Bringing the client and therapist together in virtual reality telepresence exposure therapy

D J Roberts, A J Fairchild, S Campion, A S Garcia, R Wolff

University of Salford, Salford, UK

www.salford.ac.uk/research/health-sciences/research-groups/virtual-reality

ABSTRACT

We present a technology demonstrator of the potential utility of our telepresence approach to supporting tele-therapy, in which client and remote therapist are immersed together. The aim is to demonstrate an approach in which a wide range of non-verbal communication between client and therapist can be contextualised within a shared simulation, even when the therapist is in the clinic and the client at home. The ultimate goal of the approach is to help the therapist to encourage the client to face a simulated threat while keeping them grounded in the safety of the present. The approach is to allow them to use non-verbal communication grounded in both the experience of the exposure and the current surroundings. While this is not new to exposure therapy, the challenges are: 1) to do this not only when the threat is simulated; and 2) when the client and therapist are apart. The technology approach combines immersive collaborative visualisation with free viewpoint 3D video based telepresence. The potential impact is to reduce dropout rate of exposure therapy for resistant clients.

1. INTRODUCTION

Exposure therapy is an effective treatment for phobias and post-traumatic stress disorder (PTSD). Yet it suffers high dropout rates, especially in resistant populations. Drop out can come from lack of engagement, symptoms heightening at outset of therapy, or reluctance of clients to travel to the clinic. Virtual Reality Exposure Therapy (VRET) offers potential to address the first two, and its derivative tele-VRET, the latter. We argue that the typical approach of using Head Mounted Displays (HMD) in VRET and desktop displays in tele-VRET focusses attention on threat and blocks or hinders non-verbal communication (NVC).

We present a demonstrator of a new approach to tele-VRET that addresses this. Within this the therapist and client share a virtual space with the simulated threat in such a way likely to both support a wide range of contextualised NVC and promote a feeling of togetherness. In this, the client faces a life sized 3D video of the therapist moving within a virtual environment that contains emotive stimuli. In the current demonstrator the client is captured through 2D video to allow for easy deployment in the home. Both can judge where the other is looking. The therapist can move between the client and the emotive stimuli or stand to one side and gesture toward or away from it, in this way managing their attention. Our approach allows the therapist to determine if the client is looking at them, fixating on or away from the simulated threat, or following instructions to look toward the real world. A limitation of the current demonstrator is that while the therapist can move freely, gaze estimation will be effected if the client moves off the central line of their camera. A symmetrical system that captured 3D video at each side would allow both to move in any direction without fragmenting spatial context. We also demonstrate how video based reconstruction can be used to rapidly create 3D recordings of actors approaching levels of realism that traditionally take much longer capture and rework.

2. RELATED WORK

VRET has been studied across an extensive range of phobias but perhaps most deeply with Post-Traumatic Stress Disorder (PTSD). Within PTSD, VRET has demonstrated potential efficacy and appears to be more engaging to resistant groups (Gonçalves et al., 2012). Yet drop-out rates, at approaching 40%, remain similar to non-technology based exposure therapy (Gonçalves et al., 2012).

Awareness of both memories and current present-moment experience is seen to facilitate exposure in traumatised individuals (Rothschild, 2003). Conversely, "immersive virtual environments can break the deep everyday connection between where our senses tell us that we are and where we actually are located and whom we are with" (Sanchez-Vives and Slater, 2005). Rothschild explains how the therapist uses non-verbal

communication to detect fixation and bring attention back to the present. Yet VRET typically uses Head HMDs (Gonçalves et al., 2012) that completely block both the present surroundings and therapist from view.

