recreates a 3D CGI copy of a remote person that can be viewed from any perspective but was derived from multiple discrete 2D video streams. Before integration of *withyou*, the mona lisa effect meant that a pulled chair would need to exactly align to a remote camera if mutual gaze were to be supported. However, there are shortcomings of both the retro-reflected material and of free-viewpoint video, that might impact on estimation of gaze. Specifically, the retro-reflective material suffers image bleed between spatial channels, and free-viewpoint video introduces artifacts. As one discrete step in development we present an early prototype and pilot that test these combined shortcomings do not prohibit reasonable gaze estimation.



Figure 1: Experimental deployment in a semi-public space, testing elements of the system.

### ***Example of potential utility***

To intuitively outline the utility of the new system to business users a hypothetical meeting is convened to discuss whether a project is likely to overrun. This meeting comes about without warning; an executive in the company is in the building on other business and gets into a conversation in the social space with a developer. The product owner who might be expected to oversee such matters is offsite but can be present through a single Telethrone which is installed in the space. The executive invites two clients to join him. As the meeting begins another developer from the product team passes and is signalled to join by the tele-present product owner. Now five people are physically present in the space, while one is tele-present. The members of the meeting sit down and naturally cluster into two contiguous groupings: The two clients and the executive, and the three members of the product team. This distribution can be seen in Figure 2. As the meeting evolves it becomes necessary for the product team to clarify a technical point amongst themselves. In the posited Telephone supported meeting it is

possible for the two developers to form a ‘huddle’ with the product owner. This meeting sub group can fork the conversation, while the clients and executive continue with other matters. The described Telethron system with tracking of the chairs and 3D reconstruction of the associated viewpoints can support this evolution of a meeting.

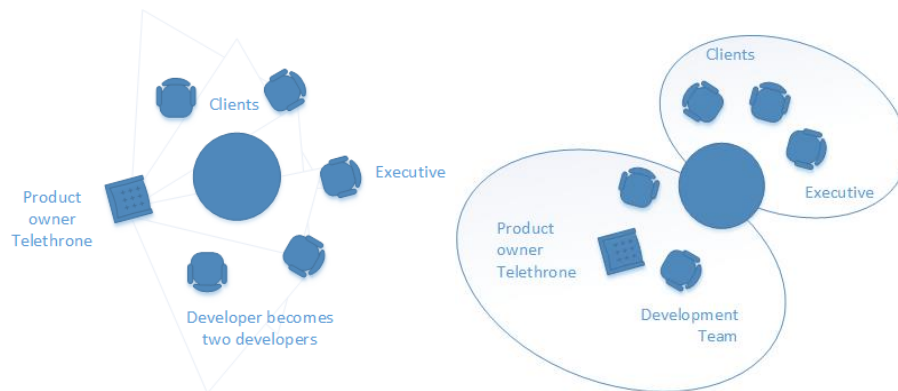


Figure 2. A normal ad-hoc business meeting evolves over time and is supported by the Telethron (square chair).

## 2 Literature review

Communicating both attention and appearance has been a long standing challenge to computer supported collaborative working (Roberts et al., 2015). Telepresence solutions aimed at collaborative working have traditionally attempted to join remote spaces so that people in each can look into the other, seeing each other, at best, as if through a window (Roberts et al., 2015). Immersive Collaborative Virtual Environments (ICVE) have been used to join remote spaces so that they coincide (Roberts et al., 2003), and free-viewpoint video has been combined with immersive projection technology to allow people to move around spaces, seemingly together (Roberts et al., 2015). In ICVE's people in each space can have the appearance and feeling of moving around the other, however, true identity and facial expression are hard to communicate without restricting movement (Roberts et al., 2015).

Situated displays have attempted to give the impression of placing people within each other's room by contextualising the display, or engineering it such that the background of the remote setting is omitted (Oyekoya, Steptoe, & Steed, 2012). In approaches to date, situated displays and capture equipment cannot be moved without painstaking alignment of cameras and projectors across the two spaces.

