The $10 Million ANA Avatar XPRIZE Competition

How It Advanced Immersive Telepresence Systems

By Sven Behnke, Julie A. Adams, and David Locke

Imagine how your life would change if you could instantly transport yourself anywhere in the world. Transcending the barriers of distance and travel time would enable many activities that are impossible today, including being present at remote events, connecting with family, friends, or colleagues, and efficiently distributing skills and hands-on expertise to locations wherever they are needed, e.g., to provide critical care and to deploy immediate emergency response in natural disaster scenarios. While the *Star Trek* transporter technology is yet to be invented, the closest thing to beaming is telepresence in an avatar robot.

INTRODUCTION

Sponsored with US$22 million by All Nippon Airways (ANA), the Avatar XPRIZE was a multiyear global competition focused on developing avatar systems that transport a human's senses, actions, and presence to a remote location in real time. XPRIZE, a nonprofit organization, announced the Avatar competition in March 2018 with a US$10 million prize purse.

The participating teams had to advance and integrate multiple emerging technologies to develop a physical, nonautonomous avatar system with which an operator can see, hear, and interact within a remote environment in a manner that feels as if they are truly there. A second objective was that a person at the remote end, the recipient, would feel that the operator was present in the avatar. The avatar systems were not operated by their developers, but by members of an international expert judging panel. Operators had only a short time to familiarize themselves with the systems before they had to perform tasks remotely in a variety of real-world scenarios, while the avatar systems had to convey a sense of presence for both the operator and the recipient. The judges also contributed to developing the guidelines, rules, and regulations that governed the competition (XPRIZE Foundation, Inc.: ANA Avatar XPRIZE Rules and Regulations).

By September 2019, 99 teams from 20 countries had registered (see the competition timeline in Figure 1). The judges selected 77 qualifying teams, who by February 2021 had to submit additional materials to be selected for the semifinals, consisting of a comprehensive paper detailing their avatar system's capabilities and their development plans, accompanied by a 15-min demonstration video showing their technology completing a self-selected scenario. The addressed use-cases included remote visits and interactions with a large variety of recipients; assisting people in need by preparing food, making drinks, and helping them get dressed; and providing health care by measuring body parameters, such as blood oxygen saturation. The 37 teams selected to participate in the semifinals ranged from enthusiasts to university groups to leading research labs and startup companies.

SEMIFINALS

The semifinal verification process required the selected teams to submit a video recorded in their own lab showing six self-selected tasks, worth 10 points toward the semifinal score.

Semifinals testing took place in September 2021 in Miami, FL, USA in two rounds. Teams had two days to set up their systems and two days for test runs. Each team had a garage work area and a room dedicated to their operator station. The operator and avatar scenario rooms were spatially separated and could only communicate over the XPRIZE-provided competition network with a full duplex, 1-Gb/s connection. Communication between the operator judge and the recipient judge was only possible through the avatar systems.

One hour of the two-hour test slots was dedicated to training the operator, a different member of the expert jury each day, on using the team's avatar system and approaches to solving the tasks. The second hour was used for the scored test run.

The systems were tested in three scenarios (see Figure 2), each worth up to 30 points. The judges evaluated the avatar systems based on four categories: operator experience (12 points), recipient

FIGURE 1. Competition timeline.

Digital Object Identifier 10.1109/MRA.2023.3310864
Date of current version: 12 December 2023

