Proceedings Paper:

Camilleri, D. and Prescott, T. orcid.org/0000-0003-4927-5390 (2018) Living in a machine: Experiencing the world through a robotic avatar. In: Vouloutsi, V., Halloy, J., Mura, A., Mangan, M., Lepora, N., Prescott, T. and Verschure, P., (eds.) Biomimetic and Biohybrid Systems. 7th International Conference, Living Machines 2018, 17-20 Jul 2018, Paris, France. Lecture Notes in Computer Science, 10928 . Springer Verlag , pp. 64-72. ISBN 9783319959719

https://doi.org/10.1007/978-3-319-95972-6_8

The final authenticated version is available online at https://doi.org/10.1007/978-3-319-95972-6_8

Living in a Machine : Experiencing the world through a robotic avatar

Daniel Camilleri and Tony Prescott

Computer Science Department, University of Sheffield,
Western Bank, Sheffield, United Kingdom
d.camilleri@sheffield.ac.uk
http://www.sheffield.ac.uk

Abstract. *Telepresence has become a main focus of research fuelled by the technological advancements in virtual reality hardware. In this paper we approach telepresence as a collection of sensory modalities and identify what the current state of the art allows with respect to user devices, robotic hardware and the mapping from one to the other.*

Keywords: Telepresence, Teleoperation, Virtual Reality

1 Introduction

The idea of one day being able to control a physical body over long distances with the possibility of simultaneously enhancing the capabilities of our body, has captured the imagination of researchers since as early as 1980 [1, 2] and even that of the public with films like Surrogates [3]. In essence, telepresence is the teleportation of your sense of presence [4] to a different location than your physical body; with your presence being contained within either an inanimate or an animate body. Moreover, recent advances in commercially available Virtual Reality (VR) hardware provide an ideal technological platform for the development of such a telepresence system.

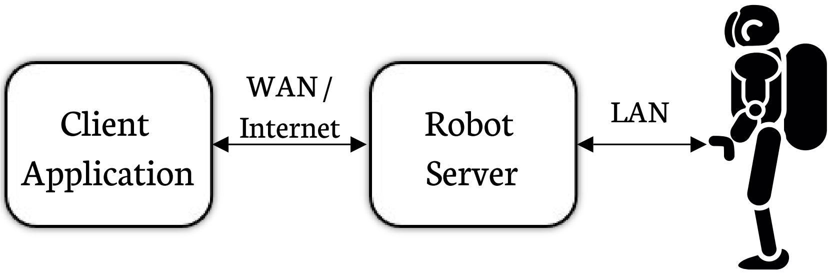

Fig. 1. System Diagram of a Telepresence System

This paper deals with the design of such a system. We are developing a telepresence application that is adaptable to the functionality of multiple VR hardware setups on the user side, as well as being able to control a wide variety of robots. Figure 1 shows the system diagram of the telepresence application with two principal components: the Client Application and the Robot Server.

Our aim for the telepresence system is not only to achieve a high level of immersion by combining multiple sensory input modalities and output modalities but also provide this experience with minimum lag and with a minimum training requirement for the user. In our paper we take a human-centric view of telepresence and for each sensory and output modality present in the human body we ask the following:

1. Which methods are technologically available to record or apply the modality to the human body
2. Which analogues exist within robotic hardware, if any
3. What sensory mapping is required from robot to client and vice versa to retain the experience of being a robot whilst allowing for natural user interaction and control

Following the practical considerations outlined for each modality in the following section, we describe the current state of our system followed by our vision for its future development.

2 Components of a Telepresence System

2.1 Vision

Vision is one of the most important sensory input modalities in the human body and thus places stringent requirements on the quality of the visual apparatus used in VR applications. The main consideration for comfortable use is the frame rate. Low frame rates in VR result in discomfort and motion sickness as vision and head motion become decoupled at frame rates below at minimum 40 frames per second (FPS) with the recommended frame rate being 90fps [5]. Furthermore the field of view (FOV) of vision is also an important component that significantly impacts user immersion [6]. Peak stereo immersion for the human eyes is at around 120°horizontal FOV and 135°vertical FOV.