Human acuity in judgement of the eye gaze of others is high (Symons et al., 2004). This keen sense of gaze is important for group social interaction (Colburn et al., 2000). Gaze awareness does not just mediate verbal communication, but rather, is a complex channel of communication in its own right (Vertegaal et al., 2001) (Otsuka et al., 2005). It also governs those who are involved in the communication at any one time (Roberts et al., 2013). Physical communication channels extend beyond the face (Nguyen & Canny, 2009) and include both micro (shrugs, hands and arms), and macro movement of the upper body (Ekman, 1993). These conversational hand gestures augment verbal communication (Krauss, et al., 1996). In multi-party conversation, body torque (the

rotation of the trunk from front facing) can convey aspects of attention and focus (Schegloff, 1998). Several decades of research and commercial activity in telecommunication and telepresence have failed to fully support transmission of all aspects of interpersonal communication (Roberts et al., 2015). Some non-verbal communication is supported in video-conferencing (VC) with limited success. The Mona Lisa effect means that in most systems both directional eye gaze and mutual eye gaze are unsupported (Vishwanath et al., 2005) (Loomis et al., 2008) as anyone looking into a camera appears to be making eye contact with all of the observers at the other end of the video channel. Commercial systems such as Cisco Telepresence Rooms cluster their cameras above the centre of three screens for meetings using their tele collaboration product. They admit that this only works well for the centrally seated observer and the brain must compensate for this (Wolff et al., 2008; Roberts et al., 2015). Therefore, in camera based telecommunication systems the angle offset from a camera “should be at most 1.2 degrees in horizontal direction, and 1.7 degrees in vertical direction to support eye contact” (Bock et al., 2008), otherwise the offset effect should be corrected. Nguyen and Canny demonstrated the Multiview system (Nguyen & Canny, 2007) and observed similar task performance in face-to-face meetings based on trust tasks, while a similar approach without spatial segmented views of the remote participant was seen to negatively impact performance. They also found that “upper-body framing improves empathy measures” (Nguyen & Canny, 2009). Pan, Steptoe and Steed found similar results with their spherical display, with a decrease in trust toward avatar mediated conversation when viewing 2D displays at oblique angles (Pan et al., 2014). Kim et al demonstrate the Telehuman system, capable of conveying both the appearance and the direction of attention of a figure projected into a tube in 2D and 3D (Kim et al., 2012). There are several methods of reconstructing the human form as a computer model in real time. It is possible to create geometric models of the human form using a technique called shape from silhouette (Grau, et al., 2007) in which multiple viewpoints from cameras allow algorithmic generation of a visual hull (Franco & Boyer, 2003). 3D video has been successfully integrated into ICVE systems that use projectors to surround the user with a virtual environment (Roberts et al., 2015). Both spatiality and appearance of the remote collaborator are maintained, supporting complex interaction (Roberts et al., 2015). Situated displays place a representation of the remote user into a space, theoretically allowing all participants to physically interact with the ‘contextual configurations’ around them (Goodwin, 2000). This is a relatively new field of research that includes Telepresence robots (Lee & Takayama, 2011), Telehuman (Kim et al., 2012), head in a jar implementations such as SphereAvatar (Pan & Steed, 2012), and Gaze Preserving Situated Multi-View Telepresence System (Pan & Steed, 2014). Telehuman is especially pertinent and brings the whole body of a standing remote user into a space via a cylindrical display with a single tracked observer viewpoint. Situated displays seek to embed the represented participant within the spatial and contextual framework of the conversation such that referential cues are better supported. This has many implications, but chief amongst these is support for a spatially faithful conversational environment supportive of gaze. SphereAvatar demonstrates that there are problems with accurate mapping, distortion, projection, and movement of the captured participants outside the limits of the capture volume (Oyekoya, Steptoe, & Steed, 2012). Retro-reflective materials such as Chromatte(tm) cloth reflect light back along the angle of incidence (Tachi,

