Beaming into the News:
A System for and Case Study of Tele-Immersive Journalism

Sameer Kishore[1], Xavi Navarro[1], Eva Dominguez[2,4], Nonny De La Peña[3,4], Mel Slater[1,5,6*]

1 Event Lab, Faculty of Psychology, University of Barcelona, Spain
2 Department of Communication, Universitat Pompeu Fabra, Barcelona, Spain
3 Annenberg School of Journalism and Communications, University of Southern California, USA.
4 Independent Journalist
5 Department of Computer Science, University College London, UK.
6 Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

*Corresponding author:
Mel Slater
Event Lab,
Universitat de Barcelona, Facultat de Psicologia,
Campus de Mundet - Edifici Teatre,
Passeig de la Vall d'Hebron 171,
08035 Barcelona
Spain
Tel: +34 93 403 9618
Email: melslater@ub.edu

Abstract

We show how a combination of virtual reality and robotics can be used to beam a physical representation of a person to a distant location, and describe an application of this system in the context of journalism. Full body motion capture data of a person is streamed and mapped in real time, onto the limbs of a humanoid robot present at the remote location. A pair of cameras in the robot's 'eyes' stream stereoscopic video back to the HMD worn by the visitor, and a two-way audio connection allows the visitor to talk to people in the remote destination. By fusing the multisensory data of the visitor with the robot, the visitor's 'consciousness' is transformed to the robot's body. This system was used by a journalist to interview a neuroscientist and a chef 900 miles distant, about food for the brain, resulting in an article published in the popular press.

1. Introduction

Ideas originally expressed in science fiction have often later been realized or have at least been an inspiration for research and technological advance. Perhaps the most famous example is the Star Trek 'communicator' as the conceptual forerunner of the cell phone. Here we consider another Star Trek concept - the idea of *beaming* - instantaneously transporting someone from one physical location to another. While it may be impossible to teleport humans or solid matter instantaneously from one place to another a physical simulation can nevertheless be realized by exploiting current ideas in cognitive neuroscience combined with immersive virtual reality, teleoperator systems and robotics. In this paper we describe a system that makes this possible, and some examples and a case study of its use in the context of an application to journalism. Conceptually the work relies on two key ideas - telepresence (or presence), the illusion of being located in a place other than where the physical body is located, and illusory virtual body ownership [1].

The term most closely associated with the illusion of being present at a place that is different from a person's true location is 'telepresence' [2]. A key aspect of this is that there needs to be a representation of the 'visitor' at the remote location. A physical representation, such as a humanoid robot in the remote location that allows the visitor to interact with the people there and perform tasks affords a physical or embodied experience compared, for example, to tele-conferencing. Technologies that allow someone to remotely control a machine or a robot are referred to as teleoperation. The combination of teleoperation and telepresence has given rise to development of several 'telepresence robots' that allow people to interact with colleagues or attend meetings remotely. However, as mentioned in [3], these robots can be thought of simply as video conferencing on wheels.

The premise of our work is that the visitors (see Inset 1) should have the perceptual illusion that their robot representation is their own body. This illusion of body ownership is influenced by several factors (see Inset 2), including appearance of the virtual/remote body, agency, and a first-person-perspective view from the eyes of the body [1]. Thus, to induce this feeling it is important to have a body that is humanoid, and to be able to control it naturally. This can be achieved with real-time motion capture, where the movements of the visitor are mapped to corresponding movements of the remote robot representation [4]. By combining all the aspects discussed above - allowing someone to see a remote location from the perspective of a remote body (a humanoid robot), have the ability to hear, touch, interact with people, and move the remote body naturally in the remote space as if it were their own - we can simulate a type of physical teleportation. In this paper we focus specifically on developing a solution in the context of journalism – where

either the journalist beams to a remote location embodied in a humanoid robot, or an interviewee beams to the location of the journalist.

Journalists have always needed to travel to remote or unusual locations in order to gain insight about a story. Having a "view from the ground," as described by a veteran World War II reporter [5] gives a journalist the crucial authenticity to connect audiences to reports of unfolding scenarios, events or stories. However, at times locations are off-limits because of inherent physical dangers, such as when the violence of a war makes it too risky to deploy there or the conditions are too inhospitable for the human body, as during a major catastrophe such as a volcanic explosion or when exploring the ocean floor or space. The ability to travel to these places embodied in a robot that corresponds to the human shape and which can be driven in parallel with natural biological body movements can extend reporting in any type of condition or environment. Moreover, good journalism also requires an ability to connect with the individuals who are being interviewed. By creating robots with an appearance compatible to the circumstance, including gender, racial and age-appropriateness, the journalist would be able to side-step a common barrier of prejudice based on the instantaneous judgments made about physical and virtual appearance.

