You Are (Not) The Robot: Variable Perspective Motion Control of a Social Telepresence Robot

Michael Suguitan
Cornell University
Ithaca, USA
mjs679@cornell.edu

Guy Hoffman
Cornell University
Ithaca, USA
hoffman@cornell.edu

ABSTRACT

COVID-19 has dramatically limited opportunities for in-person human-robot interaction research and shifted focus towards remote technologies such as telepresence robots. Telepresence robots enable rich communication and agency through their physical presence and controllability, but their screen-oriented designs and button-centric controls abstract users away from their own physicality. In this demonstration, we present a telepresence system for remotely controlling a social robot using a smartphone's motion sensors. Users can select between a first-person perspective from the robot's internal camera or a third-person perspective showing the robot's whole body. Users can also record their movements for later playback. This system has applications as an embodied remote communication platform and for crowdsourcing demonstrations of user-crafted robot movements.

KEYWORDS

human-robot interaction; telepresence; motion control; crowdsourcing

ACM Reference Format:
Michael Suguitan and Guy Hoffman. 2021. You Are (Not) The Robot: Variable Perspective Motion Control of a Social Telepresence Robot. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3411763.3451571

1 INTRODUCTION

The ongoing COVID-19 pandemic has affected how human-robot interaction (HRI) research can be safely conducted in accordance with social distancing guidelines. In response to these new constraints, the HRI community has increased interest in systems that enable remote research, such as simulators or telepresence [6, 11, 13]. Telepresence is an extensive subfield within HRI, but many platforms strongly emphasize screen-based communication for serving utilitarian functions. This neglects the importance of the design of the robot's embodiment and movements, physical affordances that are unique to robots compared to non-embodied voice- or text-based agents. Additionally, the interfaces for the telepresence robots are often game-like controllers with virtual joysticks or buttons that abstract away from the physicalities of the robot and user. Mapping the telepresence robot's movements to the user's own body can improve experiential factors such as social and spatial presence and agency [2, 9, 14].

In this interactive demonstration, we present a telepresence system for remotely motion controlling a social robot with a smartphone. Employing motion control engages the physical embodiments of both the user and the robot. We developed the system around Blossom, an open-source social robot that we have been developing in our research group. In line with the ethos of Blossom as an accessible robotics platform, we designed the remote control system to be usable by novice users. We thus chose to develop the controller around smartphones because of their functionality and accessibility. Users control the robot using the phone's orientation sensor and can choose from two perspectives: either a first-person view from a camera located in the robot's head (1PV), or a third-person view from an external camera showing the robot's whole body (3PV). Users can also record movements, save them with a descriptive name, and play back movements created by other users. While control is limited to one user at at time, other users can view the video feed and queue themselves for access to the robot. We designed the system as an interactive communication platform in response to the ongoing social distancing restrictions, but it also has further applications for crowdsourcing demonstrations of emotive robot movements for machine learning models.

2 RELATED WORK

2.1 Affective Telepresence

Many telepresence robots are designed for the pragmatic functionality of exploring a remote environment while displaying a video feed of the remote user, with the screen being the dominating design element. Sirkin and Ju found that enabling motion on a screen-based telepresence robot improved the sense of presence on both ends [15]. The use of physicality is a unique affordance that sets robots apart from static screen- or voice-based agents, and is an important modality for displaying emotive social cues. Goldberg's early telepresence robots were designed for ludic interactions, such as tending to a garden or uncovering treasures in a sandbox [5].

Some researchers have designed robots specifically for promoting social interactions at a distance. Jeong et al. designed the Fribo robot to improve social connectedness across a network of friends' homes. Each robot listens to "living noise" and notifies other robots on the network of a user's actions, such as laundering or cooking [8]. Gomez et al. used the Haru robot for transmitting "robomojis,"

CHI '21 Extended Abstracts, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8095-9/21/05.
https://doi.org/10.1145/3411763.3451571

emojis that are embodied by the robot's motion and affective animations and sounds [6]. The teddy bear Huggable robot enabled remote users to control its gaze and appendages through a web interface [16]. The MeBot telepresence robot features controllable appendages in addition to a screen displaying the remote user [1]. While these platforms afford a further level of engagement beyond static agents, using textual and button inputs circumlocutes the user's own embodiment and capacity to move around in the world.

