Commodity Telepresence with Team AVATRINA’s Nursebot in the ANA Avatar XPRIZE Finals

Joao M. C. Marques¹, Jing-Chen Peng¹, Patrick Naughton¹, Yifan Zhu¹, James S. Nam¹, and Kris Hauser¹

I. INTRODUCTION AND RELATED WORK

Telepresence shows promise for extending human skills beyond geographic boundaries, enabling long-range skill transfer through robot avatars in applications like telemedicine [1,2], search-and-rescue operations [3-4], and remote environment exploration [5-7]. Early efforts to forward this technology, like the DARPA Robotics Challenge [8], were primarily focused on the robot’s ability to accomplish locomotion and manipulation tasks in adversarial conditions while guided by trained operators. Aiming to democratize the use of telepresence beyond specialized applications, the $10M ANA Avatar XPRIZE evaluated an avatar system’s intuitiveness, immersiveness, the avatar’s social capabilities, their robustness and manipulation capacity. It consisted of 10 distinct tasks, ranging from face-to-face communication and weight sensing, to surface texture rendering and remote tool use. External judges were asked to complete the tasks using the robot after a 45-minute training session on its usage.

Our entry, AVATRINA, placed 4th in the XPRIZE finals, being among the only 4 teams, out of 17 finalists, that completed all 10 tasks at the competition. This paper presents our system, highlights our unique system features, and shares insights and lessons learned from the competition.

II. TRINA AVATAR SYSTEM FOR XPRIZE FINALS

The design philosophy espoused by our team was to use relatively inexpensive, accessible operator interfaces to democratize access to telepresence robots, which may cost $100,000 or more. We pursued this approach for the XPRIZE Semifinals as described in a prior paper [9]. We summarize the major design changes to our final system (Figure 1) in Table I, and we highlight major features of our system below.

Our avatar system is comprised of the TRINA robot and a lightweight operator station connected by a communications network. The TRINA robot can operate without a tether with 2 hours of battery life, which was deemed more than sufficient for completing the finals requirements. The robot was also redesigned to have a width of 0.7 m, which is slim enough to fit comfortably through standard doorways.

We employ 7 DoF Franka Emika Panda robots as TRINA’s arms, since their kinematic redundancy allows us to maintain a human-like appearance while tracking motion targets. In addition, their high-frequency torque control interface allows for highly compliant contact behavior. The arm-to-torso mount design was selected to maximize the robot’s dexterous workspace. This was achieved by comparing different mount options according to how effectively the arms could track pre-recorded end-effector trajectories corresponding to typical manipulation tasks.

We chose an asymmetrical design for TRINA’s grippers – with an anthropomorphic left hand (Psyonic Ability Hand) and a parallel jaw gripper (Robotiq-140) on the right hand – because different gripper designs are better suited for different tasks: the Psyonic’s power grasp can hold arbitrarily shaped objects securely, while the parallel jaw’s ease of operation and robustness makes it suitable for operation in low visibility spaces and manipulating prismatic objects. The right gripper is equipped with a LiDAR camera to scan surfaces and determine their texture, like Shin and Choi [10]. This information is presented to the operator via augmented reality (AR) visual and haptic rendering, shown in Figure 2.

Additionally, since previous research in virtual reality [11,12,13] has shown that a mismatch between the interpupillary distance (IPD) of the VR headset and the user can result in reduced depth perception, we provide the first real-time adjustable stereo baseline camera system integrated with a teleoperated mobile manipulator. Our custom stereo camera system consists of two high-resolution and field-of-view cameras (Alvium 1800 U-500C with 1.67 mm focal-length wide-angle lenses) mounted on linear actuators that allow us to adjust TRINA’s stereo baseline to match the operator’s IPD. The camera is mounted on a 3D-printed 3 DoF neck assembly which mimics the operator’s head rotation, enabling human-like head motion. Video from the cameras is transmitted via WebRTC to the headset (Valve Index), minimizing latency.

We further use AR and heads-up-display (HUD) technologies to augment the operator’s situational awareness during...
navigation: a toggleable birds-eye view camera system lets the operator see obstacles to the sides and behind the robot, while ultrasonic proximity sensors warn the operator of any imminent collision risks through directional auditory cues.

We also enhance manipulation through AR and semi-autonomous functions. We employ force hemispheres to indicate the direction and magnitude of forces felt by the arms, with the force magnitude also conveyed by controller vibration intensity. Semi-autonomous functions are demonstrated in Figure 3 alongside these AR indicators.

To enable any operator to use the robot’s full workspace, which is typically larger than a human’s, TRINA’s arms are controlled by relative positioning. This is achieved with a clutching mechanism controlled by foot pedals. A SenseGlove DK1 is used on the operator’s left hand to command the anthropomorphic gripper and provide per-finger vibrotactile feedback based on gripper-mounted pressure sensors.