2003). An everyday application of such material is high visibility jackets. They have previously been researched by Tachi for mutual tele-existence (Tachi et al., 2004). In their system the motion of users is mirrored by a tele-operated robot to which a retro-reflected view of the remote user's head is projected (from the head of the onlooking user). Krum et al describe the REFLCT system (Krum et al., 2012) which uses large retro-reflective surfaces to “provide[s] users with a personal, perspective correct view of virtual elements that can be used to present social interactions with virtual humans”. They also use helmet mounted projectors and describe a military training application in a large volume which allows faithful transmission of attention and gestures from the virtual to the real. They also briefly describe augmenting a facial mannequin by projecting onto retro-reflective projection technology (RPT) adhered to the surface. Importantly the material negates much of the effect of surface normals thereby presenting even light reflection regardless of deformations in the surface. They also point out that the optical characteristics of the material maintain polarisation and so could support passive stereoscopy. Room2room from Microsoft labs (Benko, 2016) demonstrates the utility of projecting onto furniture. The system uses a Kinect 2.5D camera to capture a remote participant, and an overhead projector combined with a Kinect to projection map a viewpoint correct image of the tele-present person onto a complex surface such as a chair. However, this is a single static view-point. The literature does not explicitly engage with the potential for inviting passing parties to join a conversation. Additionally, there is poor support for dynamic shifts in seating within a meeting. The original Telethrone (O’Hare, et al, 2016) was a situated display with fixed positions at both locations. It was thus similar to SphereAvatar (Pan & Steed, 2012) and Gaze Preserving Situated Multi-View Telepresence System (Pan & Steed, 2014). In the paper it was argued that projection of a person onto furniture was likely to be less odd (or perhaps uncanny) and more familiar than seeing a head in a jar as with SphereAvatar, or a person in a tube as in Telehuman (Kim et al., 2012). The original Telethrone was tested with two spatially distinct moving images projected from live IP based video feeds but found the results for eye gaze reconnection inconclusive (O’Hare et al, 2016). ‘*withyou*’ is an experimental capture and playback system which uses the octave multimodal suite (Roberts et al., 2015). This system uses shape-from-silhouette reconstruction (Duckworth & Roberts, 2012) to send a full 3D video polygonal hull to another rendering location. Previous tests on the system suggested that the capture and playback made it possible to judge the eye gaze of the reconstructed subjects to within limits which underpin social interaction (Roberts et al., 2013). It was developed onward for a mixed reality system with multiple sites collaborating on shared data (Fairchild et al., 2016).

### 3 Methods

In order to assess if the 3D reconstruction and eye gaze elements (Roberts et al., 2015) work with the Telethrone (O’Hare et al., 2016) a limited experiment is undertaken where the chair is decoupled from the projector in order to check whether the system allows gaze discrimination. This experiment was inspired by (Kim et al., 2012) and (Roberts estimating eye gaze). By testing the harder case of gaze estimation as someone walks past, also gives some level of confidence that gaze would be supported if a chair was pulled to any point, the simpler case. The experimental setup demonstrates the

feasibility of components of the system. This is an iterative step towards a complete system with multiple mobile chairs facing a Telethrone.

### 3.1 The Telethrone System

The new Telethrone concept moves toward simple support for social space deployments, engaging attention outside of the meeting group. Dynamic group membership might involve both someone walking toward or past and or pulling up a chair. The technology approach was to combine the earlier Telethrone situated display and *withyou* telepresence system. By doing so virtual cameras join the sets of real cameras and real projectors, removing the need for the two to have the same spatial arrangement. Viewpoint is rendered according to position of onlooker. This means tracking someone walking past, or tracking a seat that is moved up.

Further refinements of the *withyou* capture system have been undertaken for integration with Telethrone. Better projection gives higher effective pixel density across the board. Vertical and horizontal resolution is around twice that previously available for the face and eyes throughout the capture and display pipeline. A new texturing technique is also employed which picks and applies the best texture for the viewpoint. Additionally, the texturing camera is at eye level, which distortion demonstrated in (Roberts et al., 2015). The Telethrone image generator reconstructs a visual hull then the appropriate textures are selected and blended depending on the viewing angle supplied by the Vicon optical tracking system and associated VRPN software network transmission.