Tele-VRET has been demonstrated but uses desktop interfaces through which avatars representing client and therapist come together in a world, all shrunken to fit within a small monitor. Such systems support little non-verbal communication or feeling of togetherness (Roberts et al., 2015a). People seem to react to life-sized virtual humans as if real, following natural patterns that relate gaze and interpersonal distance (Bailenson et al., 2001). Subtle changes in gaze and posture of virtual humans alters people's comfort (Pertaub et al., 2002). People respond naturally to virtual avatars in distributed immersive collaborative environments (Steed et al., 2005). We have extended such systems to support mutual eye-gaze (Roberts et al., 2009). However, these avatars still do not look like the person whose movements they copy and do not reproduce faithful facial expressions. We have thus developed 3D video telepresence to communicate both what someone looks like and what they are looking at (Roberts et al., 2015a). This technology produces live 3D graphical copies of people, and any items around them, into another space.

Others combined video based reconstruction with an immersive display (Gross et al., 2003) demonstrating how spatial and visual qualities could be better balanced. However, visual and temporal qualities were still some way behind what could be achieved with motion tracked avatars. Since then, visual qualities of video based reconstruction have significantly improved (Grau et al., 2007), (Waizenegger et al., 2011). Recent (Divorra et al., 2010) and current (Steed et al., 2012), (Garcia et al., 2015) funded EU research focuses on spatial telepresence. The potential utility of our approach in collaborative work has been demonstrated within the realm of space science and exploration (Garcia et al., 2015, Roberts et al., 2015b). This technology could be used to join clinic and home.

3. OUR TELEPRESENCE SYSTEM

The ultimate aim of our telepresence system is to situate people from different physical locations into a shared simulated context within which they communicate through a wide range of non-verbal resources. Unlike 2D video based approaches, each can see where the other is looking as they move. This system has been described before (Roberts 2013). Here we summarise what it tries to solve, its approach and current state.

Unlike spoken word, NVC and its use in social interaction is inherently spatial. Just as words link together to provide meaning, so do various non-verbal signals, along with their context. In the natural world, gaze, interpersonal distance and other non-verbal cues of familiarity are linked and used to allow people to manage relationships with each other. Even board room meetings typically start and end with people going up to each other, making eye contact, smiling and sometimes tapping a shoulder or shaking hands. It is these things that grow the trust between people needed for an effective meeting.

Video conferencing supports some of NVC useful in promoting trust and togetherness. Such technology ranges from Skype on a phone to carefully aligned screens and cameras around a table. 2D Video however, loses much of the spatial grounding for NVC. Spatial context can only be accurately determined within the space of the observed, rather than across the spaces of the interactants. While cameras and screens can be aligned to support some approximation of gaze interaction, this only begins to work when people remain in the centre line of the camera. Problems of aligning camera and image of face and the Mona Lisa effect greatly limit this approach and restrict support for relationships between gaze and interpersonal distance. Video conferencing can be said to faithfully communicate visual but not spatial qualities of non-verbal behaviour.

Immersive Collaborative Virtual Environments (ICVE) offer the other extreme, where non-verbal communication between interactants can be situated in a shared virtual context but at the expense of visual faithfulness and many subtleties, such as facial expression. In such a system, people in different displays can move around a shared context together, seeing each other as life sized CGI avatars. We have previously extended ICVE with eye gaze (Roberts VR'09). Such a system theoretically supports the relationship between personal space and eye gaze although this has not been tested with rigour. ICVEs can be said to faithfully communicate spatial but not visual aspects of non-verbal behaviour.

Numerous video based approaches to reconstructing humans have been applied to telepresence. In theory, these should be able to faithfully communicate both visual and spatial qualities of non-verbal communication. However, balancing the two, especially with temporal qualities remains challenging (Roberts, 2013). This is the challenge that our telepresence system is set against. Specifically we want to faithfully communicate both visual and spatial aspects of non-verbal communication to within the limits of their use in non-touch interaction. This means being able to, for example, look someone in the eye and see if they smile as you enter their personal space, perhaps from the side.

