This is a repository copy of *Improving the visual comfort of virtual reality telepresence for robotics*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/154067/

Version: Accepted Version

Proceedings Paper:

Cash, H. and Prescott, T.J. orcid.org/0000-0003-4927-5390 (2019) Improving the visual comfort of virtual reality telepresence for robotics. In: Salichs, M.A., Ge, S.S., Barakova, E.I., Cabibihan, J.-J., Wagner, A.R., Castro-González, A. and He, H., (eds.) Social Robotics. ICSR 2019: International Conference on Social Robotics, 26-29 Nov 2019, Madrid, Spain. Lecture Notes in Computer Science (11876). Springer International Publishing , pp. 697-706. ISBN 9783030358877

https://doi.org/10.1007/978-3-030-35888-4_65

This is a post-peer-review, pre-copyedit version of an article published in ICSR 2019. The final authenticated version is available online at:
http://dx.doi.org/10.1007/978-3-030-35888-4_65

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Improving the Visual Comfort of Virtual Reality Telepresence for Robotics

Harvey Cash, Tony J. Prescott

The University of Sheffield, Department of Computer Science and Sheffield Robotics
`hcash1@sheffield.ac.uk, t.j.prescott@sheffield.ac.uk`
`https://www.sheffield.ac.uk/dcs`

Abstract. Telepresence technologies enable users to exhibit a presence in a remote location, through the use of sensors, networks and robotics. State-of-the-art telepresence research swaps conventional desktop monitors for Virtual Reality (VR) headsets, in order to increase the user's immersion in the remote environment, though often at the cost of increased nausea and oculomotor discomfort. We describe a novel method for telepresence via VR, aimed at improving comfort, by accounting for discrepancies between robot and user head pose. This is achieved through a "decoupled" image projection technique, whereby the user is able to look across captured imagery rendered to a virtual display plane. Evaluated against conventional projection techniques, in a controlled study involving 19 participants, decoupled image projection significantly reduced mean perceived nausea and oculomotor discomfort while also improving immersiveness and the perceived sensation of presence.

Keywords: Robot Telepresence · Virtual Reality · Visual Comfort.

1 Introduction

Telepresence encompasses a broad field of research, characterised by two main technological challenges: allowing users to *perceive* a remote environment, and allowing users to *affect* that remote environment. Modern state-of-the-art telepresence research often employs Virtual Reality (VR) headsets and robotics (see, e.g. Martinez et al. [5]). While the feeling of remote presence is enhanced by the immersive nature of perception through a VR headset, the user's actions are measured by some manner of input method (from keyboards and controllers, through to more sophisticated motion tracking techniques). The obtained information is then used to control mechanical actuators in the remote location. With this technology, a user might hope to perceive and affect remote objects in as intuitive a manner as would be possible were they directly present themselves. Telepresence at this level would have wide reaching applications: from enabling specialists to perform work where they are needed most, regardless of distance (e.g. remote surgery), to enabling workers to operate safely in hazardous conditions as of yet too complex for completely autonomous solutions (e.g. search and rescue), to supporting variable autonomy where human operators take control of otherwise autonomous robot systems for short periods. The latter application could be particularly significant for social robotics, allowing human operators to augment the currently limited interaction capabilities of robots.

As modern telepresence systems seek to incorporate VR headsets, in order to enhance the sensation of being remotely present, they trade away the comfort of conventional desktop monitors. While merely frustrating when observed on a desktop monitor, the technical limitations of telepresence (robot DoF, video frame rate, network latency, etc.) also contribute to nausea and oculomotor discomfort (e.g. eye strain) [4, 2] and reduced presence [6] in VR users (see [8] for a recent review). Increased discomfort effectively limits the length of time a user can spend telepresent. These limitations become especially crippling when considered in the commercial space, where the use of high quality robotics and cameras may be prohibited by cost. Nausea and oculomotor discomfort in VR systems are caused primarily by discrepancies between what the user *should* see, given the position and orientation of their eyes, and what they are *actually* shown by their Head Mounted Display (HMD). We therefore sought to design and develop an image projection technique that could reduce this view discrepancy, circumventing those hardware limitations contributing to it, and thereby improving the visual comfort of VR telepresence.