The analogue to vision in robots is cameras. However different robot systems have different camera arrangements, camera types (RGB/Depth/RGBD), framerates and often times multiple cameras per robot. Therefore in mapping the visual input of the robot to that of a human, the primary consideration, as with the user application, is that the frame rate must not fall below the 25fps threshold. The second consideration is that of camera arrangement.

In the case of stereo cameras with an adjustable vergence such as those in the iCub robot [7], one can map these cameras directly to the left and right eye as long as the vergence of the cameras is soft controlled to keep the object in the middle of the camera FOV, in focus, independent of distance.

Conversely if vergence control for stereo cameras is not possible or in the case of arbitrary RGB camera arrangements, the best immersion is achieved via stitching all the cameras together and providing this image to both eyes simultaneously. In this case it is important to equalise the colour responses of the cameras to provide a unified experience of all the camera inputs. This approach however removes any depth perception but provides better immersion when stereo cameras with vergence control are not available. Furthermore, by taking into consideration the respective translation, orientation and scaling of all cameras with respect to the head joint of the robot, one can present a scaled, stitched image to the user that is at the correct scale. Applying the correct scaling in turn provides a substitute mode of depth perception through previous experiences. The disadvantage of scaling images to correspond with user experience however, is that the user's field of view could be sparsely filled which reduces immersion.

Finally in the case of Depth or RGBD cameras, the best approach is to visualise the point cloud to scale within the user application providing the best possible perception of depth. This can however be computationally expensive so care must be taken not to drastically affect the frame rate.

2.2 Audition

Audition is another one of the primary senses. This can be stimulated by stereo headphones and the design of its experience has 2 factors: continuity and lag. Intermittent audio results in a loss of immersion but even worse is the presence of lag with respect to the visual input. The human brain allows up to 150ms of lag for audio with respect to vision but can only tolerate a lag of 30ms for vision with respect to audio. [8, 9] Thus care must be taken in designing the throughput lag of the two systems.

On the robot side, similar to vision, a robot can also have multiple microphone inputs rather than a more simple binaural input. Thus in order to provide the complete range of auditory inputs to the user, one needs to also get the translation and orientation of the microphones with respect to the head and then compute the binaural equivalent of the spatialised auditory input.

2.3 Gustation and Olfaction

In the user space there is research being undertaken for the emulation of olfaction and gustation however with the exception of a single consumer device for a limited olfactory sense [10] the availability and usability of these devices is still poor. This turns out not to be an issue because the state of the technology for robots is also very far off and while specialised devices exist for olfaction [11, 12] and gustation [13, 14], their adoption in humanoid robots is virtually non-existent. Furthermore the impact of these senses on telepresence immersion is of yet unknown.

2.4 Somatosensation

Somatosensation is the last of the primary senses and just like gustation and olfaction, the artificial stimulation of touch is still an emerging technology with available commercial products using vibrating motors [15–17] to replicate the sensation of touch. These products are however limited in their resolution and coverage, thus the second generation tactile simulation products [18, 19] are using micro-fluidics to overcome this limitation and provide a much higher tactile resolution, possibly allowing for the sensation of texture, as well as higher body coverage.

The importance of somatosensation as an input modality cannot be overstated in the pursuit of high levels of immersion and minimal user training. This is because together with vision, studies regarding the rubber hand illusion [20, 21] and the Pinocchio illusion [22] have shown that vision and touch combined, significantly accelerate the construction of and adaptation to different body schemas [23, 24]. We hypothesize that this accelerated learning means that users can more naturally control their synthetic body.

On the robot side, somatosensation is oftenly overlooked with very few commercial systems providing a significant level of coverage with the exception of iCub and ... There is however a shift towards providing increased somatosensory coverage in robots such as Pepper[25] and Nao [26].

