HUMANOID ROBOTS

iCub3 avatar system: Enabling remote fully immersive embodiment of humanoid robots

Stefano Dafarra1*, Ugo Pattacini2, Giulio Romualdi1, Lorenzo Rapetti1, Riccardo Grieco1, Kourosh Darvish1†, Gianluca Milani1, Enrico Valli1, Ines Sorrentino1,3, Paolo Maria Viceconte1,4, Alessandro Scalzo2, Silvio Traversaro1, Carlotta Sartore1,3, Mohamed Elobaid1, Nuno Guedelha1, Connor Herron5, Alexander Leonessa5, Francesco Draicchio6, Giorgio Metta6, Marco Maggiali2, Daniele Pucci1,3

We present an avatar system designed to facilitate the embodiment of humanoid robots by human operators, validated through iCub3, a humanoid developed at the Istituto Italiano di Tecnologia. More precisely, the paper makes two contributions: First, we present the humanoid iCub3 as a robotic avatar that integrates the latest significant improvements after about 15 years of development of the iCub series. Second, we present a versatile avatar system enabling humans to embody humanoid robots encompassing aspects such as locomotion, manipulation, voice, and facial expressions with comprehensive sensory feedback including visual, auditory, haptic, weight, and touch modalities. We validated the system by implementing several avatar architecture instances, each tailored to specific requirements. First, we evaluated the optimized architecture for verbal, nonverbal, and physical interactions with a remote recipient. This testing involved the operator in Genoa and the avatar in the Biennale di Venezia, Venice—about 290 kilometers away—thus allowing the operator to visit the Italian art exhibition remotely. Second, we evaluated the optimized architecture for a recipient physical collaboration and public engagement on stage, live, at the We Make Future show, a prominent world digital innovation festival. In this instance, the operator was situated in Genoa while the avatar operated in Rimini—about 300 kilometers away—interacting with a recipient who entrusted the avatar with a payload to carry on stage before an audience of approximately 2000 spectators. Third, we present the architecture implemented by the iCub Team for the ANA Avatar XPrize competition.

INTRODUCTION

The emergence of biological disasters and the evolution of digital virtual ecosystems necessitate the advancement of avatar technologies, enabling humans to inhabit either remote real locations or immersive virtual realities (VRs). The coronavirus disease 2019 (COVID-19) pandemic, for instance, highlighted the immature state of avatar technologies to facilitate effective human operations in distant locations (1). Analogously, the renewed interest of the engineering community in VR systems is also fueled by the increasing applications of digital and virtual ecosystems across various sectors (2). The renewed impetus is underscored by initiatives like the All Nippon Airways (ANA) Avatar XPrize, a $10 million competition (3) dedicated to creating avatar systems capable of transporting human presence to a remote real location in real time. This paper contributes to the development of technologies and methods that empower humans to embody physical humanoid robot avatars for real-time operations in remote locations.

When attempting to create physical avatars, one is tempted to apply the state of the art on telexistence (4). A telexistence system allows for transferring, and possibly augmenting, the skills of the human operator to a robotic avatar. Intuitiveness is a key feature of the system, trading the autonomy of the robotic avatar for the capabilities of the human operator to cope with unforeseen circumstances. Through the system, the operator is connected to the remote location while interacting with the environment or engaging with another person, referred to as “recipient.” Cybernetic avatars can also have effects at the societal level, allowing people to contribute to society without constraints (5, 6).

Physical avatar technologies benefit from the state of the art in telexistence. A physical avatar system is mainly composed of three components that are often the constituents of telexistence systems: the physical avatar, often a robot capable of navigating the environment; the operator system, which is in charge of retargeting and teleperception; and the communication layer, enabling communications between the avatar and the operator system.

Physical avatars are often implemented with robots capable of locomotion. Typical solutions include multi-legged or wheeled robots (7–10). In contexts where remote interaction with humans is crucial, humanoid robot avatars show great potential for existing and future applications. The robot’s human likeness increases its acceptability, its social interaction performances, and the clarity of its intentions (11). Also, when compared with wheeled or multilegged robots, a bipedal system design can perform more complex movements in reduced and confined spaces. Humanoid robots thus represent an optimal starting point for a platform to embody humans in terms of locomotion, manipulation, verbal, and nonverbal interaction, allowing an operator to have direct control over the whole body of the robot (12–14). The bipedal humanoid design, however, poses additional challenges because of the inherent instability of the robotic system. This complexity can be handled by letting the robot autonomously

---

*Corresponding author. Email: stefano.dafarra@iit.it
†Present address: Computer Science and Robotics Institute, University of Toronto, Toronto, Canada.
control its stability while achieving the desired tasks commanded by the remote operator, who may provide only high-level commands (for example, walking references) (15–17). In this case, the robot autonomously stabilizes the desired walking pattern, and the lower-body motion of the robot is not synchronized with the operator’s movements.

While navigating and manipulating the environment, the robot can sense its surroundings through specialized touch sensors. Touch feedback can have a noticeable effect on the manipulation capabilities of an avatar system (18), but it can also enable teleoperated physical interaction (19), giving rise to social implications. In the context of social presence, other avatar characteristics gain relevance, such as the control of facial expressions (20, 21).

The operator system often consists of a set of wearable devices and algorithms in charge of the so-called retargeting and teleperception features (22). The devices are often VR commercial products (23–25) or motion capture systems (13, 14). To achieve bilateral feedback teleoperation, it is possible to leverage ad hoc designed exoskeletons (8, 26). However, exoskeletons can be very invasive, thus constraining the operator’s motion in sometimes narrow envelopes.