2 Approach

2.1 Decoupled Image Projection

The discrepancies noted above are most often caused by the limited DoF and range of motion available to the robot, network latency, and frame rate of camera imagery. One strategy to reduce visual discomfort, then, is to circumvent these hardware limitations in software - transforming received camera imagery based on known pose information in order to reduce the discrepancy between head pose and view pose (i.e. position and orientation of perspective shown by the VR headset). Thus, a *decoupled* image projection technique may be appropriate. Figure 1 illustrates the difference between conventional 'coupled' image projection and the 'decoupled' technique. Rather than feeding camera imagery immediately to a user's VR Headset, images are first rendered to a virtual display plane, which exists some distance in front of the user.

Decoupled image projection transforms the display plane by the head pose of the robot, allowing the user's perspective to move freely across the image, entirely independent of any robot hardware limitations. This is illustrated at times t1 and t2 of Figure 1, wherein the user's vision moves synchronously with the user's head direction, without remaining anchored to the robot imagery. Responsibility for bringing camera imagery to the centre of the user's view therefore falls to the head tracking system. If the user looks to the right but the robot fails to move accordingly, they are nonetheless shown a corresponding change in perspective.

Decoupling increases the orientation-time accuracy of imagery rendered to the headset. In this way, head motion occurring between frames rendered to the

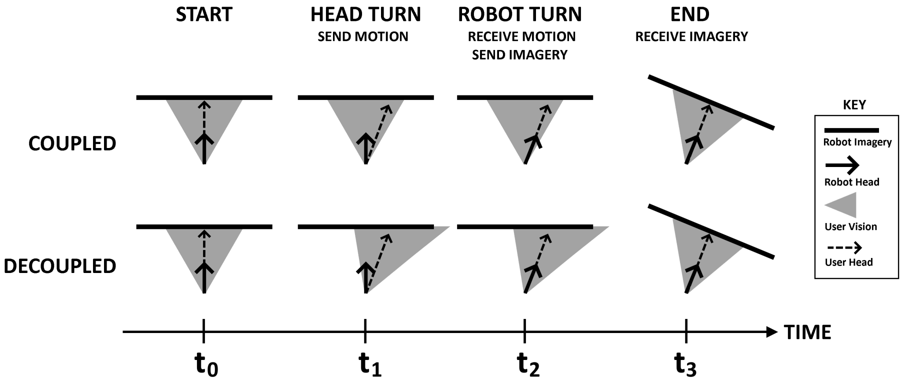

Fig. 1. Comparison of Coupled and Decoupled Image Projection

virtual display is limited only to the refresh rate of the VR Headset itself (typically 90Hz), which is significantly smoother than the 15fps frame rate typically supplied by commercially available robot cameras. This technique has been used by Aykut et al. [1] as a means of tackling network latency through recording wide angle camera imagery in excess of the field of view of the user. While success in that study was evaluated through the proportion of the headset display that remained filled with pixel data, here we evaluated the technique against its capacity for improving the comfort of telepresence, specifically our study addressed two questions: (i) To what degree does decoupled image projection improve the visual comfort of VR telepresence, over conventional coupled projection? (ii) How does the immersiveness of the experience vary between these scenarios?

2.2 Implementation

Figure 2 depicts a robot and client (local desktop computer) separated by a network and the *Decoupled Image Projection Pipeline* we designed to improve user comfort. The position and texture of the virtual display is updated whenever frames are available from the robot's cameras thus depending on robot specifications. Movement of the user's head occurring within the 3D environment, decoupled from the robot's hardware limitations (i.e. frame rate, DoF), executes at a typical refresh rate of 90Hz. Sampling imagery from the Virtual Display and projecting it to the VR Headset thus provides a smooth correspondence between user head orientation and the perspective they are shown.

The developed technique was designed with commercially available robots and robotic avatars in mind. The Consequential Robotics MIRO robot [7] was selected as a testbed in part due to its limitations: (i) the robot has 2 DoF of head movement, yaw and pitch, where each axis has an angular range less than that of a human; (ii) head position varies with orientation differently than in humans, due to differing neck physiology; (iii) MIRO has two stereo-separated