### 3.2 Experiment

The captured subject in Figure 3 was asked to focus their attention away from their body centreline to marked points in the octave either 45 degrees to their right (best case condition) or 66.5 degrees to their right (worst case condition). This seemingly over accurate angle was chosen as it is a clearly seen location in the geometry of the capture system and was selected as the maximum one might expect to deviate from the body centreline in social conversation.



Figure 3: Capture of the subject in the octave system.

This reconstruction system has not previously been tested when projecting onto RPT so repeating an element of the *withyou* experiment in a way which tests the optical

characteristics of the Chromatte cloth provides a firm foundation for further investigation. The captured model was played back through the two projectors seen in Figure 1. The furthest projector was set to display a static frame of the reconstructed subject with them appearing to look directly at the projector, which would be a seat in the full Telethrone system. This provides an image on the surface which scatters 5% of incident light in all directions. This ghostly overlay is visible as an overlay on the other projected images (O'Hare et al, 2016). Walking participants are head tracked as they try to align to the eye gaze of a static reconstructed model that continuously adjusts for their head position. This is similar to methods employed in "Estimating the gaze of a virtuality human" (Roberts et al, 2013) and Telehuman (Kim et al., 2012).

### **3.2.1 Hypotheses**

H1: that the ability of the experimental participants to judge the gaze direction of the reconstructed avatar is statistically similar to comparable experiments in literature in a best case experimental condition. Eyes are rendered in the centre of the head and directed toward the centre of the retro-reflected light cone. No discernible crosstalk from the second projector is visible in this condition.

H2: that the ability of the experimental participants to judge the gaze direction of the reconstructed avatar is significantly worse than the best case condition in a 'worst case' experimental condition. Head, body and eyes are all unaligned with respect to one another and the attention is directed to the edge of the retro-reflected light cone ensuring that cross talk occurs between the projectors.

### **3.2.2 Variables**

The independent variable is how much the confederate deviates their view from their body centreline, alongside how central to the retro-reflected light cone their deflected view is.

In the best case condition the capture gaze, eyes and head are aligned. The reconstructed eye vector is central to the cone of reflected light from the Telethrone surface resulting in a clear image.

In the worst case condition the capture is combined head and eye gaze with the reconstructed eye vector to the edge of the cone of reflected light from the Telethrone surface. This increases the effect of the crosstalk.

The dependent variable is the angle offset from the correct simulated view vector which the participants settle on during the experiment.

### **3.2.3 Participants**

n = 39 participants were recruited for a between subjects experiment with two conditions from Social VR 2016 workshop at MediaCityUK in Salford in line with ethical approval CST 15/03. The first 19 subjects performed the best case condition, while 21 subsequent attendees performed the worst case scenario.

### **3.2.4 Scope**

In the current Telethrone prototype the projector is mounted behind and above the head of the seated onlooker as they view the Telethrone. This position generates a tracked viewpoint update. The whole system can be moved around the floor on wheels to take up a different viewpoint on the Telethrone. The prototype chair with these tracked projectors is currently unwieldy, so for the purposes of the experiment we tested a component where the tracking was decoupled from the projection. This allowed the participants to move unencumbered when finding their position of mutual eye gaze.

This tests the readiness of the reconstruction for the broader Telethron system. The utilisation of a static model, streamed from disk ensures repeatability.

### 3.2.5 Procedure

Each participant wore a tracked hat which presented their location to the renderer and the logging system. They were instructed to stand between the two projectors facing the Telethron. In this location they were presented with a blend of two spatially distinct views onto the Chromatte cloth. They were requested to walk slowly along a line demarked by a barrier (shown in Figure 1). This took them into the projection frustum of the projector displaying the reconstructed and tracked image, which was continuously rendered as perspective correct based upon the VRPN tracking data. The effect was that the participants gradually felt that they are walking more into the head and eye gaze of the projected subject.