2. Background

2.1. Telerobotics and Robotic Embodiment

One of the first of the recent batch of new and relatively low cost telepresence robots is described in [6] called PRoP (Personal Roving Presence). It is a setup with a single video camera and a microphone, mounted on top of a pole that also has an LCD screen, which is attached to a drive base. Since then, many similar robots have been developed and manufactured commercially. Some robots have added features, such as the ability to express non-verbal cues, and have been shown to lead to a better quality of interaction. A case study [3] compared the various telepresence robots currently on the market by placing them in social scenarios and observing how they were perceived by the people interacting with them. Several commercially available telepresence robots allow the user to work remotely, controlling the robot and interacting with colleagues at the workplace by using a desktop-based application. Another telepresence robot, built specifically for the medical field allows elderly people to communicate with healthcare professionals [7].

It is evident that since the origin of this concept, all subsequent advances have been similar to their predecessors in terms of basic shape, control and communication paradigms. The one aspect that is consistent among all such telepresence robots developed to date is the way the robot is used by the person controlling it. In almost all the cases, the person has a desktop screen with either a joystick or a keyboard to control the robot. Although users have reported a feeling

of presence, the key component missing in these types of system is the illusory sensation that the body of the robot is their own.

As is evident in the literature (see Inset 2), it is important to consider the appearance of the body being used to elicit a full-body ownership illusion. Even though the humanoid robot that we have used for beaming has a morphology resembling that of a human it is nevertheless robotic in appearance. However, it has been shown in another study that this humanoid robot can be used to successfully induce the illusion of body ownership [8] and this has also been used successfully in robotic embodiment [9]. However, in [10] it was found that changing perspective affects the body ownership in an android robot, and that it can be induced using either first-person or third-person-perspective.

Affording visitors to see through the 'eyes' of the robot, and providing them with a natural and congruent method of controlling the limbs of a humanoid robot, we increase the likelihood of a high sense of agency and body ownership with respect to the remote robot body, based on research into body ownership and agency in virtual reality. Hence in our system the visitor may not only feel present in the remote location, but can also have the illusion of ownership over the physical body in the remote destination, which they can use to interact with people, move around and manipulate the remote environment as if they were attending with their biological body. Thus, not only does the robot provide a physical representation of the person, it also acts as a surrogate body that is driven by the visitors as if it were their real body.

Inset 1: Beaming

Beaming involves transmitting a digital representation of a person from one place to another, but where they can be physically embodied in the remote place. Applications from various fields have been developed that have exploited this technology, such as acting rehearsals [1,2], teaching music remotely and medical rehabilitation [3]. It has even supported the teleportation of a human into the cage of a rat, allowing real-time interaction between rat and human, each at their own scale [4]. The following terms are used throughout this paper [5]:

- *Visitor*: The person who 'travels' to the remote location.
- *Transporter*: This is the system used to 'beam' the visitor - requiring a high-resolution wide field-of-view and head-tracked head mounted display (HMD), full-body tracking systems and high-end audio devices. This technology is required to capture multisensory data of the visitor, and transmit it to the destination, as well as to digitally display the destination to the visitor.
- *Destination*: The destination is the remote physical location to where the visitor is transported. Here, it is required for the visitor to be represented in some way, for example, as a humanoid robot.
- *Locals*: The people present in the remote destination who can interact with the visitor are referred to as locals. Ideally the locals should not be encumbered by any equipment in order to be able to see and interact with the remote visitor.