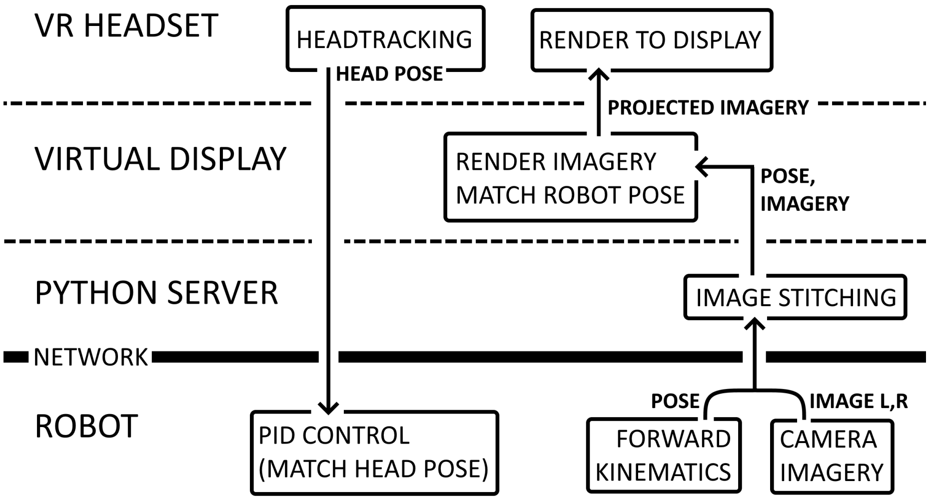

Fig. 2. Decoupled Image Projection Pipeline. The pose of the robot's head is calculated through forward kinematics, and passed together with the left and right eye images over the network to the client. The client performs image stitching to combine the left and right imagery within a local Python server, before passing the robot head pose and combined imagery on to the 3D software environment. A virtual display plane is transformed to match the robot pose, and the combined imagery rendered onto it.

eye cameras, with an effective Inter-Pupillary Distance different to humans; (iv) MIRO cameras are limited to a maximum of 1280x720 pixel resolution at 15 frames per second. These limitations are similar to those of other commercially available social robot platforms, making MIRO suitably representative. Initial development and evaluation has taken place on a simulated robot within Unity3D with the aim of transferring the VR telepresence system to the physical robot platform for further evaluation and testing in the near future.

2.3 Experimental hypothesis

In addition to the issues of discrepant head pose as just discussed, telepresence systems typically suffer from time-lags that can increase discomfort and reduce the experience of immersiveness. In the current study we therefore compared coupled and decoupled projection in two settings, one with zero time-lag (fast) and a second with a 400ms (slow). The experimental hypothesis was that decoupled projection would lead to improved comfort and a greater sense of immersiveness than conventional projection, and that these effects would be more pronounced when the system responsiveness was slow.

3 Methods

3.1 Participants

19 participants were recruited for the evaluation experiment, age range 20-35 (mean 23, s.d. 3.9), gender ratio 13:6 (male:female), all were students or researchers at the University of Sheffield recruited by personal contact. All participants reported having normal vision. All participants signed a consent form and had access to bottled water during the experiment (to minimise any discomfort arising from dehydration), participants were not paid.

3.2 Apparatus

Care was taken to design the simulated robot (see Figure 3) such that interfacing with it presented the same challenges involved in interfacing with a physical robot: including limited DoF, limited ranges of motion, and network latency. A single virtual eye camera, centrally located, was used as a surrogate for panoramic stitched imagery from the MiRo robot's stereo cameras.

In order to ensure that participants attended visually during telepresence immersion, they were presented with a number of Mahjong tiles, and were asked to find the two that match. Participants controlled which tiles are selected using an XBOX controller. Sets of tiles were generated randomly within the constraints that exactly two tiles matched on each trial.

The experiment took place within a bespoke human-robot interaction laboratory at the University of Sheffield. The room was set up such that participants were unable to see the software being used to run the experiment (see Figure 4).

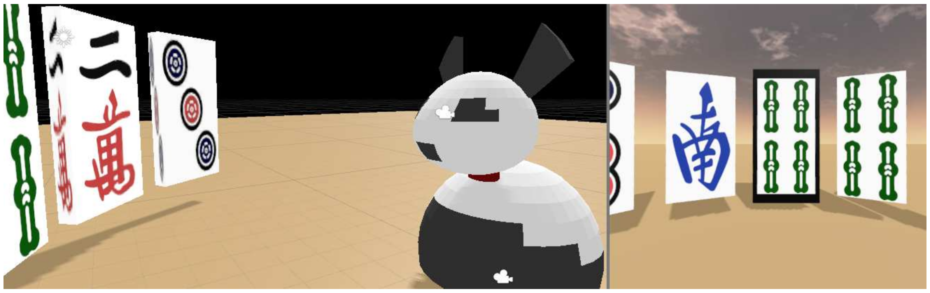

Fig. 3. Simulated MiRo robot and environment in Unity3D (left: robot, right: user perspective)

3.3 Design and procedure

The experiment followed a within-subjects design with four conditions as shown in Table 1. Participants were informed that the study would be to compare the

Fig. 4. Room Setup for Final Evaluation Experiment

relative comfort of different virtual reality experiences, but were not given any information as to exactly what would be varied between tests. Condition order was randomised for each participant, and tile permutations were randomised within each condition. The time spent in each condition (see Table 1) was 3 minutes, and the rest time between conditions was 6 minutes - in order to reduce carried-over discomfort from previous conditions.

