Commodity Telepresence with the AvaTRINA Nursebot in the ANA Avatar XPRIZE Semifinals

Joao M. C. Marques*, Patrick Naughton*, Yifan Zhu*, Neil Malhotra†, and Kris Hauser*
*Department of Computer Science, University of Illinois, Urbana-Champaign
†VRotors, Inc., Pasadena, CA

I. Introduction

Telepresence [1], or telexistence [2], envisions the use of immersive vision and rich haptic feedback to enable a human operator to feel as though they were actually present in a remote robot's environment. The ANA Avatar XPRIZE[1] aims to accelerate the development of these technologies, improving the quality and variety of haptic sensing and rendering devices and promoting fundamental research in system integration, networking, and Virtual Reality to create responsive, immersive and intuitive telepresence systems. We describe our experiences, insights, and future plans developing the AvaTRINA system, a collaboration between the University of Illinois and vRotors[2] and a finalist in the XPRIZE competition.

II. Related Work

Recent international robotics competitions have tested real-world performance of robot technology, and although some focused purely on autonomy [3, 4], the DARPA Robotics Challenge highlighted the need for teleoperation to embed human perception and judgement in complex environments and tasks, such as during search and rescue operations[5]. However, this challenge encouraged teams to use semi-autonomous and non-immersive interfaces for controlling robots due to degraded communication conditions [6, 7]. The ANA Avatar XPRIZE does not impose such communications limits and instead stresses robot avatar capabilities and operator immersion.

Avatar systems have been studied before and during the XPRIZE. The TELESAR VI robot [8] has a humanoid design, with mobile torso, arms and legs and focuses heavily on haptic immersion, being equipped with temperature and vibration sensing and rendering in its end-effectors but is incapable of locomotion, being also plagued by low arm payloads due to its constrained design. Wang et al. [9] demonstrate a whole-body exoskeleton teleoperation rig that enables highly dynamic and forceful motion, such as breaking a door with an axe. Closest to our application, Klamt et al. [10] present a highly mobile teleoperated robot which provides force-feedback to its operator through robotic arms attached to their hands. This robot's design was later improved with a more humanoid base, exoskeleton-based finger force and vibrotactile haptics, a facial reconstruction pipeline and a novel stereo rendering technique in VR to become the leading entry in the semifinals of the ANA Avatar Xprize [11].

(a) AvaTRINA Robot (b) Operator rig and view

Fig. 1: The AvaTRINA robot (left) and the operator-side apparatus and his view during a task execution (right)

III. System Design

AvaTRINA aims to use commodity VR input devices to be minimally obtrusive to the operator, avoiding tethers and anchors to provide an unburdened user experience. Moreover, we envision that the future of avatars will be many-user / few-robot deployments, which is facilitated by the use of low-cost user interfaces. The semifinals of the competition demanded a novice operator be trained to operate the system within one hour, during which the system also had to be simultaneously set up and then execute 3 scenarios in the following hour. Considering that the operator could be asked operate the robot for up to two consecutive hours, ergonomics were a primary concern. Hence, the operator interface was kept minimal, eschewing exoskeletons and complex haptic devices, both for ease of learning and to minimize operator fatigue. Further, we designed the robot to carry upwards of 3kg on each arm to enable performing daily household manipulation tasks.

A. Robot and Operator Hardware

The AvaTRINA robot consists of two 6-DOF cobot robot arms (UR5-e) with force-torque sensors, one 4-DOF 3 fingered haptic gripper (RightHand Reflex Takktile 2) and one parallel jaw gripper (Robotiq 2f-140), one custom-built pan-tilt head assembly and one omni-directional base (Waypoint Robotics

[1] https://www.xprize.org/prizes/avatar
[2] https://www.vrotors.com/

Vector), seen in Figure 1a. The Reflex gripper has under-actuated fingers with sparsely spread pressure-sensitive pads, our primary source of texture feedback.