2.5 Thermoception

Keeping on the subject of the skin, thermoception is an essential sense for human survival because the operation of our bodies is only viable within a restricted range of temperatures. This is however different in the case of robots which can withstand a larger temperature range and is thus an example of surpassing our physical limitations with the use of synthetic avatars. Its importance for telepresence has not yet been demonstrated and thus is possibly low. That being said, most of the companies investigating the use of micro-fluidics for second generation somatosensation are also investigating its use for thermoception with individual micro-fluidic circuits being able to heat up or cool down. [18, 19]

2.6 Proprioception

Compared with most other sensory modalities, proprioception is an interoceptive rather than an exteroceptive sense. This means that the feeling is contained within the body and cannot therefore be externally emulated. Thus in the user space this sense can only be replicated through technologies such as hand tracking in order to preserve the mapping learnt for hand-eye coordination. This is another crucial piece of the immersion puzzle, as without the preservation of hand-eye coordination, the experience feels unnatural and quickly discourages the user. This is evident from the incredible jump in immersion available with the latest generation VR devices that allow for hand control which preserves the visual-proprioceptive loop.

Conservation of this sense between the user and the robot space is however complex because of three principal factors. These are: different body scales, different joint configurations and improper visual scaling between the robot and the user.

2.7 Force Perception

Complementary to proprioception, the perception of force or force feedback is another aspect that ties in with proprioception and somatosensation. Consider a user that only has somatosensory feedback from the robot. In the presence of an obstacle, the robot's hand stops however that of the user is still free thus breaking the visual-proprioceptive loop. Thus the perception of force is also very important for a highly immersive telepresence application. Solutions for applying force-feedback in the user space resort to the use of actuated exoskeletons [27] with varying degrees of complexity.

On the other hand, force perception within the robot space can be acquired via the motor torque readings at the various joints after multiplying with the respective robot link lengths.

2.8 Equilibrioception

This sense, much like proprioception, is also interoceptive and thus difficult to trick without moving the whole body. Different solutions for the user space exist [28–30] but none of which allow for a standing experience except for the one being currently developed by HaptX [19] which is akin to a hoisted exoskeleton.

Equilibrioception in the robot space is provided via a gyroscope and accelerometer which indicate the current orientation of the robot with respect to gravity. This sense is present in all robots however its usefulness for telepresence is low except in the case of the teleoperation of bipedal robots.

2.9 Nociception

Starting from the robot space, the sensation of pain can be engineered by assigning different failure modes of the robot to localised sensations of pain. In the user space, pain cannot be simulated because it is another interoceptive sense and furthermore some may argue that the simulation of pain is unethical. However, setting aside the ethical implications, low levels of nociception would provide the user with a means of adapting behaviour to limitations arising due to malfunctions within the robot's body.

This could be especially crucial in the case of sensitive telepresence scenarios where a malfunctioning robot is inaccessible and thus a replacement is either impossible such as the case of a robot within a hazardous environment or outside of the time frame of a time sensitive operation such as search and rescue.

2.10 Motor Control

Motor control in the case of telepresence looks at maintaining the visual-proprioceptive loop between the robot and the user and thus is concerned with the same hardware as proprioception in both the user and robot space . The key consideration here is the lag between user movement and the visual feedback of that movement, which can break immersion, if above as little as 30ms. [31]

2.11 Speech

Speech is easy to replicate by streaming microphone input in the user space to a speaker in the robot space. The main consideration here, much like auditory input, is continuity and lag. Furthermore it has been shown that listening to an echo of yourself talking results in an unnatural feeling [32] and thus speech needs to be cancelled out from auditory input [33] in the robot space as is the case with most video conferencing applications.

2.12 Emotions

Emotions are something we can also replicate within the robot space by using colours, sounds or affective faces that allow for the expression of emotion. In the user space, one could either follow the route of Facebook Spaces [34] which assigns combinations of buttons on VR controllers to different emotions or something more involved like reading emotive states from EEG signals [35]. The latter would allow for a more natural expression of emotion by the user as the button combination route is more difficult to learn. However, much like gustation and olfaction the impact of this modality on immersion is yet unclear.

2.13 Approach

In the previous section we have laid out most of the theoretical framework required to build a wholly immersive telepresence application and in this section we will describe our approach and the results we have achieved so far in developing this application.