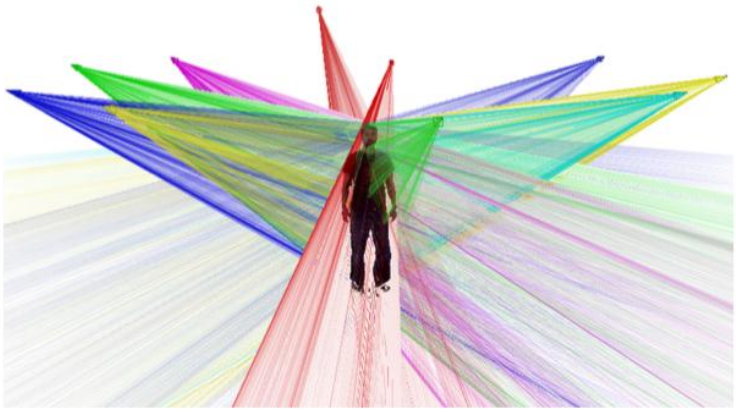

Figure 1. *3D reconstruction of a human in our telepresence system, showing lines from each camera derived from silhouettes.*

Our approach combines real time free viewpoint video with large projection displays. An end to end description of the system is given in (Roberts et al., 2015a). It adopts the video based construction approach of visual hull, using our parallel adaptation (Duckworth and Roberts, 2014) of the EPVH algorithm (Franco and Boyer, 2003). Users stand within an immersive display system and are captured by surrounding cameras, figure 1. Silhouettes from the images are then used to shape carve a form, onto which the original images are textured. This live textured model can then be sent to another immersive display system to be placed within the spatial context of a shared simulation and another user.

We have built many prototype versions that between them demonstrate that all the fundamental requirements are achievable with our approach. However, we have not yet built a single version that fully meets all. At this point in time, we are able to build demonstrators of principle and undertake perceptual experimentation such as (Roberts et al., 2013). However, we have yet to build complete an end-to-end symmetrical system that would demonstrate a sufficient balance of visual, spatial and temporal interaction to support meaningful behavioural experimentation. This paper presents a novel demonstrator.

4. DEMONSTRATION OF THIS SYSTEM APPLIED TO VIRTUAL REALITY TELEPRESENCE EXPOSURE THERAPY

We now describe: the problem we are trying to solve, our general approach, an example scenario, the technology set up, and the limitations.

The problem that we are trying to solve is managing the emotional distance to threat while: 1) the threat is simulated; 2) the client and therapist are in different buildings. The approach we are taking is inspired by Rothschild (Rothschild, 2003) who attempts to mediate a client's awareness of threat and safety of the present, making use of verbal and non-verbal communication.

Our approach is to share a virtual context through large displays while using video based reconstruction to recreate both the therapist and, in this case, the threat. In another case the threat might be completely virtual. The concept is that the therapist can interpret both attention and emotion of the client through non-verbal signals and use non-verbal communication to direct the client's attention and, by doing so, manage emotion their emotion.

In this scenario, the shared virtual environment represents a non-threatening place. The therapy scenario is one of social anxiety. The people in Figures 2 and 3 are authors playing out parts. The three parts being played are: therapist, client and threatening other. In Figure 1 the "client" looks straight at a threat that has just approach through a door. In Figure 3, the "therapist" steps between them and uses gesture and gaze to direct the client's attention to a neutral object, the sofa.

To demonstrate this principle and primary issues we have created an asymmetric system by linking two large displays with two different kinds of mediums (Figure 4 & 5). Asymmetric telepresence systems have been used to demonstrate the impact of differences in VR technology on collaboration (Slater et al., 2000) (Roberts et al., 2003). Our demonstration does not attempt to address every issue but does attempt to demonstrate the key issues and the fundamental qualities of our approach towards addressing them. The client side uses very simple technology that would be relatively straight forward and inexpensive to replicate in the home. The key components are a large flat screen onto which the shared virtual context is displayed and a camera. The therapist side is more complicated but could still be replicated within a clinic without excessive disruption or expense.