The head houses a ZED-mini camera and a display for the operator's face. AvaTRINA is also equipped with a microphone and speakers and wireless e-stops. Further, it has an on-board computer with a Ryzen 5 3600X processor and an Nvidia RTX 2060 GPU, all attached to a fixed torso. The robot can also be fitted with Realsense L515 cameras on its wrists and torso for enabling SLAM and semi-autonomous functions.

The operator hardware, seen in Figure 1b consisted of a Meta Quest 2 headset and its controllers, chosen due to their standalone operation, one camera attached to the headset, wireless earphones with built-in microphones and a laptop.

B. Software, Control and Haptics

Figure 2 provides a schematic of Trina's software stack, which is modular and centralized, allowing both autonomous and semi-autonomous operation [12]. Arm control is done using Cartesian impedance control with a clutching design, which bypasses global scaling issues present in anchored systems [11]. The robot's end effectors (EE) do not track the VR controllers position until the clutch button is pressed. While the button remains pressed, the initial poses of the EEs and controllers are registered and the relative transform between the current controller position and its initial position is applied to the EEs initial pose to set the arm controller's Cartesian target. Finer control is provided through a motion-scaling adjustable parameter. The Base's velocity was controlled using the right analog stick of the Quest controllers.

Haptic rendering was performed via the Quest's vibration on the controllers through two means: The first increased the intensity of a monotonous vibration on the respective controller based on the low-pass filtered force-torque reading in the end-effectors, whereas texture was conveyed using vibrational patterns based on the derivatives of the pressure readings on the Takktile gripper's fingers as it was dragged along an object's surface.

The streaming module sends a 3840x1080 RTMP stereo stream to the headset (seen in Figure 1b) and communication between recipient and the operator was performed via a Zoom call, where the operator's video was created using a facial video reconstruction pipeline. This pipeline merged a pose-aligned looping video of the operator without the headset with the live lower face stream from the face camera, which was then input on the network in Siarohin et al. [13] in zero-shot fashion to animate an image of the operator without a headset, not relying on in-headset gaze tracking, unlike Schwarz et al. [11].

IV. Competition Evaluation

The 3 scenarios of the semifinals aimed at evaluating three basic components of a telexistence device: The first one consisted in collaborative assembling a shape-matching puzzle, testing manipulation accuracy and collaboration ability; the second emulated a brief business interaction, evaluating social

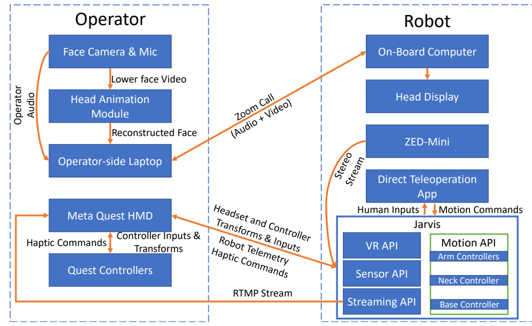

Fig. 2: AvaTRINA communication diagram

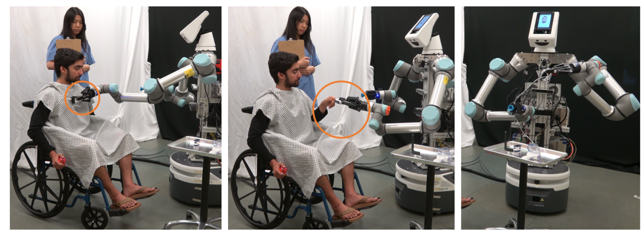

Fig. 3: Pre-semifinals AvaTRINA robot in mock medical appointment performing auscultation (left), compliant human contact taking pulse oxymeter measurements (center) and dextrous bimanual manipulation of deformable objects (right)

interaction ability; and the third focused on describing the texture and weight of a manipulated object, evaluating the quality and depth of haptic feedback. Each scenario, worth 30 points, was scored based on a task-based completion checklist and questionnaires on the subjective experiences of the operator and recipient. A showcase video was also submitted to highlight the robot's strengths, seen in Figure 3, the highlight of which being the teleoperated auscultation, later featured in subsequent work on autonomous auscultation [14].