2.2 Motion and Viewpoint Control

Rather than use text inputs or game controllers as proxies for controlling robots, proprioceptive motion controls provide a more direct translation from the user's embodiment to the robot's embodiment, enhancing their sense of self-location and agency [10]. Free choice between first- and third-person is a common user option in video games and virtual reality, with several works showing that first-person perspectives increase immersion and a sense of body ownership, while third-person offers heightened spatial awareness [3, 4, 7]. Sakashita et al. used a virtual reality system to remotely puppeteer a robot while providing the user with a first-person view [14]. Some works have focused on the accessible motion control afforded by smartphones to provide a more immersive telepresence experience. Ainasoja et al. compared motion- and touch-based interfaces for controlling a Double self-balancing telepresence robot, and found that a hybrid motion-touch interface (motion for left-right steering, touch for forward-reverse) was most preferred [2]. Jonggil et al. implemented smartphone motion control for a mobile camera robot and found that motion controls improved the user's sense of presence, synchronicity with the robot, and understanding of the remote space compared to touch-based controls [9]. These robots used were utilitarian in design and function (roving mobile platforms with screens and cameras), and the user perspectives were constrained to 1PV.

3 TECHNICAL IMPLEMENTATION

We built upon prior works with a motion controlled telepresence platform using a social robot (Figure 1). The users can freely switch between 1PV and 3PV perspectives. In this section we detail the technical implementation of the robot and user interfaces.

3.1 Robot

We used our lab's Blossom robot, an open-source social robot (Figure 1, left) [18]. Blossom features 4 DoFs: yaw, pitch, roll, and vertical translation. Although the robot's DoFs are limited compared to more complex embodiments, the head motion is enough for creating expressive gestures. Having a minimally expressive embodiment also avoids complexity that would steepen the learning curve for new users. For 1PV, we embedded a small USB camera inside the robot's head, in front of one of its ears.

3.2 User Interfaces

To bolster the system's accessibility, we built the application as a mobile browser experience instead of a standalone application. This enabled us to iterate quickly and access a rich library of functionality through APIs while obviating the need for external downloads on the user's device.

We created two interfaces to accommodate the two viewpoints (Figure 1, right). The user can freely select between two viewpoints: 1PV from the camera in the robot's head shown on the phone, and 3PV from the host computer's webcam shown on the desktop browser (3PV). Users access the interface from a public URL (`blossombot.com`). We implemented username and password authorization to prevent unwarranted access to the video streams and for managing the queue of users waiting to control the robot.

3.2.1 *Mobile Interface.*

The mobile interface consists of a video feed showing 1PV and a simple layout of buttons for controlling the robot (Figure 1, center). The layout was inspired by existing controlling and recording applications, such as voice recorders and remote camera applications. Below the video feed is a slider for controlling the height of the robot. Control is toggled with an on-off switch. Users can record and save movements with a large microphone-style recording button. The robot can be reoriented using a calibration button; this resets the robot's yaw orientation relative to the phone's current compass heading. Below the control row are a selector and play button for replaying created movements.

3.2.2 *Desktop Interface.*

The desktop interface consists only of a video screen showing 3PV (Figure 1, right).

3.3 Backend

3.3.1 *Communication.*

The robot is connected to the host computer, which also serves the interface (Figure 1, left). We use `ngrok` to enable communication across the internet from the user to the host computer and robot.

3.3.2 *Motion control.*

We use a kinematic model of the robot to translate the phone's orientation into the angular poses for the robot's head (Figure 2). The controller uses the `DeviceOrientation` API to report motion events. The phone's inertial measurement unit (IMU) records its pose as Tait-Bryan angles about the phone's reference frame. Due to the IMU's inability to track translation, we implement height control with a slider on the interface. In 1PV, the phone and robot axes are aligned as if the phone's camera were looking through the robot's eyes (Figure 2, right). When switching from 1PV to 3PV, the motion is mirrored horizontally to accommodate the front-facing view of the robot, as if the user were facing a physical mirror. In 3PV, yawing or rolling the phone to the left from the user's perspective moves the robot to its right, and vice-versa; this transformation is equivalent to inverting the phone's reference frame (Figure 2, left).