Figure 2. *The client in the foreground is approached by a threatening other. The threat is a pre-recorded 3D reconstruction of someone approaching aggressively.*

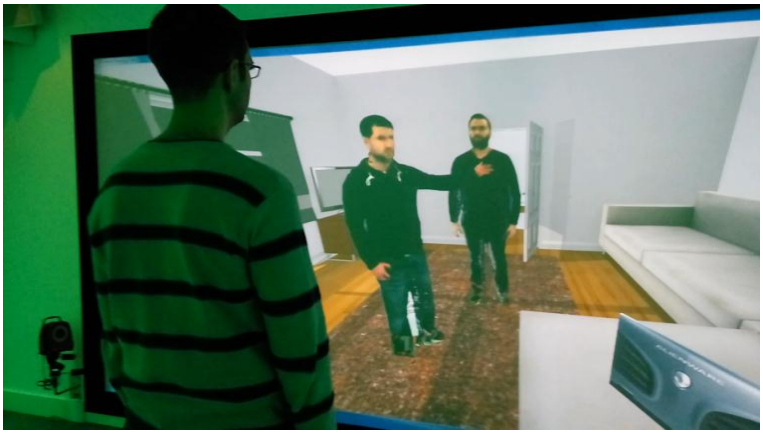

Figure 3. *A mock up of a client fixating on a virtual threat and the therapist stepping in front of it. The therapist (centre) is reconstructed in real time across the telepresence link.*

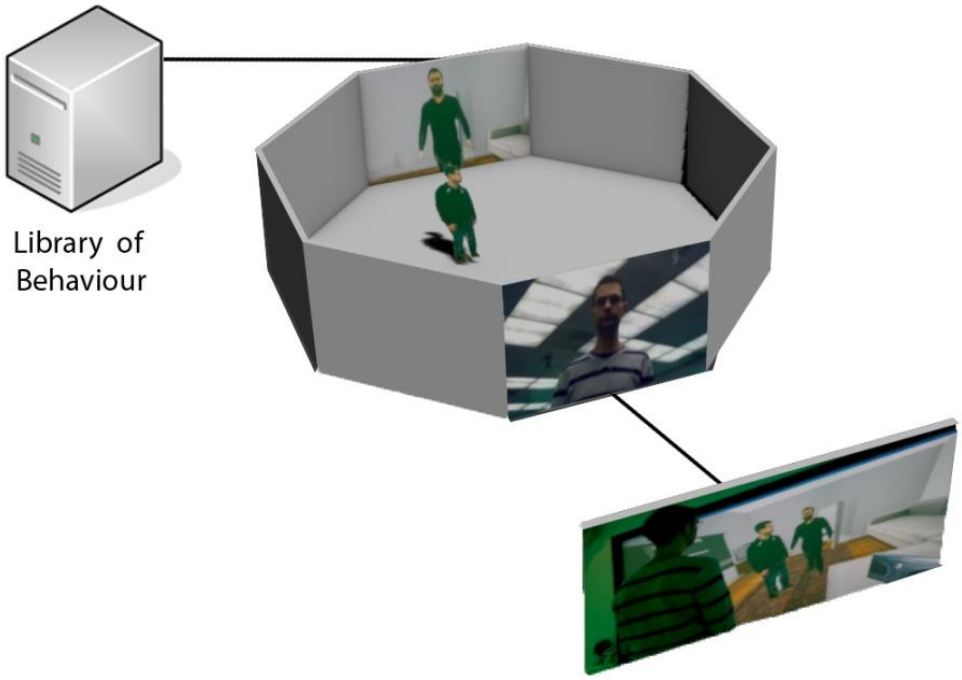

Figure 4. *Diagram of the asymmetric telepresence system built for this demonstrator.*

Proc. 11th Intl Conf. Disability, Virtual Reality & Associated Technologies
Los Angeles, California, USA, 20–22 Sept. 2016

The two fundamental differences are the use of two screens and a ring of cameras. The face on view of the "client" is transmitted via skype to a display wall in front of the "therapist". The rear portion of the partially shared virtual environment is displayed behind the therapist. A ring of cameras looks down at the therapist from above the screens. Each is angled so that while capturing the therapist moving within a portion of the space, neither screen is seen. This allows us to use a simpler and faster method of background segmentation that does not need to account for moving images. Between these two displays, the therapist can look ahead to see the client and behind to see the back of the virtual room the client looks into. The "therapist" appears in the foreground of the partially shared virtual space, as seen by the "client".