Table 1. Table of Experimental Conditions

Condition Name	Vision Method	Simulated Network Delay (Round Trip)
COUPLED-FAST	Coupled	0ms
COUPLED-SLOW	Coupled	400ms
DECOUPLED-FAST	Decoupled	0ms
DECOUPLED-SLOW	Decoupled	400ms

At the start of the experiment, and after each condition, participants were asked to self-assess their experience of nausea and oculomotor discomfort through the Simulator Sickness Questionnaire (SSQ) shown in Table 2, based on the original by R. Kennedy [3]. Each question was answered on a 0-3 linear scale (none to severe) and results combined to produce nausea (sum of items 1, 6-8, 12-16) and oculomotor (sum of 2-5, 9-11) scores at the start of the experiment (baseline) and after each condition (post-test). This allowed adjusted nausea and oculomotor scores to be calculated for each condition by subtracting the baseline from the post-test score. Participants were also asked to rate their feelings of immersion (described as “If fully immersed, you would be so engaged that you forget you are wearing the [VR] headset.”) and presence (described as “If fully present, you would feel as though you were *actually* there with the tiles, in that world.”) as either none, slight, moderate, or high (questions 0a and Ob in Table 2). These were also mapped into a linear score of 0 - 3, and a combined ‘quality’ score of Immersion * Presence calculated (Figure 8).

After all conditions had been completed participants were debriefed and given an opportunity to provide feedback.

Table 2. Table of Immersion and Simulator Sickness Questions

0a. How Immersive was the experience?	0b. How Present did you feel?
1. General Discomfort	2. Fatigue
3. Headache	4. Eye Strain
5. Difficulty Focusing	6. Salivation Increasing
7. Sweating	8. Nausea
9. Difficulty Concentrating	10. Fullness of the Head
11. Blurred Vision	12. Dizziness with eyes open
13. Dizziness with eyes closed	14. Vertigo
15. Stomach Awareness	16. Burping

4 Results

All participants completed the experiment. Figure 5 shows the means and standard deviations for SSQ scores by condition. Across both nausea and oculomotor scores, and for both zero and 400ms delays, there is a clear trend for decoupled image projection to be perceived as more comfortable than coupled.

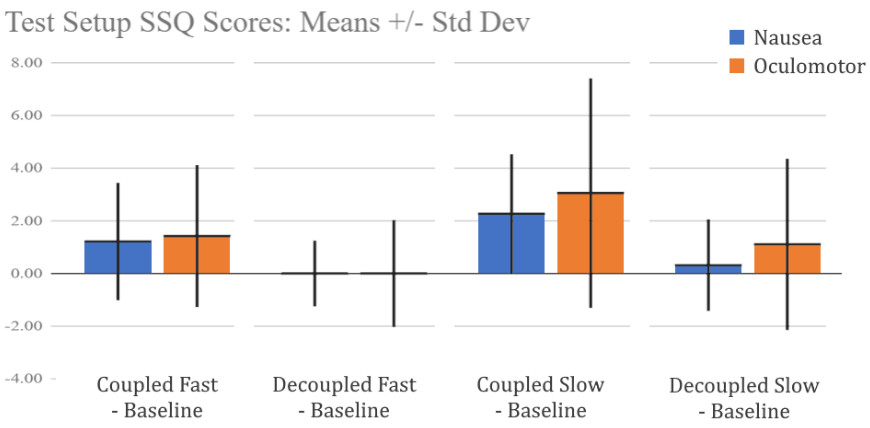

Fig. 5. Nausea and Oculomotor Adjusted SSQ Scores

In order to further compare the coupled and decoupled display techniques, we subtract coupled from decoupled for each participant's nausea and oculomotor scores. If the result is positive, the discomfort caused by coupled can be considered to be higher than that caused by decoupled. These comparisons are shown in Figures 6 and 7 for the fast and slow conditions respectively. Each dot shows the score of one or more participants, as indicated by the participant number. That all but two (of nineteen) participants are in the upper right quadrant indicates that this difference is neutral or positive for both oculomotor and nausea and therefore that the decoupled condition was largely experienced as more comfortable than coupled.