[1] J.-M. Normand, B. Spanlang, F. Tecchia, M. Carrozzino, D. Swapp, and M. Slater, "Full Body Acting Rehearsal in a Networked Virtual Environment - A Case Study," *PRESENCE - Teleoperators and Virtual Environments,* vol. 21, pp. 229-243, 2012.
[2] W. Steptoe, J. M. Normand, O. Oyekoya, F. Pece, E. Giannopoulos, F. Tecchia, *et al.*, "Acting in Collaborative Multimodal Mixed Reality Environments," *Presence-Teleoperators and Virtual Environments,* vol. 21, pp. 406-422, 2012.
[3] D. Perez-Marcos, M. Solazzi, W. Steptoe, O. Oyekoya, A. Frisoli, T. Weyrich, *et al.*, "A fully immersive set-up for remote interaction and neurorehabilitation based on virtual body ownership," *Front. Neur.*, vol. 3:110. doi, 2012.
[4] J. M. Normand, M. V. Sanchez-Vives, C. Waechter, E. Giannopoulos, B. Grosswindhager, B. Spanlang, *et al.*, "Beaming into the Rat World: Enabling Real-Time Interaction between Rat and Human Each at Their Own Scale," *PLoS ONE,* vol. 7, p. e48331, 2012.
[5] A. Steed, W. Steptoe, W. Oyekoya, F. Pece, T. Weyrich, J. Kautz, *et al.*, "Beaming: An Asymmetric Telepresence System," *IEEE Computer Graphics and Applications,* vol. 32, pp. 10-17, 2012.

Inset 2: Embodiment and Body-Ownership Illusion

To illustrate this idea of body ownership, consider the work of Petkova and Ehrsson [1], who described an experiment where a mannequin body was used to apparently substitute the real body of the participants. This was achieved by mounting a pair of cameras on top of the mannequin body looking down towards it. These cameras fed a stereo head-mounted display worn by the participants, so that when they looked down towards their real body they would see the mannequin body in its place. This led to participants experiencing the perceptual illusion that the mannequin body was their own, provided that tactile stimulation that was seen to be applied to the mannequin body was synchronously applied to their real body.

Other studies have shown that although the appearance of the body matters, it is not crucial for invoking the illusion [2]. Results have shown that males can successfully have this illusion in virtual reality with respect to a female body, thin men with respect to a fat body, adults to a child body, light-skinned people to a dark-skinned body, and even with respect to a highly asymmetrical body [3]. However, it appears that the body has to have human characteristics for this illusion to occur [4]. The effect of varying visual perspective, visuomotor correlation and agency have all been studied in detail, and it has been shown that a visuomotor correlation has a strong influence on inducing the illusion of ownership [5].

[1] V. I. Petkova and H. H. Ehrsson, "If I Were You : Perceptual Illusion of Body Swapping," *PLoS ONE,* vol. 3, p. e3832, 2008.
[2] A. Maselli and M. Slater, "The building blocks of the full body ownership illusion," *Frontiers in Human Neuroscience,* vol. 7, 2013.
[3] M. Slater and M. Sanchez-Vives, "Transcending the Self in Immersive Virtual Reality," *Computer (Long. Beach. Calif).,* pp. 24–30, 2014.
[4] M. Tsakiris, L. Carpenter, D. James, and A. Fotopoulou, "Hands only illusion: multisensory integration elicits sense of ownership for body parts but not for non-corporeal objects," *Experimental Brain Research,* pp. 1-10, 2009.
[5] E. Kokkinara and M. Slater, "Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion," *Perception,* vol. 43, pp. 43 – 58, 2014.

3. Materials and Methods

3.1. The Transporter

The Transporter is the part of the system that the visitor uses to transport him- or herself to the remote destination. This can be considered from two points of view. The first is what is required to display the remote destination to the visitor. For this purpose, a stereoscopic 3D video feed from two cameras separated by a standard interocular distance at the destination is streamed in real time via the Internet to the head-mounted display (HMD) worn by the visitor. The HMD that has been used for the various applications of this system is the NVIS nVisor SX111, although the Oculus Rift has also been successfully incorporated into the system. The audio captured from the destination is also streamed to the Transporter in real-time and played back through high quality headphones.

The second aspect concerns transferring the visitor's behavior to the humanoid robot representing the visitor at the destination. To provide the visitor with the most natural method of controlling the robot, the limb motions of the visitor are tracked in real-time. The head of the visitor is also tracked with 6 degrees of freedom by an Intersense 900, which is attached to the HMD itself, while the body can be tracked by a variety of full-body motion capture systems. For our system, we have used the Xsens Inertial Motion Capture system most frequently, although other commercially available systems such as Arena Optitrack or Microsoft Kinect can also be used. These tracking systems can be used to capture the position and orientation of each limb of the visitor in real-time. Additionally, they also supply spatial information of the visitor with respect to the world. Once we have this information, it is used as an input for a system that was developed specifically to convert this tracking data and map it to the humanoid robot at the destination, in real time [4]. This system is also what differentiates this setup from other telepresence arrangements, as it facilitates a much more natural method of controlling the robot. Instead of using a desktop screen with a joystick or keyboard, participants have the ability to move the robot's body by moving their real body. The software application that manages all the data exchange and the rendering in the HMD has been programmed in the XVR system [11].