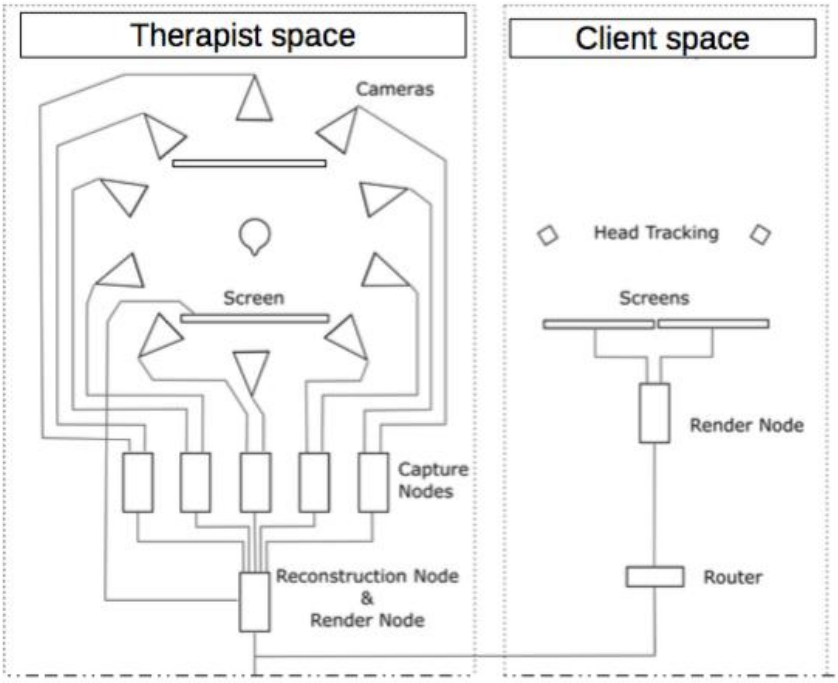

Figure 5. *System Architecture of this demonstrator.*

5. DISCUSSION

An ideal VRET and teleVRET system would allow client and therapist to be immersed together within the simulation without restricting NVC and its contextualisation in any way. Currently this could nearly be supported for co-located VRET using large immersive display systems. All apart from one potentially significant problem, the need to ware 3D glasses that hide the eyes. TeleVRET is more challenging, not least as it requires easily deployable, unobtrusive, easily maintained, low cost solutions for the home end. We have presented a pragmatic approach to this that uses today's technologies in a novel way. It may be many years before technology is available that allows people in different places to seamlessly share each other's spaces. At present we must make a compromise between freedom of movement across the shared simulation against complexity of system and the issues of each complexity.

The demonstrator we have presented is meant to convey concept, pragmatic approach and issues rather than an ultimate system. It demonstrates a range of technologies put together in a pragmatic way. Both simpler and more advanced approaches could be derived from this. The most advanced would allow a full sharing of virtual context in which client and therapist could move around together. The current and simpler versions provide a partial sharing of context which imposes restrictions on movement within the shared space. However, the demonstrated and simpler approaches are far more deployable, affordable and maintainable given current technology. Our approach could be described as partial as it does not allow both parties to move across the full extent of the shared space. However, we felt it was more important to show a practical solution achievable today within people's homes. Until fully immersive stereo can be achieved without stereo, there has to be a compromise between freedom of movement across shared space and ability to determine eye gaze. We were able to support a "therapist" judging the gaze of a "client" and moving between the client and the "simulated threat" he gazed at, and gesturing to a less threatening part of the simulation.