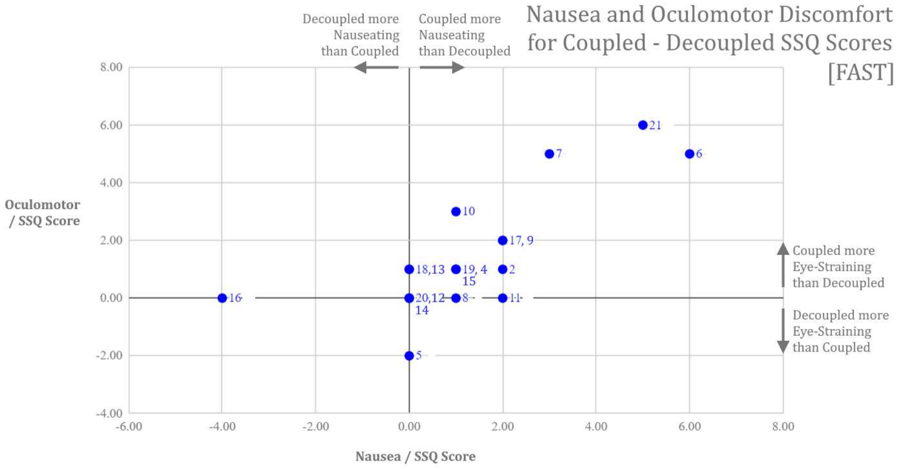

Fig. 6. Zero Delay Coupled minus Decoupled Comparison

Two-Way Repeated Measures ANOVAs were conducted for the dependent variables of per-condition nausea and oculomotor score (adjusted by initial baselines) and analysed for the effects of image projection technique, latency, and their interaction. Results indicate that decoupled image projection significantly decreased oculomotor discomfort ($F(1,18) = 23.31$, $p < 0.000$, $\eta_p^2 = 0.564$) and nausea ($F(1,18) = 18.09$, $p < 0.000$, $\eta_p^2 = 0.501$) compared to coupled. There was also an effect of latency (fast better than slow) for oculomotor ($F(1,18) = 5.48$, $p = 0.031$, $\eta_p^2 = 0.233$) though this failed to reach significance for nausea ($F(1,18) = 3.29$, $p = 0.086$, $\eta_p^2 = 0.155$). The interaction between image projection technique and latency was not found to be significant in either case (oculomotor: $F(1,18) = 0.43$, $p = 0.521$, $\eta_p^2 = 0.023$; nausea: $F(1,18) = 1.72$, $p = 0.206$, $\eta_p^2 = 0.087$). This indicates that the effect of decoupled as an improvement over coupled may be relatively independent of time delay in relation to these measures.