The audio, video and motion capture streams are transferred from the Transporter to the destination using a high-bandwidth Internet connection. Similarly the stereo video stream from the remote cameras, and the audio stream are transferred back to the visitor's Transporter. It is critical to have all these various streams synchronized so that there is no latency or lag in the communication. In terms of network usage, the stereo video streams require the most amount of bandwidth. Thus, the video frames are compressed using the VP8 format prior to streaming. Since it is possible that the head and body of the visitor might be tracked using different tracking systems, at each frame an entire packet is constructed with new values of motion data. Finally, this array of integer values is sent at each frame to update the limb angles of the robot at the destination. The limb values of the robot are updated depending on the maximum refresh rate of

the robot. The intermediate values between the current position of the limb and new set of values are interpolated by the robot's internal processor. Since the rate at which new values are streamed is much higher than the refresh rate of the robot the robot will always work with the latest motion capture data, but the interpolation will always lead to smooth movements. However, this does mean that it might miss small subtle movements in between such updates.

3.2.   The Destination

At the destination, the visitor is represented by a robot through which he or she can interact with the locals in a natural way, and without encumbering the locals with the requirement to use or wear any special equipment. Hence, the essential component of the setup at the destination is the humanoid robot that acts as the substitute body for the visitor. The robot that we have used for the system is the Robothespian, manufactured by Engineered Arts, UK. This is a 180 cm tall humanoid robot, with two legs, a torso, two arms, and a head. The joints of the robot's upper limbs are pneumatic, while the torso and head, each with three degrees of freedom, move with a DC motor. The shoulders have three degrees of freedom, the elbows have one degree of freedom, and the forearm has the ability to rotate along its own axis as well. The wrist, which has one degree of freedom, is left at its default value, as the tracking systems that we used do not provide tracking information of hand rotation. Due to the lack of sensors in the hands of the robot, haptic feedback to the visitor when the robot's hand is touched is not possible. However, the locals can physically interact directly with the robot itself (for example, shake hands), which is not possible using traditional means of remote communication.

To allow the visitor to see from the robot's perspective, two Microsoft HD-3000 webcams are mounted on the robot's forehead. The feed from these two cameras is streamed to the transporter, and rendered in the HMD of the visitor, who therefore can see the destination in stereo. Since the head of the robot is also directly mapped and controlled by the head movements of visitors they can also see the surrogate robot body when looking towards themselves from a first person perspective. Hence, if the visitor looks down, they see the robot body instead of their own, or looking at a mirror in the destination the visitor will see a reflection of the robot body - which due to the real-time motion capture and mapping moves as they move, subject to the restrictions discussed above. The robot also has a built-in speaker and an omnidirectional microphone that are used by a built-in Skype API for two-way audio communication between the transporter and destination. Furthermore, the Skype API also detects incoming audio, which is used for simple lip-syncing by moving the robot's lower jaw based on the amplitude of the incoming sound.

The lower half of the robot is fixed in place, thus, the robot cannot walk on its own. However, to facilitate movements, specifically translation and rotation on the ground, a programmable platform was custom-built for this specific purpose, on which the robot was mounted. Using the method described in [4], new values for the position and orientation of each limb in the upper arm and head are sent to the robot at every frame. Additionally, the torso of the visitor is also

tracked and this information is used to compute the movement of the visitor in 2D space, which is subsequently streamed to the platform concurrently. Figure 1 shows the schema for this system.

Through this setup, the visitor is able to see the remote destination in stereoscopic 3D, and has head-based sensorimotor contingencies since the head of the robot moves in correspondence with the visitor's head moves. The sensorimotor contingencies contribute to provide a rich multisensory experience and to the sense of presence in the remote location [12]. Furthermore, the movement of the arms of the visitor and the robot are congruent, which can also be seen by the visitor through the 'eyes' of the robot, engendering agency.