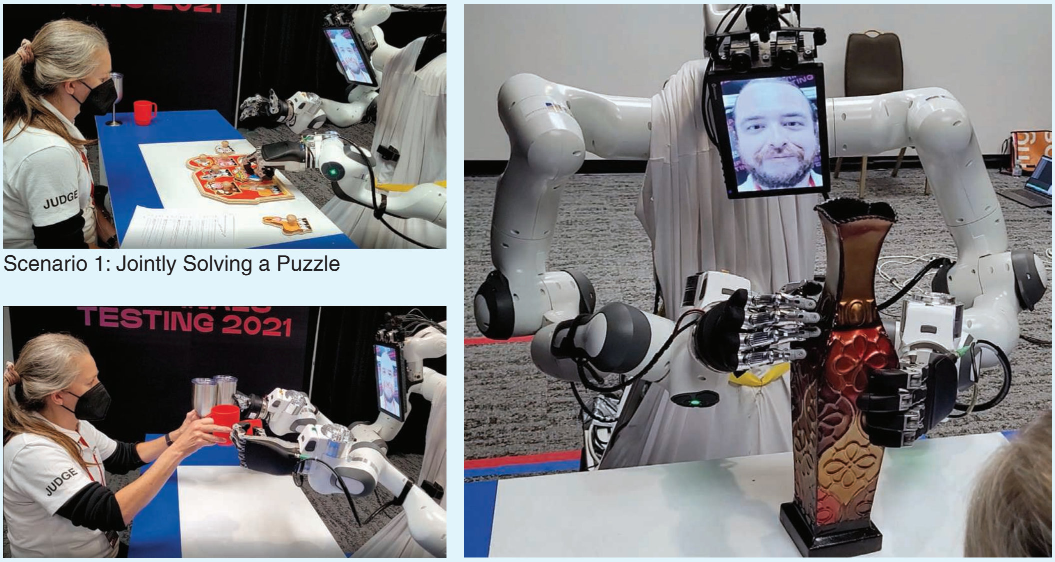
experience (eight points), avatar ability (six points), and overall system (four points). Twenty-nine teams participated in the semifinals event. Six additional teams that were unable to travel to Miami were visited by XPRIZE personnel and judges.

The avatar systems varied widely in complexity and capability. Avatar robots had one or two manipulator arms with grippers of varying dexterity: from simple to human-like five-finger hands. The avatars often had a movable head with cameras and microphones. Some teams also displayed the operator's face or animated their facial expressions in the avatar's head. Most avatar robots moved on wheels, but some used two legs for walking. Most operator stations used head-mounted displays (HMDs) to provide the visualization. Often, the operator used handheld virtual reality (VR) controllers or data gloves to teleoperate the avatar. Some teams captured operator inputs via upper-body and hand exoskeletons, providing the operator with force and haptic feedback.

A total of 214 scored scenarios were completed. Thirty-four teams completed at least one run and 31 teams completed all six scenarios. Twenty teams qualified for the finals. Prize money of US$2 million was paid in equal portions to the top 15 teams that tested in Miami. The first-ranked team, NimbRo, scored 99/100 points. Overall, most of the experience scores were very high, with the exceptions of the difficulty for the operator to feel haptics and forces remotely (54% of points scored), to see and hear clearly what was going on in the remote space (77%), and to sense their own position and movements in the remote space (79%).

FINALS

The finals were held from 1 to 5 November 2022 at the Long Beach Convention and Entertainment Center, Los Angeles County, CA, USA. Two qualified teams merged and two dropped out, leaving 17 teams from 10 countries. All of the avatar robots participating in the finals are shown in Figure 3.



Scenario 1: Jointly Solving a Puzzle

Scenario 2: Celebrating a Business Deal

Scenario 3: Exploring an Artifact

FIGURE 2. Semifinals scenarios.