Finally, we deploy an adapted live talking head animation machine learning pipeline to render a live headset-free video of the operator on the robot [9], similar to [14], but not requiring on-site training or extensive headset modifications.

### III. XPRIZE Finals Performance

During the finals, our team was able to complete all 10 tasks, being one of only 4 teams to do so. During the first competition run, the judge operated TRINA to complete all 10 tasks in a time of 24:47. The second competition run was executed at a faster pace, but a hardware emergency stop was triggered on the final task, ending the run.

Schwarz et al. [15] provides a detailed per-task breakdown of each successful team’s performance, and notes that our robot had the slowest locomotion system of these four teams. This was due to the Waypoint Vector base’s built-in obstacle detection system generating spurious detections on the dark course, limiting its maximum speed. Execution was also slowed by a network disconnection during Task 6 that required an operator station restart. The operator noted that the semi-autonomous texture sensing mode was helpful in reaching and identifying the rocks effectively.

During run 2 the operator started the last task (object identification through texture) at the 12:17 mark, much faster than the previous day. However, while reaching to feel the object’s texture, TRINA’s arm collided with the environment with high force, causing it to engage mechanical brakes. Since the arm could not be remotely recovered, this prevented the operator from completing the task. We note that other successful teams also faced emergency stops [15] [16], but recovered from them through system robustness or redundancy.

### IV. Conclusion & Lessons Learned

The relatively successful performance of AVATRINA indicates that immersive and effective teleoperation can be achieved through minimal use of haptic load force feedback (LFF), which uses external devices to apply load forces to the operator’s arm. AVATRINA was one of the four teams to complete all tasks and had no LFF devices, while Pollen Robotics (second place) had a simple wearable LFF device and the other two used costly exoskeletons. This suggests that rich user interfaces with augmented reality and operator assistance are viable low-cost alternatives to LFF.

We also highlight our team’s use of semi-autonomous functions to assist with both routine and specialized tasks. Our modular software interface [9] allowed us to quickly develop a task-specific texture scanning mode for the finals, which greatly simplified the task of reaching and feeling the rocks, but note that integration of semi-autonomous tasks in shared autonomy remains an active area of research [17] [18].

### TABLE I: Major TRINA changes from XPRIZE Semifinals to Finals.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Semifinals</th>
<th>Finals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power (run time)</td>
<td>Power tether</td>
<td>1534 Wh battery (≈ 2 h)</td>
</tr>
<tr>
<td>Communications</td>
<td>Ethernet tether</td>
<td>WiFi</td>
</tr>
<tr>
<td>Neck Assembly</td>
<td>2 DoF</td>
<td>3 DoF</td>
</tr>
<tr>
<td>Eye Cameras</td>
<td>ZED-Mini</td>
<td>IPD-Adjustable Stereo</td>
</tr>
<tr>
<td>VR System</td>
<td>Meta Quest 2</td>
<td>Valve Index + PC</td>
</tr>
<tr>
<td>Left Gripper</td>
<td>Parallel Jaw</td>
<td>6 DoF Ability Hand</td>
</tr>
<tr>
<td>Right Gripper</td>
<td>4 DoF Jaw</td>
<td>Parallel Jaw</td>
</tr>
<tr>
<td>Arms</td>
<td>6 DoF UR5-e</td>
<td>7 DoF Pandas</td>
</tr>
<tr>
<td>Force Haptics</td>
<td>Vibrotactile</td>
<td>Vibrotactile + AR force-spheres</td>
</tr>
<tr>
<td>Texture Sensing</td>
<td>Pressure Sensors</td>
<td>LiDAR Camera</td>
</tr>
<tr>
<td>Texture Rendering</td>
<td>Vibrotactile</td>
<td>Vibrotactile, auditory, AR Mesh</td>
</tr>
<tr>
<td>Arm Control</td>
<td>Oculus Controllers</td>
<td>Valve Index Controller, Senseglove</td>
</tr>
<tr>
<td>Assistive Modes</td>
<td>High Precision, Homing, Base Speed, Texture</td>
<td></td>
</tr>
<tr>
<td>Video Streaming</td>
<td>RTMP, WebRTC</td>
<td></td>
</tr>
<tr>
<td>Arm Clutching</td>
<td>Oculus Triggers</td>
<td>Foot Pedals</td>
</tr>
</tbody>
</table>

Fig. 2: A rock being scanned (left) and reconstructed as an AR mesh displayed in the headset (right). The operator can also “feel” the virtual object as the controller moves across it, as we map the reconstructed texture to vibrotactile and audio feedback.

Fig. 3: (Left) View from the operator’s perspective, showing force-feedback spheres while lifting a heavy object. (Right) Heads-up display menu enables semi-autonomous functionalities: arm “homing” (right), texture-sensing mode (left), and base speed adjustment (center) icons.
REFERENCES