3.3.3 *Video.*

For the video streams, we used `WebRTC`, the standard for online audiovisual communication. `WebRTC` manages the handshaking for broadcasting the video stream to several watchers.

3.4 Preliminary Deployments

We have tested the system with several novice users without prior knowledge of the platform or robotics in general. Even without prior experience, the users were able to quickly learn the control method. The few confusing aspects came from the controls mirroring when switching from 1PV to 3PV, and we found that users often neglected the rolling motion in 1PV.

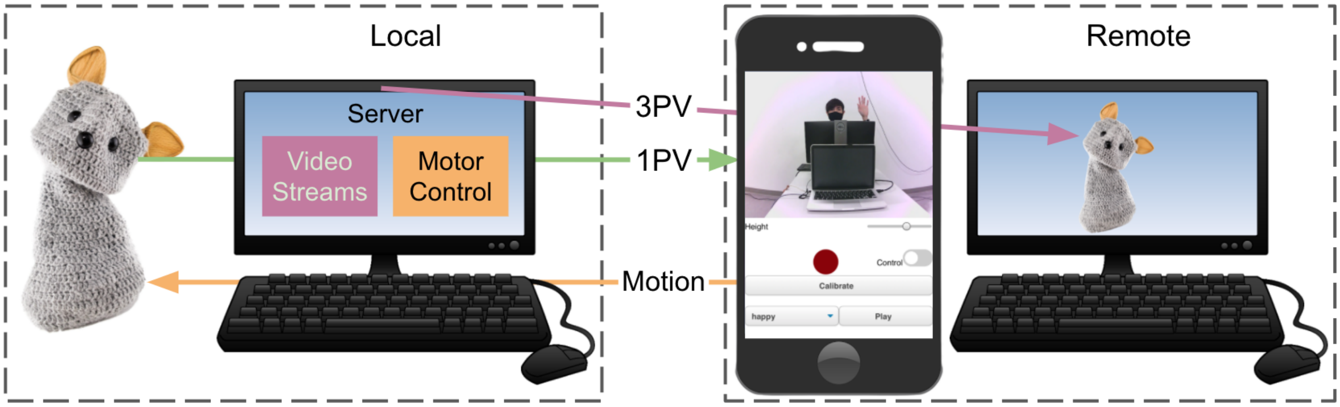

Figure 1: System architecture. The first- (1PV) and third-person views (3PV) from the local robot and desktop (left) are transmitted to the remote phone and desktop (right). The phone's motion data is transmitted to the local computer to control the robot's motors.

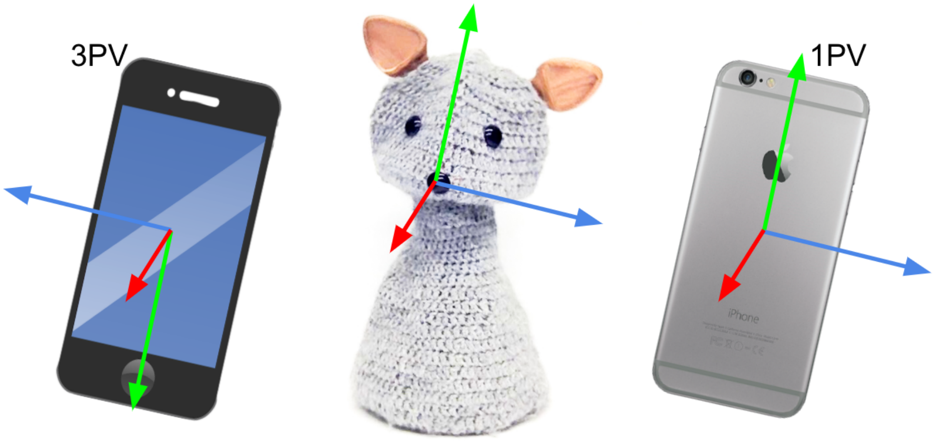

Figure 2: The alignment of the robot's and phone's reference frames. The motion is mirrored in the third-person view (3PV) to accommodate the perspective of looking straight at the robot, which is equivalent to inverting the phone's reference frame.