Our entry scored 26.5, 28, 28.3 in scenarios 1, 2 and 3, respectively, placing 4$^{\text{th}}$ place in the semifinals overall. We observe that despite the operator-light design, the haptics focused scenario 3 was the highest rated, suggesting that tetherless vibrotactile interfaces might provide sufficient immersion with minimal discomfort. This aligns with literature indicating that while haptic feedback does improve task performance and teleoperator experience [15, 16, 17, 18], there seems to be diminishing returns on the richness of this feedback [16].

Operator comments also provide insights on improvements to be made before the finals. During the scenarios, the operators expressed concerns over high video streaming latency and poor visibility during teleoperation. To address these,we are moving to a WebRTC video pipeline using hardware accelerated encoding, adopting more nimble anthropomorphic grippers to reduce occlusions, providing richer haptic sensing, and adding a 3-DOF neck mechanism for added head mobility.

Poor depth perception was also identified as an issue - so we are investigating using actuated cameras on the head assembly to match the operator's interpupillary distance and investigating different stereo rendering techniques to improve stereo perception, such as Schwarz and Behnke [19] and Kuo et al. [20].

References

[1] J. V. Draper, D. B. Kaber, and J. M. Usher, "Telepresence," *Human Factors*, vol. 40, no. 3, pp. 354–375, 1998. [Online]. Available: https://doi.org/10.1518/001872098779591386

[2] S. Tachi, K. Minamizawa, M. Furukawa, and C. L. Fernando, "Telexistence — from 1980 to 2012," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 10 2012, pp. 5440–5441. [Online]. Available: https://ieeexplore.ieee.org/document/6386296

[3] N. Correll, K. E. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J. M. Romano, and P. R. Wurman, "Analysis and Observations From the First Amazon Picking Challenge," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 1, pp. 172–188, 2018. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7583659

[4] T. Rouček, M. Pecka, P. Čížek, T. Pet\vríček, J. Bayer, V. Šalanský, D. He\vrt, M. Petrlík, T. Báča, V. Spurný, F. Pomerleau, V. Kubelka, J. Faigl, K. Zimmermann, M. Saska, T. Svoboda, and T. Krajník, "DARPA Subterranean Challenge: Multi-robotic Exploration of Underground Environments," in *Modelling and Simulation for Autonomous Systems*, J. Mazal, A. Fagiolini, and P. Vasik, Eds. Cham: Springer International Publishing, 2020, pp. 274–290. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-43890-6_22

[5] E. Krotkov, D. Hackett, L. Jackel, M. Perschbacher, J. Pippine, J. Strauss, G. Pratt, and C. Orlowski, "The DARPA Robotics Challenge Finals: Results and Perspectives," *Journal of Field Robotics*, vol. 34, no. 2, pp. 229–240, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21683

[6] C. G. Atkeson, P. W. B. Benzun, N. Banerjee, D. Berenson, C. P. Bove, X. Cui, M. DeDonato, R. Du, S. Feng, P. Franklin, M. Gennert, J. P. Graff, P. He, A. Jaeger, J. Kim, K. Knoedler, L. Li, C. Liu, X. Long, T. Padir, F. Polido, G. G. Tighe, and X. Xinjilefu, "What Happened at the DARPA Robotics Challenge Finals," in *The DARPA Robotics Challenge Finals: Humanoid Robots To The Rescue*, M. Spenko, S. Buerger, and K. Iagnemma, Eds. Cham: Springer International Publishing, 2018, pp. 667–684. [Online]. Available: https://doi.org/10.1007/978-3-319-74666-1_17

[7] H. A. Yanco, A. Norton, W. Ober, D. Shane, A. Skinner, and J. Vice, "Analysis of Human-robot Interaction at the DARPA Robotics Challenge Trials," *Journal of Field Robotics*, vol. 32, no. 3, pp. 420–444, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21568