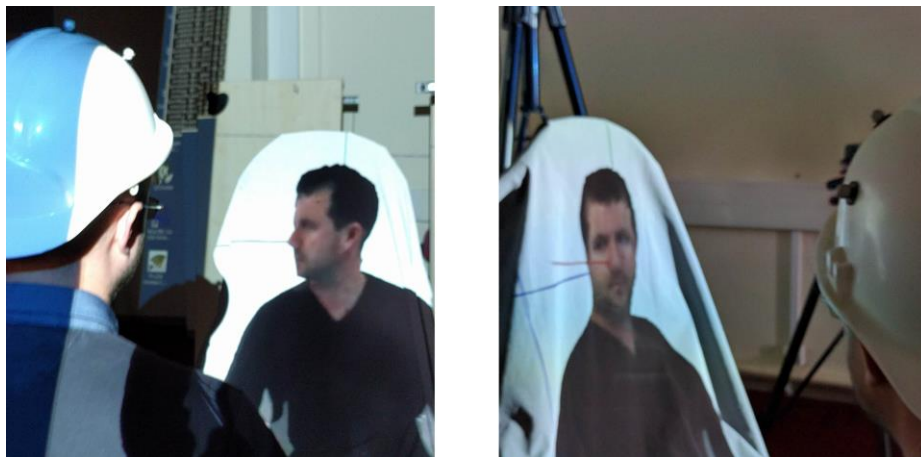


Figure 4: Participant wearing the tracked hat, front view (LHS) and looking to the right (RHS). When they were satisfied that they were in the correct position, they signalled to the experimenter who immediately stopped the data logging to record their position.

## 4 Findings

For the best case condition  $n=18$  the deviation from correct angle is  $Mdn=-0.78$   $M=-0.50$   $StDev=6.28$   $Q1=-4.00$   $Q3=2.46$ . Shapiro-Wilk significance of 0.087 suggests that the data is normally distributed. For the worst case  $n=21$  the deviation from correct angle is  $Mdn=8.27$   $M=7.74$   $StDev=4.48$   $Q1=3.78$   $Q3=10.86$ . Shapiro-Wilk significance of 0.419 indicates that the data is normally distributed.

Wilcoxon ranked signed comparison between conditions shows significant difference between accuracy in best and worst cases  $z=-2.765$   $p=0.006$  while an independent sample t-test likewise shows a difference with  $t=-4.769$   $p=.000$ . Figure 5 shows box plots comparing the two conditions.



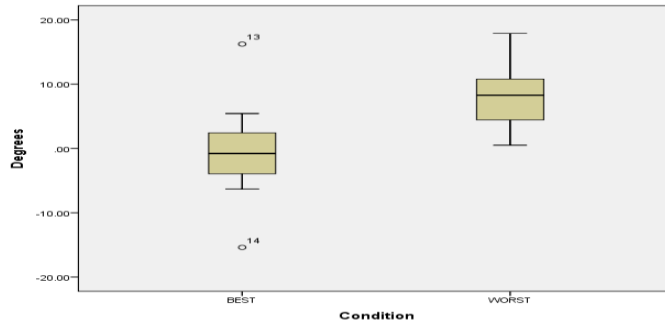


Figure 5: Accuracy of the two conditions compared as a box plot. In the best case condition the median accuracy is very close while worst case consistently undershoots the target.