As with any remote technology, latency adversely affects usability. Compared to video or text communication, the real-time motion control interface induces an expectation of synchronization that exacerbates the latency between the motion and the video. The latency accumulates, potentially locking up the robot until all of the motion data is dumped at once. To attenuate the effects of latency, we implemented a timeout from when the client interface sends its orientation data and when the host computer receives the data packet. A timestamp difference greater than two seconds results in the data being ignored by the host computer. Additionally, to avoid further slowdowns, we found it best to dedicate a separate computer for running the backend while relegating another computer for handling ancillary tasks such as communicating with users.

4 USE CASES

In this section we detail the plans for the live demonstration and other potential use cases of the platform.

4.1 Demonstration

For our demonstration, remote users will control the robot using their smartphones. Users will navigate to the interface's webpage on their phones and log in with provided credentials. There they will see the 1PV video feed from the robot's head and control the robot. Users can record movement sequences and play back movements created by other users. Optionally, users can open the desktop interface on their computers to view the 3PV video feed showing the robot's whole body. To prevent several users fighting for simultaneous control of the robot, we limit the control access to the first user in a queue. The current user must relinquish control to pop themselves from the queue and pass control to the next user.

4.2 Other Use Cases

4.2.1 Crowdsourcing affective robot movement demonstrations. The primary research-facing application of the system is a crowdsourcing platform for obtaining demonstrations of affective robot movements. In prior work, we collected a dataset of user-crafted movements as an input to a generative autoencoder neural network

[17]. We used the neural network to generate synthetic movements samples, and found that they were largely comparable to the user-crafted samples in terms of realism and conveyed emotion. While we were able to collect quality movements from novice users, this method is constrained by physical access to the robot. Mandlekar et al. addressed this location barrier by implementing smartphone teleoperation of a robot arm to enable crowdsourcing of demonstrations for grasping tasks [12]. We are currently deploying this system in a similar capacity, though our application is more affective in nature. Compared to a functional task with objective goals, the subjectivity of our application will likely yield large variation in the resulting dataset.

4.2.2 Accessible embodied remote communication. Blossom could be used as an accessible telepresence robot platform. We designed Blossom to be open-source and reproducible, and several research groups and hobbyists have created their own Blossoms (hardware and software is available at `github.com/hrc2/blossom-public`). The system could be a low-cost telepresence robot that can be augmented by end users (e.g. screens, audio communication).

4.2.3 Generalization for remotely controlling other robots. Other works employed smartphones for robot control, indicating their viability as control interfaces [2, 9, 12]. Developing a wrapper library that enables different robots to be controlled by common hardware (smartphones, keyboards, mice) would alleviate the limited access to physical robots and enable a wider userbase to participate in robotics research.

5 CONCLUSION

We presented a telepresence system for remotely motion controlling a social robot with a smartphone. In designing the controller, we used motion controls to emphasize the user's own physicality to heighten their sense of presence and spatial awareness in the remote location. Our system enables robot accessibility to novice users, in terms of both the minimal hardware required and by enabling users to remotely control a robot even with the ongoing social distancing restrictions. We hope that this work provides an engaging interactive experience and inspires other roboticists to develop systems for enabling remote access for their robots.

REFERENCES

[1] Sigurdur Adalgeirsson and Cynthia Breazeal. 2010. MeBot: A Robotic Platform for Socially Embodied Telepresence. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 15–22. https://doi.org/10.1109/HRI.2010.5453272

[2] Antti E. Ainasoja, Said Pertuz, and Joni-Kristian Kämäräinen. 2019. Smartphone Teleoperation for Self-balancing Telepresence Robots.. In *VISIGRAPP (4: VISAPP)*. 561–568. https://doi.org/10.5220/0007406405610568

[3] Alena Denisova and Paul Cairns. 2015. First Person vs. Third Person Perspective in Digital Games: Do Player Preferences Affect Immersion?. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 145–148. https://doi.org/10.1145/2702123.2702256

[4] Henrique Galvan Debarba, Sidney Bovet, Roy Salomon, Olaf Blanke, Bruno Herbelin, and Ronan Boulic. 2017. Characterizing First and Third Person Viewpoints and Their Alternation for Embodied Interaction in Virtual Reality. *PLOS ONE* 12, 12 (12 2017), 1–19. https://doi.org/10.1371/journal.pone.0190109

[5] Ken Goldberg. 2001. *The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet*. MIT Press.