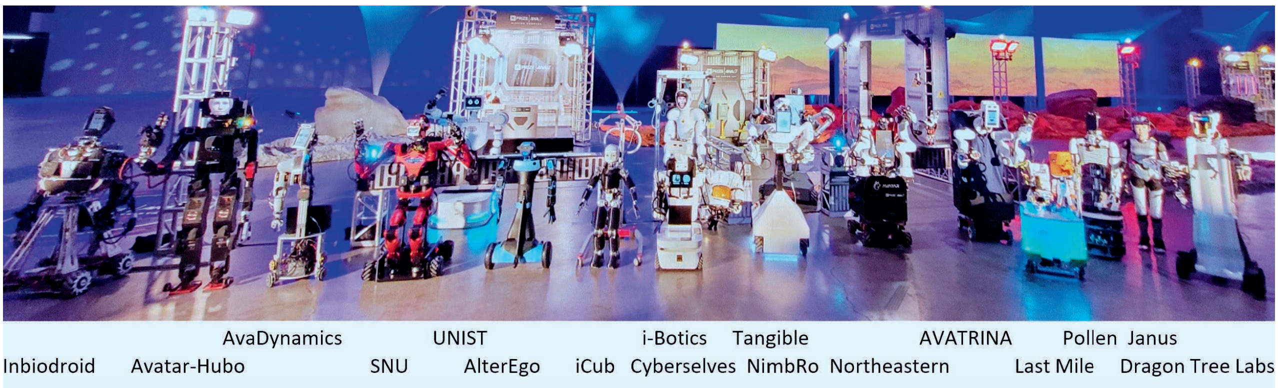


FIGURE 3. Avatar robots of all 17 ANA Avatar XPRIZE finals teams in front of the competition arena in Long Beach, CA.

Teams were given garage space and two days to set-up their systems. The avatar robots were tested on a single test track, shown in Figure 3. Five operator control rooms were shared by multiple teams; thus, an additional requirement was that the operator station had to be moved and set up quickly. Furthermore, the avatar robots had to be untethered, i.e., powered by batteries and communicating only via the XPRIZE-provided Wi-Fi.

> "AVATAR ROBOTS HAD ONE OR TWO MANIPULATOR ARMS WITH GRIPPERS OF VARYING DEXTERITY: FROM SIMPLE TO HUMAN-LIKE FIVE-FINGER HANDS."

The finals testing required the avatar to complete a mission that consisted of 10 tasks, described in Table 1, resembling a mission on a distant planet. The tasks had to be completed in the specified order and included navigation among obstacles, multimodal communication with a human recipient, simple and complex object manipulation, judging the weight of objects, using a power drill, and discriminating texture or roughness of stones. Figure 4 shows these tasks. The maximum mission time was 25 min and completion of all tasks was required to be eligible for the grand prize.

Completion of each task was worth one point. Operator and recipient judges scored their experience of being present in the remote space and of perceiving the presence of the remote operator, respectively, using the criteria listed in Table 2. Each criterion was assessed on a scale of never/poor (0 points), sometimes/fair (0.5 points), and always/good (1 point). Thus, up to five experience points could be earned.

Finals testing began with a scored qualifying day on which the judges operated all teams' avatars on the test course. The two finals testing days were open to the public, with more than 2,100 visitors and many media representatives in attendance and thousands watching the live stream of the first XPRIZE public testing event since the 2004 Ansari XPRIZE. The finals live-stream recording can be seen at https://youtu.be/lOnV1Go6Op0.

Three teams completed all 10 tasks on the first test day, with NimbRo in the lead (15 points), followed by Pollen Robotics (14.5 points), and AVATRINA (14.5 points). The top 12 teams advanced to Day 2. Table 3 reports the final results. The better of the two scores was retained, with the completion time as the tie breaker.

Four teams improved their task score on Day 2, three teams had the same score on both days, and five teams had a better score on Day 1. The average change in task score was one point, indicating a consistent evaluation of the avatar systems despite different judges, system changes between the days, and good or bad luck.

The task completion rates reported in Table 1 indicate that the top 12 teams solved all four tasks in the connectivity domain on both days. The first task that multiple teams failed to complete was Task 6, which involved discriminating between objects of different weight. The drop in completion rates for Tasks 8–10 was caused by two factors: lack of time and task difficulty. Specifically, grasping and using the power drill to remove a screw was successful only 67% of the time, and identifying a rough stone by feeling its texture and retrieving it had only 60% success rate.

The experience scores given by the operator and recipient judges were high, with an average score of 4.5 points, a minimum score of 3.5, and four teams receiving a perfect score of 5. This result indicates that, overall, the judges were quite satisfied with their experience. The recipient judges rarely deducted half a point, but the operator judges deducted points significantly more frequently.

Four teams completed all 10 tasks, with two receiving perfect experience scores from both judges. The fastest team, NimbRo, took an average of only 35 s per task, which is not much slower

TABLE 1. Finals tasks.