4. Tele-Immersive Journalism

4.1. Tele-Immersive Journalism Used for News about the system

The system has been used several times for the purposes of tele-immersive journalism, and has been extensively demonstrated in the media. The BBC was the first news channel to cover this, where they carried out an interview using the system itself (URL 1). The journalist was present at University College London (the destination). A Robothespian was located there that represented the remote visitor (the interviewee, Dr Sanchez-Vives) who was physically located in Barcelona, Spain, and was 'beamed' to London, where she was interviewed about the system.

The system was also used by TV3 based in Catalonia and the Spanish Newspaper La Vanguardia, where the relationship between interviewer and interviewee was reversed (URLs 2-3). This time, the two journalists were physically present in Barcelona, Spain, and were 'beamed' to University College London where they controlled the Robothespian, and conducted an interview about the system with one of the researchers located in London. The news article was printed in La Vanguardia while TV3 aired it as part of a science documentary.

4.2. Tele-Immersive Journalism to Report News

While the demonstrations mentioned above were a first of their kind and received widespread news and media coverage, the interviews had always been about the system itself. The first time that this setup was used by a journalist to not just experience or showcase the system, but to actually apply this technology for conducting interviews about other issues was by journalist Nonny de la Peña.

In a session that lasted for about three hours, Nonny de la Peña 'beamed' to Barcelona, Spain from Los Angeles, USA, and conducted two sets of interviews. The first was with a researcher

(Dr Javier Martínez Picado) whose team had recently discovered an important result in HIV research. The second was to conduct a debate amongst three students who were pro-, anti- or neutral about Catalonia's bid for independence from Spain. The debate was led and moderated by her remotely, i.e., as embodied in the robot. This event was broadcast live on Barcelona TV and an impromptu interview was also conducted where a journalist from BTV asked Nonny de la Peña, while embodied in the robot, about future applications of this technology (URL 4).

Nonny de la Peña utilized the system once again, when she beamed from London to Barcelona to conduct an interview with Dr. Perla Kaliman about her research and book regarding types of food and cooking that are good for the brain (Cocina para tu Mente, by Perla Kaliman and Miguel Aguilar, 2014). In this case the interview using this system was published in traditional news media. The article, which focused solely on the substantive issue of food for the brain rather than the system that was used, was published in the newspaper Latino LA (URL 5).

In this case the visitor (the journalist) was physically present at University College London where we had set up a laboratory as a Transporter system. The Transporter was equipped with a high resolution, large field of view stereo NVIS nVisor SX111 HMD. Since this HMD weighs about 1.3 Kg, a long uninterrupted session could be uncomfortable. Therefore, the session was carried out in parts, with regular breaks of 5-10 minutes at the discretion of the visitor. The body of the visitor was tracked in real-time by the Xsens MVN motion capture suit, and an Intersense 900 head tracker was used to track the orientation of the visitor's head. The position and orientation values were processed using the algorithm described in [4]. High-quality headphones and microphone were used to facilitate two-way audio communication.

The destination was the University of Barcelona, Spain. A Robothespian with specifications as described in the Materials section was used as the physical representation of the visitor. The speaker and microphone built in to the robot provided the necessary hardware for the audio communication. Additionally, a mirror was also placed in front of the robot. The setup explained above can be seen in Figure 2 and also shown in video http://youtu.be/ry6dmWB34qI.

4.3.   Reports of the Experience

Six people who took part during the course of the two sessions that included the interview with the HIV researcher and the debate about Catalonia returned a questionnaire. This was based on standard questions related to telepresence to judge the extent to which they felt as if they were in the same physical space as the journalist with whom they were interacting. Additionally, other questions related to Human-Robot Interaction with a humanoid robot were included from the standard NARS (Negative Attitudes Towards Robots Scale) questionnaire. These questions were aimed at retrieving information about how comfortable locals felt while interacting with a robot, and if they perceived the experience to be positive. All questions were scored on a 7-point Likert