This is not the first time that immersive projection technology has been used in VRET. However, we are unaware of a publication describing its use to support non-verbal togetherness of client and therapist or

communication between them. This is not the first time that immersion and life sized avatars have been used to improve feelings of togetherness or contextualise non-verbal communication. For example, we have previously described our technology approach to faithfully communicating both appearance and attention by combining immersive displays with free viewpoint 3D video based avatars. We have also previously described its application to collaborative working. This is the first time its potential application to exposure therapy has been described.

6. CONCLUSION

The primary contribution of this paper is demonstrating how to support the kind of non-verbal communication used between client and therapist in exposure therapy, firstly when the stimuli, and secondly the other, appear through technology. The methodological contribution is using video based reconstruction in tele-therapy for the first time.

We have demonstrated how video-based reconstruction could potentially be used in virtual reality telepresence exposure therapy. This potential utility is in three parts:

- Making the client feel less alone within a threatening simulation. This is because it supports the range of non-verbal resources used to manage social distance and build feelings of trust and rapport.
- Helping the therapist to manage the client's anxiety and attention. Contextualisation of non-verbal communication is necessary for both.
- Potential utility in creation of visually realistic virtual humans, rapid enough to fit within a course of therapy. Conventional approaches take weeks of authoring.

We have sort to demonstrate concept, pragmatic approach and issues:

- The concept is that the therapist and client can be situated together within the simulation, to allow most of the range of non-verbal communication used by many therapists to manage a client's distance to threat.
- The approach is to combine 3D free-viewpoint video based reconstruction with large display systems and simulated environments.
- The issues are around compromise between complexity and deployability of the system.

Rather than demonstrating an approach that maximises the level of sharing of the simulation, we have demonstrated an asymmetric and pragmatic approach that is less complex, cheaper, more deployable and likely to better retain grounding in the real world. This asymmetry also allows us to demonstrate the impact of technology choices.

The potential impact of this approach is in reducing dropout rates of exposure therapy. This is important as dropout rates of 40% are not uncommon in resistant populations. Furthermore, as symptoms typically increase at the beginning of a course of exposure therapy, clients can dropout with negative health impacts. We argue that by allowing clients to both use virtual reality exposure therapy and work with a therapist at home, reduces the risk of non-attendance to therapy sessions. This could impact not only on success rate of treatment but in reducing costs to health providers through reducing missed appointments. We further argue that allowing the therapist and client to see each other and estimate what the other is looking at, would help to manage the grounding of the client in the safety of the present. This again has potential to reduce dropout rates by reducing the risk of retraumatisation and improving the relationship between client and therapist. While remote therapy can be done with conventional video conferencing and CGI avatars, the levels of non-verbal communication used within a clinical therapy session are not supported. Our approach has the fundamental properties to support them much better. Our demonstrator shows both the issues and the principles of the solution.

Acknowledgements: The authors with to thank Charlie Moritz from Freedom form Torture, Allan Barret from Pennine NHS Care Trust, Warren Mansell from University of Manchester and Linda Durbrow-Marshal from University of Salford for helping us understand the relationship and interaction between client and therapist and what needs to change in VRET to accommodate this. We also wish to thank the technology team at Salford that have helped in the past to develop the telepresence system, including Toby Duckworth, Carl Moore and Rob Aspin and John O'hare.

7. REFERENCES

Bailenson, J N, Blascovich, J, Beall, A C & Loomis, J M (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence,* 10**,** 583-598.

Divorra, O, Civit, J, Zuo, F, Belt, H, Feldmann, I, Chreer, O, Yellin, E, Ijsselsteijn, W, Van Eijk, R & Espinola, D (2010). Towards 3D-aware telepresence: Working on technologies behind the scene. *Proc. ACM CSCW: New Frontiers in Telepresence*.

Duckworth, T & Roberts, D J (2014). Parallel processing for real-time 3D reconstruction from video streams. *Journal of Real-Time Image Processing,* 9**,** 427-445.

Fairchild, A J, Campion, S P, Garcia, A, Wolff, R, Fernando, T & Roberts, D J ---2015. A Mixed Reality Telepresence System for Collaborative Space Operation.