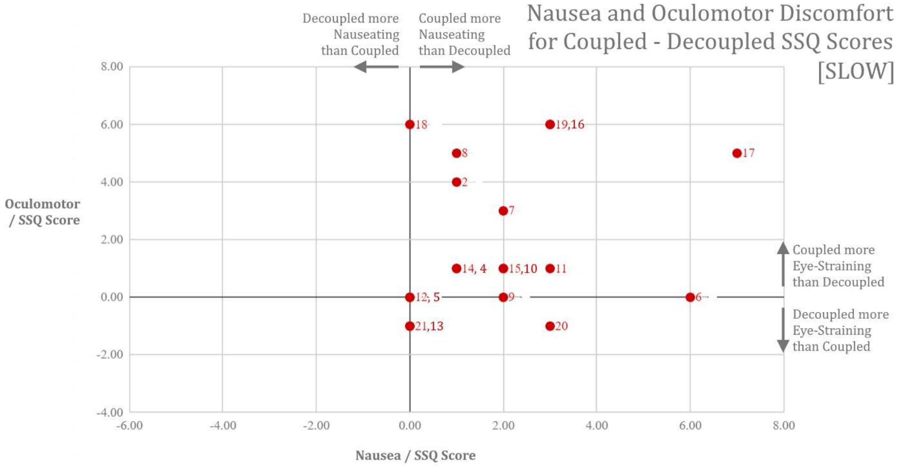

Fig. 7. 400ms Delay Coupled minus Decoupled Comparison

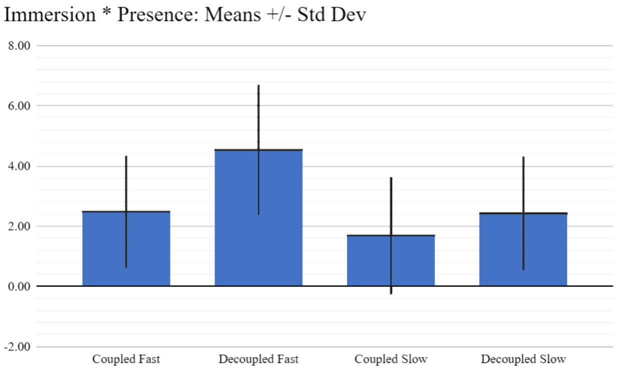

Fig. 8. Combined Immersion and Presence Scores

Figure 8 shows the combined Immersion * Presence score by condition. In both latency conditions the decoupled system was, on average, rated as more immersive than the coupled one. Two way repeated measures ANOVA showed significant main effects due to both technique (F(1,18) = 8.39, p = 0.01, $\eta_p^2 = 0.318$) and latency (F(1,18) = 15.76, p = 0.001, $\eta_p^2 = 0.467$) and a significant technique*latency interaction (F(1,18) = 5.48, p = 0.031, $\eta_p^2 = 0.233$).

From feedback collected during and following the experiment, many participants reported feeling that their field of view had been constricted in the decoupled/slow condition, and that they felt distracted by the movement of the virtual display. We suggest that this perceived distraction, made worse by increased network latency, could reduce oculomotor comfort.

5 Conclusions and Further Work

At both fast (0ms) and slow (400ms round trip) network latencies, decoupled image projection was shown, on average, to be less nauseating and more visually comfortable than coupled. Our results also show that decoupled image projection improves the perceived quality of telepresence over coupled image projection. The partial eta squared (η_p^2) values in the range 0.3 to 0.5 show these to be strong effects similar in impact to having a significant delay in network latency. Future studies might usefully include objective measures of nausea and oculomotor discomfort (rather than subjective ratings) since subjective self-assessment of comfort tends to be highly variable. Participant feedback and oculomotor discomfort scores suggest that limiting the user's field of view by cropping or blurring the edges of the display plane, such that its orientation and motion are less noticeable, could also increase comfort and immersion.

The design and implementation of the decoupled image projection technique proceeded with the limitations of commercially available robotics in mind at all times. We therefore consider that the developed technique is both generalisable (not limited to the specifications of any given robot) and suitable for broad application in future research and in applications of robot telepresence.

References

[1] T. Aykut et al. "A stereoscopic vision system with delay compensation for 360 [degree] remote reality." In: *Proceedings of the on Thematic Workshops of ACM Multimedia*. Oct. 2017. DOI: `10.1145/3126686.3126751`.

[2] E.M.Kolasinski. *Simulator sickness in virtual environments (technical report 1027)*. Tech. rep. U.S. Army Research Inst. - Behavioural and Social Sci, 1995.

[3] R. S. Kennedy et al. "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness". In: *International Journal of Aviation Psychology* 3 (1993), pp. 203–220.

[4] L. Marc et al. "Visual discomfort and visual fatigue of stereoscopic displays: a review." In: *Journal of Imaging Sci. and Tech.* 53 (May 2009),

[5] U. Martinez-Hernandez, L. W. Boorman, and T. J. Prescott. "Telepresence: immersion with the iCub humanoid robot and the oculus rift." In: *Conf. on Biomimetic and Biohybrid Systems*. July 2017. DOI: `10.1007/978-3-319-22979-9_46`.

[6] M.Slater and M.Usoh. *An experimental exploration of presence in virtual environments*. Tech. rep. 689. University College London, Department of Computer Science, 1993, pp. 3–4.

[7] Consequential Robotics. *MiRo Project*. URL: `http://consequentialrobotics.com/miro/` (visited on 12/02/2018).

[8] S. Weech and and M. Barnett-Cowan S. Kenny. "Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review". In: *Frontiers in Psychology* 10 (2019), p. 158. DOI: `10.3389/fpsyg.2019.00158`. URL: `https://www.frontiersin.org/article/10.3389/fpsyg.2019.00158`.