TASK	DESCRIPTION
CONNECTIVITY: HUMAN-TO-HUMAN CONNECTION	
Task 1 100%	The avatar robot maneuvers to the mission control desk.
Task 2 100%	The Avatar reports to the mission commander and introduces themselves.
Task 3 100%	The avatar receives the mission details and confirms them with the mission commander.
Task 4 100%	The avatar activates a switch which opens the station door.
EXPLORATION: THE NEW ERA OF TRAVEL	
Task 5 96%	The avatar exits the mission control room through the door and travels across the planet to the next task.
Task 6 79%	The avatar must identify the full power canisters that are among empty canisters.
Task 7 71%	The avatar places the correct canister into the designated slot which triggers the lighting of the next task zone.
SKILLS TRANSFER: EXPERTISE WITH NO BOUNDARIES	
Task 8 63%	The avatar navigates along the planet's surface to arrive at the next task.
Task 9 42%	The avatar must use the drill to remove the door.
Task 10 25%	The avatar must reach through the barrier to identify the rough textured rock and retrieve it.

The percentages report task completion rates of the 12 best teams.

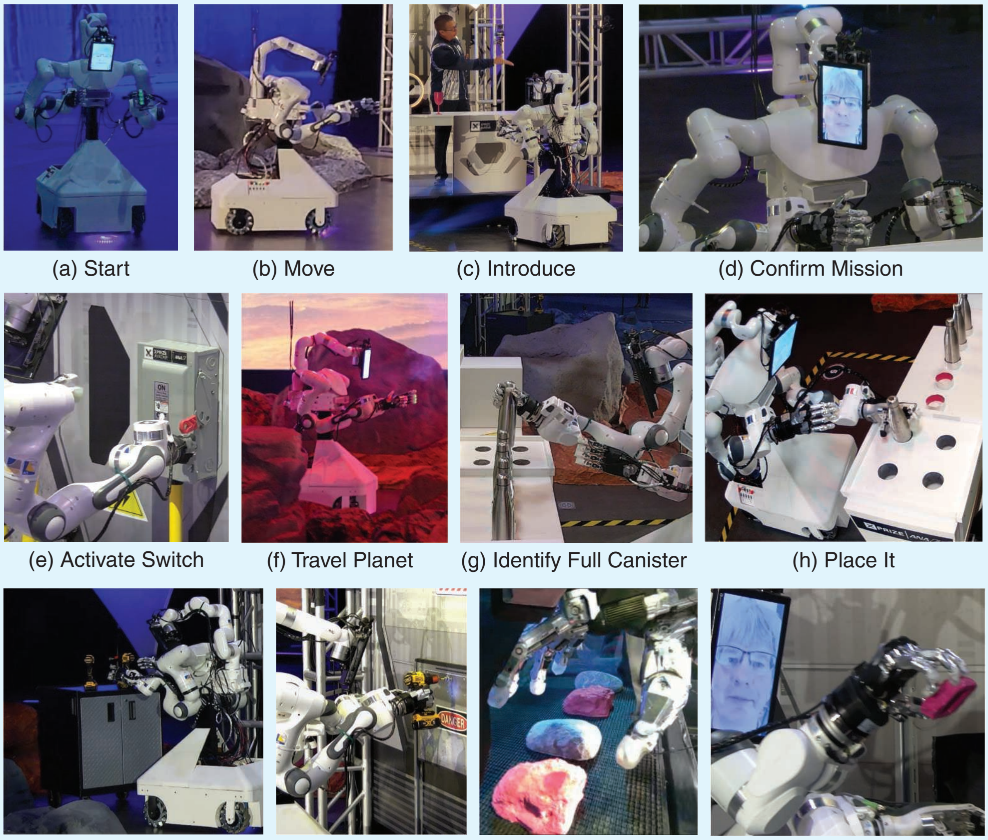


FIGURE 4. Finals tasks; see descriptions in Table 1.

than a human performing the tasks. The top three teams received a prize of US$5 million, US$2 million, and US$1 million.

Overall, finals testing worked well and the developed avatar systems were properly evaluated. Avatar system reliability was an issue, though. During 32% of the test runs, it took a significant amount of time for the avatars to start moving. There were also several stoppages and delays during the tests. The largest single cause of failure was network connectivity issues, but cable disconnections, software freezes, etc. also occurred.

Another issue for some teams was the operator's situational awareness. For example, it was not always easy for the operator to choose an appropriate communication distance to the recipient. Some avatar robots collided with test course objects and obstacles, sometimes resulting in falls. Overall, wheeled robots had a clear advantage over walking robots.