Franco, J-S & Boyer, E. (2003) Published. Exact polyhedral visual hulls. British Machine Vision Conference (BMVC'03). 329-338.

Garcia, A, Roberts, D, Fernando, T, Bar, C, Wolff, R, Dodiya, J, Engelke, W & Gerndt, A. (2015) Published. A collaborative workspace architecture for strengthening collaboration among space scientists. Aerospace Institute of Electrical and Electronics Engineers.

Gonçalves, R, Pedrozo, A L, Coutinho, E S F, Figueira, I & Ventura, P (2012). Efficacy of virtual reality exposure therapy in the treatment of PTSD: a systematic review. *PloS one,* 7**,** e48469.

Grau, O, Hilton, A, Kilner, J, Miller, G, Sargeant, T & Starck, J (2007). A free-viewpoint video system for visualization of sport scenes. *Motion Imaging Journal, SMPTE,* 116**,** 213-219.

Gross, M, Würmlin, S, Naef, M, Lamboray, E, Spagno, C, Kunz, A, Koller-Meier, E, Svoboda, T, Van Gool, L & Lang, S. (2003) Published. blue-c: a spatially immersive display and 3D video portal for telepresence. ACM Transactions on Graphics (TOG). ACM, 819-827.

Pertaub, D-P, Slater, M & Barker, C (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and virtual environments,* 11**,** 68-78.

Roberts, D, Wolff, R, Otto, O & Steed, A (2003). Constructing a Gazebo: supporting teamwork in a tightly coupled, distributed task in virtual reality. *Presence: Teleoperators and Virtual Environments,* 12**,** 644-657.

Roberts, D, Wolff, R, Rae, J, Steed, A, Aspin, R, Mcintyre, M, Pena, A, Oyekoya, O & Steptoe, W. (2009) Published. Communicating eye-gaze across a distance: Comparing an eye-gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together. Virtual reality conference, 2009. VR 2009. IEEE. IEEE, 135-142.

Roberts, D J, Fairchild, A J, Campion, S P, O'hare, J, Moore, C M, Aspin, R, Duckworth, T, Gasparello, P & Tecchia, F (2015a). withyou—An Experimental End-to-End Telepresence System Using Video-Based Reconstruction. *Selected Topics in Signal Processing, IEEE Journal of,* 9**,** 562-574.

Roberts, D J, Garcia, A S, Dodiya, J, Wolff, R, Fairchild, A J & Fernando, T. (2015b) Published. Collaborative telepresence workspaces for space operation and science. Virtual Reality (VR), 2015 IEEE. IEEE, 275-276.

Roberts, D J, Rae, J, Duckworth, T W, Moore, C M & Aspin, R (2013). Estimating the gaze of a virtuality human. *Visualization and Computer Graphics, IEEE Transactions on,* 19**,** 681-690.

Rothschild, B (2003). *The body remembers casebook: Unifying methods and models in the treatment of trauma and PTSD*, WW Norton & Company.

Sanchez-Vives, M V & Slater, M (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience,* 6**,** 332-339.

Slater, M, Sadagic, A, Usoh, M & Schroeder, R (2000). Small-group behavior in a virtual and real environment: A comparative study. *Presence,* 9**,** 37-51.

Steed, A, Roberts, D, Schroeder, R & Heldal, I. (2005) Published. Interaction between users of immersion projection technology systems. HCI International 2005, the 11th International Conference on Human Computer Interaction. 22-27.

Steed, A, Tecchia, F, Bergamasco, M, Slater, M, Steptoe, W, Oyekoya, W, Pece, F, Weyrich, T, Kautz, J & Friedman, D (2012). Beaming: an asymmetric telepresence system. *IEEE computer graphics and applications***,** 10-17.

Waizenegger, W, Feldmann, I & Schreer, O. (2011) Published. Real-time patch sweeping for high-quality depth estimation in 3D video conferencing applications. IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 78710E-78710E-10.