[8] S. Tachi, Y. Inoue, and F. Kato, "TELESAR VI: Telexistence Surrogate Anthropomorphic Robot VI," *International Journal of Humanoid Robotics*, vol. 17, no. 05, p. 2050019, 2020. [Online]. Available: https://doi.org/10.1142/S021984362050019X

[9] A. Wang, J. Ramos, J. Mayo, W. Ubellacker, J. Cheung, and S. Kim, "The HERMES humanoid system: A platform for full-body teleoperation with balance feedback," in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, 2015, pp. 730–737. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7363451/

[10] T. Klamt, M. Schwarz, C. Lenz, L. Baccelliere, D. Buongiorno, T. Cichon, A. DiGuardo, D. Droeschel, M. Gabardi, M. Kamedula, N. Kashiri, A. Laurenzi, D. Leonardis, L. Muratore, D. Pavlichenko, A. S. Periyasamy, D. Rodriguez, M. Solazzi, A. Frisoli, M. Gustmann, J. Roßmann, U. Süss, N. G. Tsagarakis, and S. Behnke, "Remote mobile manipulation with the centauro robot: Full-body telepresence and autonomous operator assistance," *Journal of Field Robotics*, vol. 37, no. 5, pp. 889–919, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21895

[11] M. Schwarz, C. Lenz, A. Rochow, M. Schreiber, and S. Behnke, "NimbRo Avatar: Interactive Immersive Telepresence with Force-Feedback Telemanipulation," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 5312–5319. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9636191/

[12] P. Naughton and K. Hauser, "Structured Action Prediction for Teleoperation in Open Worlds," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3099–3105, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9691823/

[13] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe, "First Order Motion Model for Image Animation," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d\textquotesingle Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper/2019/file/31c0b36aef265d9221af80872ceb62f9-Paper.pdf

[14] Y. Zhu, A. Smith, and K. Hauser, "Automated Heart and Lung Auscultation in Robotic Physical Examinations," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4204–4211, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9707618

[15] M. Chu, Z. Cui, A. Zhang, J. Yao, C. Tang, Z. Fu, A. Nathan, and S. Gao, "Multisensory Fusion, Haptic and Visual Feedback Teleoperation System under IoT Framework," *IEEE Internet of Things Journal*, p. 1, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9758071

[16] J. G. W. Wildenbeest, D. A. Abbink, C. J. M. Heemskerk, F. C. T. van der Helm, and H. Boessenkool, "The Impact of Haptic Feedback Quality on the Performance of Teleoperated Assembly Tasks," *IEEE Transactions on Haptics*, vol. 6, no. 2, pp. 242–252, 4 2013. [Online]. Available: https://ieeexplore.ieee.org/

abstract/document/6200270/
[17] E. Triantafyllidis, C. Mcgreavy, J. Gu, and Z. Li, "Study of Multimodal Interfaces and the Improvements on Teleoperation," *IEEE Access*, vol. 8, pp. 78 213–78 227, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9076603
[18] I. A. Kuling, K. Gijsbertse, B. N. Krom, K. J. van Teeffelen, and J. B. F. van Erp, "Haptic Feedback in a Teleoperated Box & Blocks Task," in *Haptics: Science, Technology, Applications*, I. Nisky, J. Hartcher-O'Brien, M. Wiertlewski, and J. Smeets, Eds. Cham: Springer International Publishing, 2020, pp. 96–104. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-58147-3_11
[19] M. Schwarz and S. Behnke, "Low-Latency Immersive 6D Televisualization with Spherical Rendering," in *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*, 2021, pp. 320–325. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9555797/
[20] C.-Y. Kuo, C.-C. Huang, C.-H. Tsai, Y.-S. Shi, and S. Smith, "Development of an immersive SLAM-based VR system for teleoperation of a mobile manipulator in an unknown environment," *Computers in Industry*, vol. 132, p. 103502, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166361521001093