[6] Randy Gomez, Deborah Szapiro, Luis Merino, Heike Brock, Keisuke Nakamura, and Selma Sabanovic. 2020. Emoji to Robomoji: Exploring Affective Telepresence Through Haru. In *International Conference on Social Robotics*. Springer, 652–663. https://doi.org/10.1007/978-3-030-62056-1_54

[7] Geoffrey Gorisse, Olivier Christmann, Etienne Armand Amato, and Simon Richir. 2017. First- and Third-Person Perspectives in Immersive Virtual Environments: Presence and Performance Analysis of Embodied Users. *Frontiers in Robotics and AI* 4 (2017), 33. https://doi.org/10.3389/frobt.2017.00033

[8] Kwangmin Jeong, Jihyun Sung, Hae-Sung Lee, Aram Kim, Hyemi Kim, Chanmi Park, Yuin Jeong, JeeHang Lee, and Jinwoo Kim. 2018. Fribo: A Social Networking Robot for Increasing Social Connectedness through Sharing Daily Home Activities from Living Noise Data. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL, USA) *(HRI '18)*. Association for Computing Machinery, New York, NY, USA, 114–122. https://doi.org/10.1145/3171221.3171254

[9] Ahn Jonggil and Gerard Jounghyun Kim. 2018. SPRinT: A Mixed Approach to a Hand-Held Robot Interface for Telepresence. *International Journal of Social Robotics* 10, 4 (2018), 537–552. https://doi.org/10.1007/s12369-017-0463-2

[10] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (2012), 373–387. https://doi.org/10.1162/PRES_a_00124

[11] Finley Lau, Deepak Gopinath, and Brenna D. Argall. 2020. A JavaScript Framework for Crowdsourced Human-Robot Interaction Experiments: RemoteHRI. In *Association for the Advancement of Artificial Intelligence*.

[12] Ajay Mandlekar, Yuke Zhu, Animesh Garg, Jonathan Booher, Max Spero, Albert Tung, Julian Gao, John Emmons, Anchit Gupta, Emre Orbay, Silvio Savarese, and Li Fei-Fei. 2018. RoboTurk: A Crowdsourcing Platform for Robotic Skill Learning through Imitation. In *Proceedings of The 2nd Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 87)*. PMLR, 879–893. http://proceedings.mlr.press/v87/mandlekar18a.html

[13] Robin R. Murphy, Vignesh Babu Manjunath Gandudi, and Justin Adams. 2020. Applications of Robots for COVID-19 Response. arXiv:2008.06976 [cs.RO]

[14] Mose Sakashita, Tatsuya Minagawa, Amy Koike, Ippei Suzuki, Keisuke Kawahara, and Yoichi Ochiai. 2017. You as a Puppet: Evaluation of Telepresence User Interface for Puppetry. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 217–228. https://doi.org/10.1145/3126594.3126608

[15] David Sirkin and Wendy Ju. 2012. Consistency in Physical and On-Screen Action Improves Perceptions of Telepresence Robots. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (Boston, Massachusetts, USA) *(HRI '12)*. Association for Computing Machinery, New York, NY, USA, 57–64. https://doi.org/10.1145/2157689.2157699

[16] Walter Dan Stiehl, Jun Ki Lee, Cynthia Breazeal, Marco Nalin, Angelica Morandi, and Alberto Sanna. 2009. The Huggable: A Platform for Research in Robotic Companions for Pediatric Care. In *Proceedings of the 8th International Conference on Interaction Design and Children* (Como, Italy) *(IDC '09)*. Association for Computing Machinery, New York, NY, USA, 317–320. https://doi.org/10.1145/1551788.1551872

[17] Michael Suguitan, Randy Gomez, and Guy Hoffman. 2020. MoveAE: Modifying Affective Robot Movements Using Classifying Variational Autoencoders. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 481–489. https://doi.org/10.1145/3319502.3374807

[18] Michael Suguitan and Guy Hoffman. 2019. Blossom: A Handcrafted Open-Source Robot. *ACM Transactions on Human-Robot Interaction* 8, 1, Article 2 (March 2019), 27 pages. https://doi.org/10.1145/3310356