Scale, with 1 being total disagreement and 7 being total agreement with the statement. The questions and the scores are shown in Table 1, and Figure 3 shows a summary graph. Three out of the 6 respondents recorded a score of at least 4 in response to how much they felt as if they were interacting with a person rather than a robot (Q1), with 4 people scoring at least 4 that they felt that they were in the same space as the person controlling the robot (Q2) and that they were comfortable in the presence of the robot (Q5). This indicates the degree to which the locals felt that they were having a normal conversation with the person controlling the robot, instead of focusing on the technology that was involved. Five out of the 6 people scored at least 4 regarding the feeling that they would be able to solve tasks together with the robot (Q6). Even though this response could be specific to the robot that was used, it is still a positive trend that shows that people are willing to accept a robot surrogate body for collaboration. In relation to negative aspects only 2 out of the 6 scored at least 4 regarding their fear of making mistakes or breaking something (Q4), and similarly 2 felt threatened by the robot (Q7). As with the previous question, these responses could also be specific to the robot that was used. However, although the robot was humanoid and has a morphology resembling that of a human, it is still very robotic in its appearance. Their perception of this specific robot, or any other model could be based on their previous experiences and acceptance of robots in general, which could have been influenced by factors such as popular media. Finally 4 out of 6 scored at least 4 in agreement with the idea of using the robot for journalism (Q3). This response could be understood as an indication of how useful the locals perceived the interaction to be, as the people who answered the questionnaire had all experienced the system as an application of journalism. No general conclusions can be drawn from this small sample, and the variation between individuals was high, but given that this was the first ever trial of this type of human-human social interaction via a robot, and given the fact that there were no facial expressions (a feature greatly missed and commented on by the HIV researcher interviewed), the results are encouraging for future applications.

The journalist Nonny de la Peña approached the task at hand with the same intention as when doing an interview with her biological body rather than conducting the conversation on the phone or through Skype. The interview questions were prepared prior to the event and included reading background information and exchanging emails as would typically be done in preparation for reporting any piece. On the day of the interview, prior to donning the Xsens motion capture suit and HMD, de la Peña reviewed her research and committed interview questions to memory.

There were constraints for de la Peña in operating the robot body, such as overcorrection from head turns or hand movement and a viewpoint that was much taller than her normal height. She says, "I can only describe the experience as trying to do a sit up for the first time – you have a concept of how to do it but no muscles to actually perform the task. My entire being had to work in a unified effort to occupy and embody a 'second' self so I could conduct the type of interviews I have done over the past twenty years." However, within approximately fifteen minutes of initiating the experience, she began to adopt the robot body as a natural parallel of her biological

body. The connection became so intense that after the first session, de la Peña notes that it took thirty minutes to stop feeling as if her biological body was still driving the robot body. Moreover, in the second reporting session with Dr. Kaliman, de la Peña reports she was more readily able to adapt the stance that allowed her to control the robot's movements utilizing the Xsens suit.

De la Peña also indicates that for several months afterwards when she recalled the experience, her body involuntarily adopted the most comfortable position for matching the robot to her natural body stance. For example, her arms would bend at the elbow, with her hands outstretched ready to wave, shake hands or gesture. Her head would look upright and her back would stiffen in order to more readily walk forward, back or to swivel from left to right. She also reports some strange and distinct connections to her "robot-self." She says, "When I first saw the interview with my robot-self in Barcelona, I was unexpectedly and strangely upset about the viewpoint of the TV crew camera because I was seeing it from the wrong angle! I actually jumped up from my desk and walked away. I had to force myself to sit back down to watch the whole video. I still cannot define for myself why this view made me so feel so uncomfortable." Moreover, she describes watching the BBC video report of Dr. Sanchez-Vives driving an identical robot. De la Peña felt the disturbing sensation as though someone were inside and occupying her own personal body.

Finally, de la Peña notes that her ability to report was enhanced in comparison to conducting interviews with current video-based technologies. Using the robot as an extension of self, de la Peña was able to shake hands, make eye contact or adjust viewpoint to address unique individuals as well as to 'walk over' and see materials provided by an interview subject. She could also act as a director by arranging the locals into the organization and location that she wanted - by pointing directly into the space, and talking to individuals and groups. By using her biological body in a similar way as if she were actually present on scene, de la Peña had a freedom to engage with the interview subject and the other locals in ways not possible with Skype or the telephone.