Judges sometimes struggled to remember how to use the systems. The operators quickly became immersed in the remote space, using the avatar robot as their own body, but if they had to communicate with team members in the operator room, that immersion broke.

Force and haptic feedback were necessary for Tasks 6 and 10 to identify a heavy object and to feel the texture of stones, but was also useful for other tasks and strongly enhanced the feeling of presence.

Two arms were not required for any task but were often helpful, e.g., to hold two canisters simultaneously to compare their weights, or to pick up or fix an object with one hand to grasp or scratch it with the other.

The ability for operators to change their viewing perspective without moving the robot base was important, as

TABLE 2. Finals experience scoring.

EXPERIENCE SCORING	CRITERION
Operator experience (3 points)	■ The avatar system enabled the operator judge to feel present in the remote space and conveyed appropriate sensory information. ■ The avatar system enabled the operator judge to clearly understand (both see and hear) the recipient. ■ The avatar system was easy and comfortable to use.
Recipient experience (2 points)	■ The avatar robot enabled the recipient judge to feel as though the remote operator was present in the space. ■ The avatar robot enabled the recipient judge to clearly understand (both see and hear) the operator.

it increased immersion and 3D scene perception, minimized occlusion, and allowed for selecting appropriate views for manipulation tasks.

Some form of animation or display of the operator's face on the avatar robot helped the recipient judge to experience the operator's presence and to see them clearly.

DEVELOPED AVATAR SYSTEMS

A large number of capable avatar systems were developed for this competition. Importantly, advances in the state-of-the-art of telepresence and mobile telemanipulation in the avatar robots, the operator interfaces, and the communication between the two were achieved. We exemplarily present the three winning teams.

NimbRo

NimbRo is the robot competition team of the Autonomous Intelligent Systems Lab at the University of Bonn, Germany.

The NimbRo avatar robot [Figure 5 (b)] has a human-like upper body with two compliant seven-degrees of freedom (DoF) arms (Franka Emika Panda) equipped with dexterous five-finger hands. The right hand (Schunk SVH) is very human-like in proportions and dexterity with 20 DoF, while the cable-driven left hand (Schunk SIH) is larger, more compliant, and has only 11 DoF. For haptic perception, the SIH fingertips are equipped with 3D Hall effect sensors. This hand's index finger is also equipped with two different microphones. The SVH fingertips are equipped with miniature switches. The hands are attached via six-axis force/torque sensors.

The robot's head is mounted on a 6-DoF robotic arm that provides full freedom of movement, mirroring the operator's head motion. The head consists of a display showing the animated operator's face, a pair of wide-angle cameras with a human-like baseline providing 4 K video streaming at 46 Hz, and a stereo microphone.

The upper body is attached to the robot base via a linear actuator, allowing manipulation at various heights, including the floor. Four Mecanum wheels provide omnidirectional driving capabilities. Two wide-angle cameras mounted on the front and the back of the robot's torso provide a birds-eye view around the robot while driving, facilitating safe obstacle avoidance. The robot is powered by a 48-V 30-Ah battery and has a fast on-board PC with Nvidia GPU and two Wi-Fi adapters.

The operator station [Figure 5(a)] is equipped with two 7-DoF arms and 20-DoF hand exoskeletons (SenseGlove DK1) that measure the operator's arm and finger movements. The hand exoskeletons are attached via six-axis force/torque sensors. The arms provide force feedback to the operator's hands and the exoskeletons exert resistance to finger closure via brakes. Further haptic feedback is provided to the fingers by vibrating actuators. The left index finger is also equipped with a linear actuator for continuous force feedback and a strong vibration actuator for haptic feedback.

The operator wears a VR HMD (HTC Valve Index) with stereo headphones, allowing full immersion in the remote situation (Figure 6). Spherical rendering is used to mitigate latency of the camera motion, preventing motion sickness. The HMD has been equipped with three additional cameras to capture the operator's mouth and eyes. Based on this, the operator's gaze direction and eyelid opening are estimated to photo-realistically animate their face in real time. The operator can control the omnidirectional

TABLE 3. Finals results.