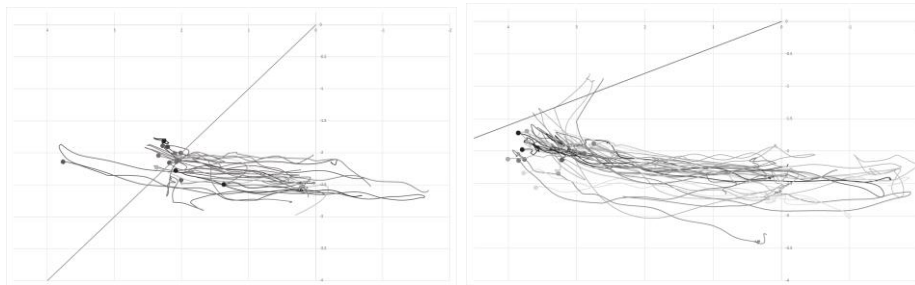


Figure 6: The walking tracks of the 18 participants in 6a and the 21 participants in 6b trying to resolve the gaze deflected 45 and 66.5 degrees from the vertical axis respectively. Participants started on the right hand side and stopped at the dots to the left. Axes are meters. The positions for the worst case condition are all short of the line.



(a) Best case condition (b) Worst case condition

Figure 7: Best case view onto the Telethrone from 7a: 45 degrees from front head and eyes aligned, no image cross talk and 7b: 66.5 degrees from front, head and eyes not aligned, cross talk from the other spatial segment is visible.

## 5 Discussion and Conclusion

All statistics were performed in SPSS comparing within the experiment and against the most similar experiments, Telehuman (Kim et al., 2012) and “Evaluating the gaze of a virtuality human” (Roberts et al., 2013). Differences in the final positions shown in Figures 6a and 6b is potentially explained by the additional difficulty in resolving the eye component of the model. All subjects stopped short of the correct position. Roberts et al used an earlier version of the free viewpoint reconstruction system and projection into an immersive environment.

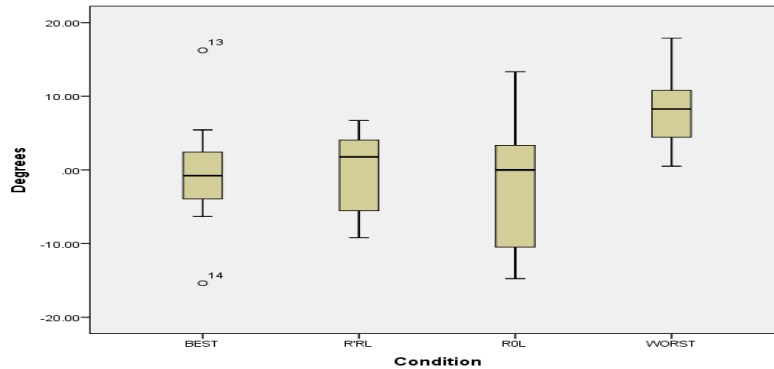


Figure 8, Comparing to “Estimating the gaze of a virtuality human”.

When comparing to their paper ‘estimating the gaze of a virtuality human’ median and standard deviations were used to create box charts from the original data (Figure 8). The rightmost bar labelled WORST is the worst case condition from the experiment while the bar to its immediate left ROL, the most analogous condition from the Roberts paper to WORST. ROL in Roberts et al has eyes aligned forward in the head, and the head turned away from the body. R’RL in Roberts is eyes, head, and body not aligned. The left most bar labelled BEST is for the experimental best condition while the second bar R’RL is for the most analogous condition to BEST in Roberts et al “estimating the gaze of a virtuality human”. The best case condition is analogous to Telehuman in their reported ‘looking at’ scenario where participants had to decide where they were being looked at. Telethron has higher mean accuracy in the best case condition with 0.85 degrees compared to 5.2 degrees. Telehuman has a far better STDEV at .89 compared 6.27. The standard deviation of Telehuman is potentially lower as many more experimental runs were performed. Plots are both positive and negative rather than absolute values as reported in Roberts’ paper to examine directionality of the system. The mean angle of deviation of the Telethron system in the best case condition compares well to Telehuman with a higher mean accuracy, and a standard deviation in the same range as the offset reported in Telehuman. Without access to the detailed Telehuman data it is hard to make detailed comparison. It is notable that the maximum negative deviation for the best case condition is less than the lower quartile deviation from the analogous condition in Roberts et al’s experiment. Better quartile ranges in both best and worst case condition are suspected to be due to the better texturing method and higher resolution. Subject 14 complained that the system did not work for him and he could not seem to resolve the gaze at all. His final