5. Discussion

We have presented a novel application that uses a combination of state-of-the-art hardware and software combining telepresence and virtual reality for beaming a person to a remote place. This significantly advances beyond existing telepresence robot setups by combining theoretical and practical knowledge from the fields of cognitive science and computer science. This system not only allows a person to immediately interact with people in a remote destination, but it also can invoke a feeling of owning a physical body, thereby giving the illusion of really being there in a physical sense. Additionally, the method of controlling this body is direct - the robot limbs and torso are moved by moving the corresponding parts of the real body. Thus, the robot body can be used for gestures and integrating other forms of non-verbal communication in the experience as

13

well. This substitution of the real body by a surrogate robotic body allows someone to see, move and interact with the environment approximately like they would if they were physically present at the destination.

The main bottleneck in terms of cost and accessibility is the hardware involved. Thanks to advances in the field of gaming hardware with the recent proliferation of low cost and high quality head-mounted displays at consumer prices some of the critical elements are widely available. This not only helps to lower the financial cost, but also allows visitors to hold longer, uninterrupted sessions, since the fatigue would be less with the new generation of lighter HMDs. Another consumer device that can be used to advantage is the Microsoft Kinect. The current version of the device is already capable of full-body real time tracking, and consequently, can be used in our system to drive the robot remotely. One advantage the Kinect has over Xsens is that the Kinect uses markerless tracking. The Xsens, which is based on inertial tracking, requires the visitor to put on a special suit and perform a pre-calibration every time. The Kinect on the other hand uses computer vision techniques to detect and track participants without any additional equipment. This consumer oriented approach is technically already possible with our system by using, for example, an Oculus Rift with a Kinect, albeit with lower quality motion capture. By combining a good quality, reasonably priced and portable HMD with a markerless and portable full-body tracking system the Transporter side of the system is both cheap and portable.

While it is already economically and practically possible to have a portable Transporter system, there is still some way to go before we have low-cost life-sized humanoid robots available commercially and universally. The cost and time required to deliver and set up a life-size humanoid robot such as the Robothespian is still very high. However , the system that we have developed has already been integrated and successfully tested with the Nao robot, manufactured by Aldebaran Robotics. As the field of humanoid robotics gains advances commercially, the feasibility and advantages that this system offers will vastly improve. The algorithm that has been used to map movements of the visitor to the robot is modular and can be extended to allow compatibility with a new robot easily, as long as the structure of the robot is humanoid. This allows us to continuously enhance the system as these robots become available. Furthermore, newer robots are being developed with a plethora of sensing devices and systems that provide rich information regarding the state of the robot. For example, the Nao comes pre-built with touch-based sensors on its hands, which could be used to drive vibrotactile devices attached to the visitor's hands. The sensor could be used to simulate haptic feedback when the visitor uses their hand to touch an object or shake a local's hand. The addition of visual-tactile haptic feedback along with first-person-perspective, visual-motor correlation and head-based sensorimotor contingencies would deliver an even stronger illusion of full-body ownership.

This system aims to provide a vision for the near future, where cities could have docking and charging stations for humanoid robots available for rent. People could instantly teleport to any part of the world by becoming embodied in the humanoid robot that they would select from the

station, and interact with other people, or even other embodied robots without ever having to spend time, money and energy in travelling. This is the ultimate vision of beaming.

Acknowledgements

List of URLs

1. A short video and article on the BBC that talks about the tele-journalism application of the BEAMING system: http://www.bbc.com/news/technology-18017745
2. An article published by Spanish newspaper La Vanguardia that used the system to conduct an interview: http://www.lavanguardia.com/vida/20120520/54296167703/teletransporte-barcelona-londres.html
3. The documentary aired by Catalan TV channel TV3 that showcases the BEAMING system in the context of Telejournalism: http://blogs.tv3.cat/quequicom.php?itemid=48881
4. An impromptu interview of journalist Nonny de la Peña conducted by BTV during a debate that she moderated while using the system, embodied as a humanoid robot: https://www.youtube.com/watch?v=FFaInCXi9Go
5. The article describing the interview using the system published in traditional news media (Latino LA), which focused solely on the substantive issue of food for the brain, rather than the system that was used: http://latinola.com/story.php?story=12654

References

[1]   M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke, "First person experience of body transfer in virtual reality," *PLOS ONE,* vol. 5, pp. e10564-e10564, 2010.
[2]   M. Minsky, "Telepresence," *Omni,* pp. 45-52, 1980.

[3]   K. M. Tsui and M. Desai, "Exploring use cases for telepresence robots," in *6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Lausanne, Switzerland, 2011, pp. 11-18.