RANK	TEAM	TIME	TASK	JUDGED	TOTAL
1	NimbRo (Germany)	5:50	10	5	15
2	Pollen Robotics (France)	10:50	10	5	15
3	Team Northeastern (USA)	21:09	10	4.5	14.5
4	AVATRINA (USA)	24:47	10	4.5	14.5
5	i-Botics (The Netherlands)	25:00	9	5	14
6	Team UNIST (Korea)	25:00	9	4.5	13.5
7	Inbiodroid (Mexico)	25:00	8	5	13
8	Team SNU (Korea)	25:00	8	4.5	12.5
9	AlterEgo (Italy)	25:00	8	4.5	12.5
10	Dragon Tree Labs (Singapore)	25:00	7	4	11
11	Avatar Hubo (USA)	25:00	6	3.5	9.5
12	Last Mile (Japan)	25:00	5	4	9

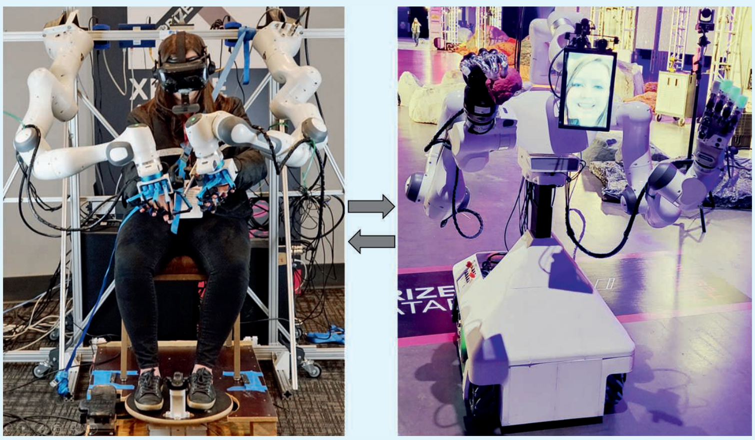


FIGURE 5. Avatar system of winning team NimbRo.

driving with a 3-DoF foot controller and upper body height with a pedal.

Separate robot operating system (ROS) cores run on the operator station and the avatar robot. They communicate bidirectionally via UDP using the nimbro_network library. The main camera stream (stereo 2,472 × 2,178 at 46 frames per second) is high-efficiency video coding-encoded (HEVC) and decoded on the GPU (NVENC), resulting in a bandwidth of 14 Mb/s. Audio is processed by the JACK Audio Connection Kit and redundantly transmitted via UDP using the OPUS audio codec. NVIDIA MAXINE is used for GPU-accelerated acoustic echo cancellation and Jamulus for team communications with operator and recipient. Control data are sent redundantly and packet loss is monitored. System operation is constantly monitored and components are autorespawned when necessary.

> "THE FINALS TESTING REQUIRED THE AVATAR TO COMPLETE A MISSION THAT CONSISTED OF 10 TASKS … RESEMBLING A MISSION ON A DISTANT PLANET."

POLLEN ROBOTICS

Pollen Robotics SAS is a telepresence startup based in Bordeaux, France. Their avatar robot [Figure 7(a)] is a version of their telepresence robot Reachy that was improved in several aspects for the competition. The upper body is human-like with two 7-DoF arms, each with a payload capacity of 3.5 kg. The arms are driven by self-designed parallel Orbita actuators: 2DoF in the shoulder and elbow and 3DoF in the wrist. The grippers have three underactuated fingers that adapt to objects. One finger is equipped with a fingernail and a microphone to provide haptic feedback. The robot head consists of a facial animation display, a stereo camera, and two microphones. It is attached to the torso via a 3-DoF actuator. The robot drives using a three-wheeled omnidirectional base with a compact, circular footprint that is equipped with a 2D horizontal lidar sensor to provide the operator with a navigation map.