position is an outlier seen on the far left hand side of Figure 12a. He made three passes through the correct eye vector. It is interesting that he was the tallest participant with the tracking data showing him to be approximately 15cm taller than the mean height of 1.55m for the group. While height might be a factor subject 13, the other outlier visible as a dark spot in the centre of Figure 12a, was close to the mean height at 1.50m. The novel contribution of the paper is the presentation of an updated version of the Telethrone system, with tracked 3D video augmenting the spatial segmentation, as well as an experiment which builds toward a deployable system. In the experimental results the best case scenario compares well against “Estimating the gaze of a virtuality human” in the similar ROL condition, and against Telehuman in the ‘Look at’ condition. This satisfies H1 and clears the way for further integration with the Telethrone prototype and a behavioural test with multiple participants. In particular it is suspected that the better texturing approach gave higher accuracy and smaller inter quartile ranges over previous experiments. The worst case scenario satisfies the H2 assertion that the accuracy would be compromised with deflected eye gaze and cross talk. While this data demonstrates less accuracy it is by no means a broken system for the support of directional gaze. Given the properties of the retro-reflective material, the approach is locally scalable with up to 5 onlookers limited by the optical characteristics of the material (only 5 spatial segments are provided). Up to 6 Telethrone per site (6 connected sites of all Telethrone) are theoretically possible provided that multiple tracked and positioned avatars can be projected from a suitable high resolution projector taking in all seats from each Telethrone. Telethrone appears to support small ad-hoc group meetings with dynamic participation and sub groups within a group. The simplicity of the system, its affordability, flexibility, and scalability seem to be appropriate for high traffic social spaces which is a less researched application for telepresence displays.

## 6 References

- Bock, S.W., Dicke, P. and Thier, P., 2008. How precise is gaze following in humans?. *Vision research*, 48(7), pp.946-957.
- Colburn, A., Cohen, M.F. and Drucker, S., 2000. The role of eye gaze in avatar mediated conversational interfaces. *Sketches and Applications, SIGGRAPH'00*.
- Duckworth, T. and Roberts, D.J., 2014. Parallel processing for real-time 3D reconstruction from video streams. *Journal of real-time image processing*, 9(3), pp.427-445.
- Ekman, P., 1993. Facial expression and emotion. *American psychologist*, 48(4), p.384.
- Fagel, S., Bailly, G., Elisei, F. and Lelong, A., 2010. October. On the importance of eye gaze in a face-to-face collaborative task. In *Proceedings of the 3rd international workshop on Affective interaction in natural environments* (pp. 81-86). ACM.
- Fairchild, A.J., Camion, S.P., García, A.S., Wolff, R., Fernando, T. and Roberts, D.J., 2016. A mixed reality telepresence system for collaborative space operation. *IEEE Transactions on Circuits and Systems for Video Technology*.
- Goodwin, C., 2000. Action and embodiment within situated human interaction. *Journal of pragmatics*, 32(10), pp.1489-1522.
- Grau, O., Thomas, G.A., Hilton, A., Kilner, J. and Starck, J., 2007. Mav, A robust free-viewpoint video system for sport scenes. In *3DTV Conference, 2007* (pp. 1-4). IEEE.
- Kim, K., Bolton, J., Girouard, A., Cooperstock, J. and Vertegaal, R., 2012. Mav, TeleHuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In *Proceedings of the SIGCHI Human factors* (pp. 2531-2540). ACM.
- Krauss, R.M., Chen, Y. and Chawla, P., 1996. Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us?. *Advances in experimental social psychology*, 28, pp.389-450.
- Krum, D.M., Suma, E.A. and Bolas, M., 2012. Augmented reality using personal projection and retroreflection. *Personal and Ubiquitous Computing*, 16(1), pp.17-26.