Client Application Starting with the Client Application, in our approach we make use of the Unity 3D game engine [36] around which we design our user experience. This game engine was chosen because of its versatility in being cross compiled to multiple operating system platforms. As well as for the presence of the Virtual Reality ToolKit (VRTK) [37] addon which allows the client VR application to be run with multiple hardware setups like Oculus Rift, HTC Vive as well as Android Daydream (formerly Cardboard). This allows us to cater the available sensory modalities based on the type of hardware being used. We've also created a C# Yarp [38] plugin for Unity 3D which is used as the communication layer.

Robot Server On the server side we are developing a universal driver in Python that is capable of interfacing with a variety of robots through the same set of functions allowing for research into the effect of robot morphology on the immersion of the user. Furthermore due to Yarp being an open protocol we employ the use of a VPN to protect sensory transmissions over the internet.

2.14 Results

So far of the 10 sensory modalities and 3 output modalities mentioned in Section 2 we have implemented vision, audition, proprioception + motor control, speech and emotion for the Pepper robot and soon for the iCub robot as well.

While we do not have a setup to measure and validate the lag in the different modalities of our application, we do have some preliminary user feedback for the current state of the system. The lag of audition with respect to vision is such that it is comfortable to watch a video via the telepresence application. As such it must be between the -30 and +150ms range. Furthermore the lag between head movement and its visual feedback has been shown to be very responsive. This responsiveness however starts to decrease as locomotion and arm control are added to the mix requiring better implementation and tuning of our motor control system. In the case of speech, the lag in transmission is currently small but has a noticeable echo.

3 Unresolved Questions and Future Work

The implementation in its current state makes use of a mainstream VR setup based on the Oculus Rift Consumer Edition with Touch Controllers [39]. This will be the topic of future work on the client side as we expand our VR hardware setup to include further sensory modalities. Some of the most important items requiring future work are:

1. Developing methods for conserving user hand-eye coordination in the robot
2. Augmenting the current user VR setup with somatosensensation and force-feedback
3. Exploring the mapping of robot somatosensation to that of the user and how this affects immersion
4. Assessing the importance of emotion for telepresence
5. Investigating the difference robot morphology has on the user experience
6. Conducting user studies to analyse the immersion of the full telepresence system and how different sensory modality sets affect the overall experience

4 Conclusion

In summary, this paper has reviewed the current state of the art in virtual reality devices for the user space, the analogous devices that currently exist in the robot space and has also explored the mapping of one to the other with the aim of high levels of immersion and natural use. Furthermore we have outlined our initial approach and results in the pursuit of putting the theory into action and identified areas which require further theoretical and practical research.

References

1. Minsky, M.: Telepresence. (1980)
2. Slater, M., Sanchez-Vives, M.V.: Enhancing our lives with immersive virtual reality. Frontiers in Robotics and AI **3** (2016) 74
3. Mostow, J.: Surrogates - movie starring bruce willis. Movie (2009)
4. Steuer, J.: Defining virtual reality: Dimensions determining telepresence. Journal of communication **42**(4) (1992) 73–93
5. Murray, J.W.: Building Virtual Reality with Unity and Steam Vr. CRC Press (2017)
6. Abrash, M.: What vr could, should, and almost certainly will be within two years. Steam Dev Days, Seattle (2014) 4
7. Metta, G., Sandini, G., Vernon, D., Natale, L., Nori, F.: The icub humanoid robot: an open platform for research in embodied cognition. In: Proceedings of the 8th workshop on performance metrics for intelligent systems, ACM (2008) 50–56
8. Van Wassenhove, V., Grant, K.W., Poeppel, D.: Temporal window of integration in auditory-visual speech perception. Neuropsychologia **45**(3) (2007) 598–607
9. Conrey, B., Pisoni, D.B.: Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. The Journal of the Acoustical Society of America **119**(6) (2006) 4065–4073