The operator station provides immersive audio–visual feedback through an HMD (Oculus Quest 2) equipped with a facial tracker (VIVE) to capture the operator's mouth area for animating the avatar's face. The Oculus VR controllers used to capture the operator's hand movements were augmented with voice-coil vibration actuators (Actronica) to provide haptic feedback. The mobile base is controlled via a joystick on one of the controllers. One-DoF actuators provide torque

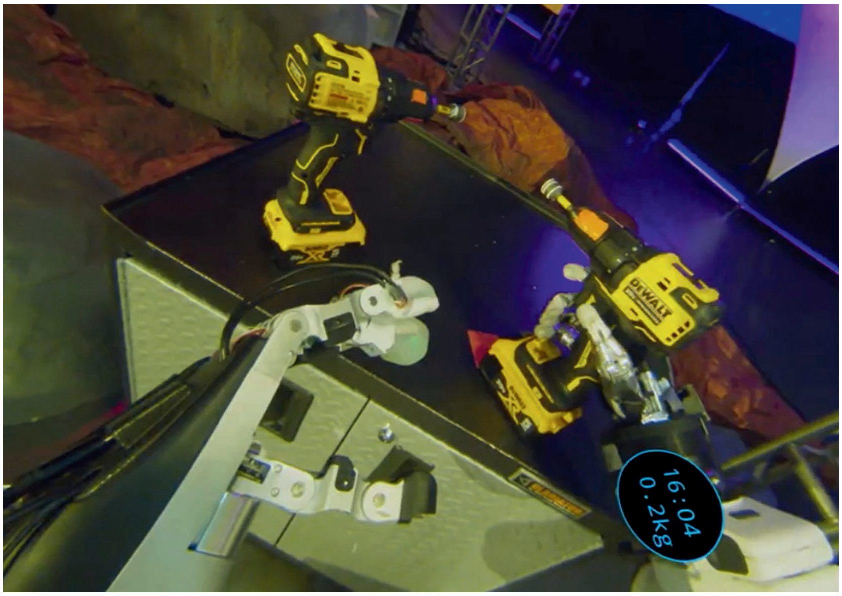
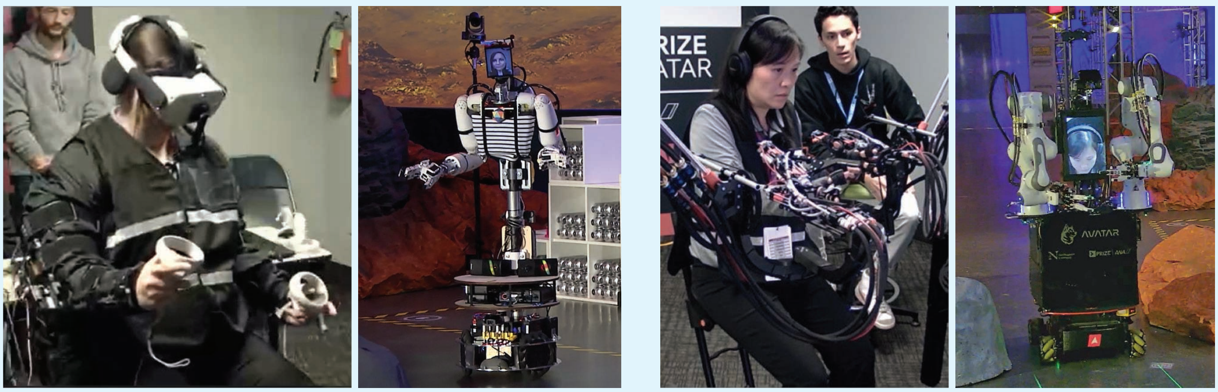


FIGURE 6. Immersive visualization of grasping a power drill.



(a) Pollen Robotics

(b) Team Northeastern

FIGURE 7. Avatar systems of (a) second-place team Pollen Robotics and (b) third-place team Team Northeastern.

feedback to the operator's elbows. WebRTC is used for bidirectional communication between the avatar robot and the operator station.

TEAM NORTHEASTERN

Team Northeastern is based at the Institute for Experiential Robotics at Northeastern University, Boston, MA, USA.

Their avatar robot Robalto [Figure 7(b)] is equipped with two 7-DoF arms (Franka Emika Panda) in a non-anthropomorphic configuration. The self-designed grippers have three fingers with low-impedance hydrostatic transmissions that provide haptic force feedback. The low-friction finger actuators and hydraulics with pressure sensors are located in the robot base. A display in the robot's center shows the operator. The robot is equipped with two wide-angle cameras: one main camera for manipulation and a top-down camera for navigation. The robot base drives omnidirectionally using four Mecanum wheels.