[4]   B. Spanlang, X. Navarro, J.-M. Normand, S. Kishore, R. Pizarro, and M. Slater, "Real Time Whole Body Motion Mapping for Avatars and Robots," in *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology - VRST '13*, New York, NY, USA, pp. 175-178.

[5]   M. Gellhorn, *The view from the ground*: Atlantic Monthly Press, 1988.

[6]   E. Paulos and J. Canny, "Social tele-embodiment: Understanding presence," *Autonomous Robots,* pp. 87-95, 2001.

[7]   T.-C. Tsai, Y.-L. Hsu, A.-I. Ma, T. King, and C.-H. Wu, "Developing a telepresence robot for interpersonal communication with the elderly in a home environment," *Telemedicine journal and e-health : the official journal of the American Telemedicine Association,* vol. 13, pp. 407-24, 2007.

[8]   S. Kishore, M. González-Franco, C. Hintemüller, C. Kapeller, C. Guger, M. Slater, *et al.*, "Comparison of SSVEP BCI and Eye Tracking for Controlling a Humanoid Robot in a Social Environment," *PRESENCE: Teleoperators and Virtual Environments,* vol. 23, pp. 242-252, 2014.

[9]   S. Nishio, K. Taura, H. Sumioka, and H. Ishiguro, "Effect of Social Interaction on Body Ownership Transfer to Teleoperated Android," *ROMAN 2013 - The 23rd IEEE International Symposium on Robot and Human Interactive Communication,* pp. 565-570, 2013.

[10]   K. Ogawa, K. Taura, S. Nishio, and H. Ishiguro, "Effect of perspective change in body ownership transfer to teleoperated android robot," *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication,* pp. 1072-1077, 2012.

[11]   F. Tecchia, M. Carrozzino, S. Bacinelli, F. Rossi, D. Vercelli, G. Marino, *et al.*, "A Flexible Framework for Wide‐Spectrum VR Development," *Presence - Teleoperators and virtual Environments,* vol. 19, pp. 302-312, 2010.

[12]   M. Slater, "Place Illusion and Plausibility can lead to realistic behaviour in immersive virtual environments," *Philos Trans R Soc Lond,* vol. 364, pp. 3549-3557, 2009.

Figure 1: An illustration of the BEAMING schema. The visitor is physically at the location of the Transporter, where they are fully body-tracked and are wearing a Head Mounted Display (HMD). The body-tracking information of the visitor is captured and streamed to the Destination over the Internet, where it is mapped on to a humanoid robot. The robot is used by the visitor as their physical surrogate at the Destination. The locals are the people present at the Destination who can interact with the robotic representation of the visitor. Two cameras separated by the standard inter-ocular distance are mounted on the humanoid robot's head that stream video in stereoscopic 3D directly to the HMD donned by the visitor.

Figure 2: The specification of the setup used for the immersive journalism interview. The visitor, journalist Nonny de la Peña, wore an Xsens body tracking suit and an NVIS nVisor SX111 HMD, with an Intersense 900 head tracker. At the destination she was represented by a custom-built Robothespian with two Microsoft HD-3000 webcams separated at a standard interocular distance, manufactured by Engineered Arts. She could view the Destination in stereoscopic 3D via the HMD. The entire communication and exchange of data took place through a regular Internet connection.

Figure 3: Box plot for the questionnaire scores. The thick horizontal lines are the medians, the boxes are the interquartile ranges, and the whiskers are the ranges. The question numbers refer to Table 1.

Table 1: The questionnaire. A score of 1 means "Totally disagree" and 7 means "Totally agree" with each of the statements. 1 – 6 are the six participants that responded to the questionnaire.

Question	1	2	3	4	5	6
Q1 I felt as if I was interacting with a real person, and not a robot.	7	1	5	3	2	4
Q2 I felt as if the person controlling the robot was in the same physical space as me.	5	4	4	3	3	6
Q3 I think it is a good idea to use a robot for journalism purposes.	3	3	4	5	5	4
Q4 I was afraid to make mistakes or break something while interacting with the robot.	1	2	5	1	5	1
Q5 I was comfortable in the presence of the robot.	7	2	5	6	2	5
Q6 I feel I could solve tasks together with this robot.	7	4	5	5	6	2
Q7 I felt threatened by the robot.	1	1	1	1	7	5