- Lee, M.K. and Takavama, I.. 2011. May. Now, I have a body: Uses and social norms for mobile remote presence in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 33-42). ACM.
- Loomis, J.M., Kelly, J.W., Pusch, M., Bailenson, J.N. and Beall, A.C.. 2008. Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception*, 37(9), pp.1443-1457.
- Maimone, A. and Fuchs, H.. 2011. October. Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on* (pp. 137-146). IEEE.
- Nguyen, D.T. and Cannv, J.. 2007. April. Multiview: improving trust in group video conferencing through spatial faithfulness. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1465-1474). ACM.
- Nguyen, D.T. and Cannv, J.. 2009. April. More than face-to-face: empathic effects of video framing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 423-432). ACM.
- O'Hare, J., Bendall, R. C. A., Rae, J., Thomas, G., Weir, B., & Roberts, D. J. 2016 Is This Seat Taken? Behavioural Analysis of the Telethrone: A Novel Situated Telepresence Display. The Eurographics Association.
- Otsuka, K., Takemae, Y. and Yamato, J.. 2005. October. A probabilistic inference of multiparty-conversation structure based on Markov-switching models of gaze patterns, head directions, and utterances. In *Proceedings of the 7th international conference on Multimodal interfaces* (pp. 191-198). ACM.
- Oyekoya, O., Steptoe, W. and Steed, A.. 2012. May. SphereAvatar: a situated display to represent a remote collaborator. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2551-2560). ACM.
- Pan, Y. and Steed, A.. 2012. October. Preserving gaze direction in teleconferencing using a camera array and a spherical display. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012* (pp. 1-4). IEEE.
- Pan, Y., Steptoe, W., & Steed, A. (2014). Comparing flat and spherical displays in a trust scenario in avatar-mediated interaction.
- Pejsa, T., Kantor, J., Benko, H., Ofek, E. and Wilson, A.. 2016. February. Room2Room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 1716-1725). ACM.
- Roberts, D. A mixed reality telepresence system for collaborative space operation.
- Roberts, D., Wolff, R., Otto, O. and Steed, A.. 2003. Constructing a Gazebo: supporting teamwork in a tightly coupled, distributed task in virtual reality. *Presence*, 12(6), pp.644-657.
- Roberts, D., Duckworth, T., Moore, C., Wolff, R., & O'Hare, J. (2009). Comparing the end to end latency of an immersive collaborative environment and a video conference.
- Roberts, D. J., Fairchild, A. J., Campion, S. P., O'Hare, J., Moore, C. M., Aspin, R., Tecchia, F. (2015). withyou—an experimental end-to-end telepresence system using video-based reconstruction. *Selected Topics in Signal Processing, IEEE Journal of*, 9(3), 562–574.
- Roberts, D. J., Rae, J., Duckworth, T. W., Moore, C. M., & Aspin, R. (2013). Estimating the gaze of a virtuality human.
- Steptoe, W., Wolff, R., Murgia, A., Guimaraes, E., Rae, J., Sharkey, P., Steed, A. (2008). Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative virtual environments.
- Symons, L. A., Lee, K., Cedrone, C. C., & Nishimura, M. (2004). What are you looking at? Acuity for triadic eye gaze. In (Vol. 131, pp. 451–69).
- Tachi, S. (2003). Telexistence and retro-reflective projection technology (rpt).
- Tachi, S., Kawakami, N., Inami, M. & Zaitzu, Y. (2000) Mutual telexistence system using retro-reflective projection technology *International Journal of Humanoid Robotics*
- Vertegaal, R., Veer, G. V. D., & Vons, H. Effects of gaze on multiparty mediated communication..
- Vertegaal, R., Slater, R., Van der Veer, G. and Niiholt, A.. 2001. March. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 301-308). ACM.
- Vishwanath, D., Girshick, A. R., & Banks, M. S. (2005). Why pictures look right when viewed from the wrong place. In (Vol. 8, pp. 1401–1410). Nature Publishing Group.
- Wolff, R., Roberts, D., Murgia, A., Murray, N., Rae, J., Steptoe, W., Sharkey, P. (2008) Communicating Eye Gaze across a Distance without Rooting Participants to the Spot.