The operator station provides force feedback to the operator's hands using two 3-DoF exoarms. The operator wears hand exoskeletons with gimbals to track hand and finger motions and to provide force feedback to the fingers. Avatar driving is controlled with an omnidirectional pad.

The operator sees the remote scene on two large screens: one in portrait mode showing the manipulation workspace and a wide-angle landscape screen on the floor showing a top–down view for navigation. The avatar projects two parallel laser lines in the forward direction onto the floor, indicating its width to the operator for navigation, which is necessary to compensate for the lack of a 3D visualization. Similarly, two lateral laser lines tracking the gripper's forward motion indicate depth for manipulation.

DISCUSSION

The ANA Avatar XPRIZE competition was a multiyear collaborative effort by the organizers, judges, and participating teams to advance the state-of-the-art in telepresence robotics. The competition was well-organized with multiple milestones and team down-selections. The competition's focus changed along the way and, hence, it was also evaluating the team's abilities to adapt to unexpected requirements. During the semifinals, the focus was on the interaction between the operators and the recipients, and the judges enjoyed testing the different avatar systems' capabilities. Teams could define their own tasks and demonstrated them in a video that contributed to the score. In contrast, the finals focused more on capabilities, task completion, and overall system reliability. Since execution time was used for scoring, judges were only able to playfully explore an avatar's capabilities during operator training and after the competition.

> "SUCCESS WAS DETERMINED NOT ONLY BY THE TECHNICAL CAPABILITIES OF THE ROBOTIC SYSTEMS, BUT ALSO BY THE INTUITIVENESS OF THE OPERATOR INTERFACES AND THE EFFECTIVENESS OF THE OPERATOR TRAINING PROCEDURES."

When designing a competition's rules and tasks, it is not easy to define the difficulty level. The competition tasks need to be ambitious, but should also be achievable by the best teams. XPRIZE and the judges struck the right balance, as evidenced by the competition results. The clear scoring criteria and judge rotations resulted in a fair and reproduceable evaluation of the developed avatar systems, even for subjective criteria, such as the feeling of being present in the remote space.

The developed systems significantly advanced the field, including the immersive visualization of a remote scene, force and haptic feedback to feel the interaction with the environment, and animation of the operators' faces by the avatar. Head movements, gaze, gestures, and facial animations contributed to perceiving the operator being in the avatar and establishing shared attention with the recipient. It remains open, however, how much human-likeness was necessary or desirable and what form avatars should have.

Rather than converging to a single solution, the competition featured a variety of different approaches that performed well. Success was determined not only by the technical capabilities of the robotic systems, but also by the intuitiveness of the operator interfaces and the effectiveness of the operator training procedures.

The largest prize purse in a robotic competition to date attracted top research groups and companies. While teams had different policies regarding the disclosure of their technical approaches and results during the competition, XPRIZE organized a workshop after the finals to share developments and discuss lessons learned. This workshop was followed-up by a session at the 2022 IEEE Telepresence Symposium and the Second Workshop Toward Robot Avatars at the 2023 IEEE International Conference on Robotics and Automation.

After the competition, participants moved on to new challenges. Some teams wish to commercialize the developed technologies, which requires identifying viable use cases. The complex avatar systems could be further developed for dangerous or hard-to-reach domains, such as space, disaster relief, or medical assistance in isolation wards. Everyday virtual travel use requires avatar systems to become simpler and more affordable. Academic groups will explore research questions raised by the competition, including how much human-likeness avatars should assume and how to balance and interface direct control and autonomy.

XPRIZE is discussing a follow-up avatar competition with potential sponsors. This potential follow-up could raise the bar in several dimensions, including imposing bandwidth restrictions and latencies, locomotion on more difficult terrain, more complex manipulation (e.g., bimanual tasks), additional interaction modalities (e.g., temperature or smell), and deeper interactions between avatars and recipients, including interpretation of subtle communication cues and direct physical contact.