10. FEELREAL, I.: Feel Real. https://feelreal.com/ (2014) Accessed: 30-03-2018.
11. Trivino, R., Gaibor, D., Mediavilla, J., Guarnan, A.V.: Challenges to embed an electronic nose on a mobile robot. In: ANDESCON, 2016 IEEE, IEEE (2016) 1–4
12. Zhang, X., Zhang, M., Sun, J., He, C.: Design of a bionic electronic nose for robot. In: Computing, Communication, Control, and Management, 2008. CCCM'08. ISECS International Colloquium on. Volume 2., IEEE (2008) 18–23
13. Latha, R.S., Lakshmi, P.: Electronic tongue: an analytical gustatory tool. Journal of advanced pharmaceutical technology & research **3**(1) (2012) 3
14. Ha, D., Sun, Q., Su, K., Wan, H., Li, H., Xu, N., Sun, F., Zhuang, L., Hu, N., Wang, P.: Recent achievements in electronic tongue and bioelectronic tongue as taste sensors. Sensors and Actuators B: Chemical **207** (2015) 1136–1146
15. bHaptics: bHaptics - Tactsuit. https://www.bhaptics.com/ Accessed: 30-03-2018.
16. VR, N.: Hardlight VR. http://www.hardlightvr.com/ (2017) Accessed: 30-03-2018.
17. Immerz, I.: KOR-FX. http://www.korfx.com/ (2014) Accessed: 30-03-2018.
18. Teslasuit: Teslasuit. https://teslasuit.io/ Accessed: 30-03-2018.
19. HaptX, I.: HaptX Gloves. https://haptx.com/ Accessed: 30-03-2018.
20. Tsakiris, M., Haggard, P.: The rubber hand illusion revisited: visuotactile integration and self-attribution. Journal of Experimental Psychology: Human Perception and Performance **31**(1) (2005) 80
21. Costantini, M., Haggard, P.: The rubber hand illusion: sensitivity and reference frame for body ownership. Consciousness and cognition **16**(2) (2007) 229–240
22. Conson, M., Mazzarella, E., Trojano, L.: Self-touch affects motor imagery: a study on posture interference effect. Experimental brain research **215**(2) (2011) 115
23. Liu, Y., Medina, J.: Influence of the body schema on multisensory integration: Evidence from the mirror box illusion. Scientific Reports **7**(1) (2017) 5060
24. Medina, J., Coslett, H.B.: From maps to form to space: touch and the body schema. Neuropsychologia **48**(3) (2010) 645–654
25. Robotics, S.: Pepper. Softbank Robotics (2016)
26. Robotics, S.: Nao. Last accessed **20** (2017)
27. Frisoli, A., Rocchi, F., Marcheschi, S., Dettori, A., Salsedo, F., Bergamasco, M.: A new force-feedback arm exoskeleton for haptic interaction in virtual environments. In: Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2005. World Haptics 2005. First Joint, IEEE (2005) 195–201
28. Project, M.: MMOne. http://mm-company.com/ Accessed: 30-03-2018.
29. VR, R.: Roto. http://www.rotovr.com/ Accessed: 30-03-2018.
30. Ltd., I.C.: YawVR. https://www.yawvr.com/ Accessed: 30-03-2018.
31. Allison, R.S., Harris, L.R., Jenkin, M., Jasiobedzka, U., Zacher, J.E.: Tolerance of temporal delay in virtual environments. In: Virtual Reality, 2001. Proceedings. IEEE, IEEE (2001) 247–254
32. Kurihara, K., Tsukada, K.: Speechjammer: A system utilizing artificial speech disturbance with delayed auditory feedback. arXiv preprint arXiv:1202.6106 (2012)
33. Stenger, A., Trautmann, L., Rabenstein, R.: Nonlinear acoustic echo cancellation with 2nd order adaptive volterra filters. In: Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on. Volume 2., IEEE (1999) 877–880
34. Facebook: facebook Spaces. https://www.facebook.com/spaces Accessed: 30-03-2018.

35. Ramirez, R., Vamvakousis, Z.: Detecting emotion from eeg signals using the emotive epoc device. In: International Conference on Brain Informatics, Springer (2012) 175–184
36. Technologies, U.: Unity 3D Game Engine. `https://unity3d.com/` Accessed: 30-03-2018.
37. VRTK: VRTK. `https://github.com/thestonefox/VRTK` Accessed: 30-03-2018.
38. Metta, G., Fitzpatrick, P., Natale, L.: Yarp: yet another robot platform. International Journal of Advanced Robotic Systems **3**(1) (2006) 8
39. Oculus, V., et al.: Oculus rift. Available from WWW:¡ http://www. oculusvr. com/rift (2015)