The communication layer connects the operator system to the physical avatar. It allows the different components of the teleexistence system to communicate with each other in spite of potentially delayed networks (27). The software suite that implements the communication layer is usually referred to as middleware. Common middlewares are the Robot Operating System (ROS) (28) and Yet Another Robot Platform (YARP) (29).

Humanoid robots have been considered for a large variety of applications ranging from rehabilitation of the elderly to interactions with autistic children (30, 31). In many applications, humanoid robots were thought of as teleoperated machines, and social implications of this control mode received attention from the scientific community (32). These studies aimed to afford the operator a sense of embodiment in the robot, reproducing solely the operator’s movements, in an attempt to make the recipients engage with a small humanoid robot. Other applications focused on the retargeting of the operator’s motions in full-size humanoids (33), but they did not consider the locomotion aspects. In some cases, the operator could also be in charge of the balancing of the robot through lower-body (34) or full-body exoskeletons (35, 36). In other cases, exoskeletons provided haptic or vibrotactile feedback on the current balance status of the robot (15, 37). In contrast to the above efforts, we present in this paper a complete avatar system where the operator’s motions are fully retargeted on the robot, including the locomotion intents. The robot is in charge of keeping its balance while haptic and vibrotactile feedback is provided to the operator when the robot interacts with the environment.

Fully fledged avatar systems often allow the operator to control the robot’s manipulation and locomotion abilities while providing visual, auditory, and haptic feedback (7). A reduced teleoperation system has also been tested with the operator being an astronaut on the International Space Station (38). In most cases, the robot was either wheeled or in a sitting configuration (4). In this paper, we present a complete avatar system exploiting a humanoid robot while also considering emotional aspects. We let the operator control the robot’s facial expressions too while receiving haptic feedback when the robot was touched. Moreover, we tested the system with the operator and the robot positioned hundreds of kilometers apart.

The ANA Avatar XPrize competition provided a testing ground for avatar systems (39). Most teams adopted a wheeled configuration (9, 10, 20) or a hybrid legged-wheeled solution (25). Concerning the equipment enabling the robot teleoperation (referred to as “operator equipment”), both light commercial VR devices (10, 25) and custom-made exoskeletons (9, 20, 22) have been adopted. Our avatar system was the only one to complete tasks in the final stage using bipedal locomotion with a lightweight set of operator devices, comprising both commercial and custom-made wearables.

The contributions of the paper are twofold. First, we present the iCub3 humanoid robot. After about 15 years of development of the iCub platform, iCub3 is the latest iCub version with increased body size, optimized for locomotion and physical interaction tasks.

Second, we present a generic avatar system that allows an operator to embody humanoid robots. The operator is given a set of lightweight and noninvasive devices, comprising iFeel, custom-made wearable technologies for motion and force tracking, developed by the Istituto Italiano di Tecnologia (IIT). The avatar system simultaneously transports the operator’s locomotion, manipulation, voice, and facial expressions to the avatar with visual, auditory, and haptic feedback (weight, touch).

We validated the avatar system using iCub3 as an avatar. The validations consisted of four implementations of the avatar system, each designed to meet different objectives. First, we conducted a remote visit of the Italian Pavilion within the Biennale dell’Architettura di Venezia (40), where the operator was in Genoa and the avatar in Venice, at about a 290-km distance. The objective here was to test the embodiment and the remote physical interaction via the iCub3 avatar system. The demonstration is visible in movie S1 and online in a longer format (41). Second, we remotely participated in the live show We Make Future (42), where the operator was in Genoa and the avatar was in Rimini, about a 300-km distance, in front of 2000 spectators. The objective here was to perform a physical collaboration task with a remote recipient while entertaining an audience. The full show is available online (43), and the iCub3 demonstration is presented also in movie S1. Third, we participated in the ANA Avatar XPrize semifinals, where the objective of the implemented architecture was to maximize the sense of presence and shared situational awareness with a recipient while having precise control of the robot end-effectors. Fourth, we attended the ANA Avatar XPrize finals, where the avatar system was focused on complex locomotion and manipulation of heavy and textured objects.

RESULTS

This section introduces the validation results of the generic avatar system architecture presented in Materials and Methods. In particular, the validation scenarios listed in the Contribution paragraph of the Introduction define different requirements and shape different objectives (Table 1).

Remote teleoperation: iCub3 explores the Biennale di Venezia

The iCub3 avatar system underwent testing in a demonstration where the operator was situated at the IIT premises in Genoa, Italy, and the iCub3 robot was positioned in the Italian Pavilion of the Biennale dell’Architettura di Venezia located in Venice, Italy. Consequently, the operator and the robot were approximately 290 km apart, “linked” through a standard fiber optic internet network. The primary objective of this validation was to establish an architecture enabling the human operator to have verbal, nonverbal, and physical interaction...
Table 1. Summary of the set of validations we used for the iCub3 avatar system. For each validation, we define a set of requirements that meet specific objectives. More specifically, a validation might require the avatar to be in a remote location with respect to the operator (remote) or at a close distance (local), typically in the same building. Another requirement is represented by the level of expertise of the operator. We define an expert operator as someone who has deep knowledge of the avatar system, whereas a naive operator has to be trained before the beginning of the validation; the training time is fixed at about 30 min. We also categorize the validations according to the skills required of the avatar in terms of locomotion, interaction, and manipulation. Given the requirements and objective, we show which avatar system might be implemented. Each system setting and algorithm is detailed in a specific paragraph in Materials and Methods.

| Validation        | Location | Operator | Locomotion capabilities | Interaction with recipient | Manipulation | Requirements | Objectives                                      | System settings and algorithms |
|-------------------|----------|----------|-------------------------|---------------------------|-------------|--------------|-------------------------------------------------|-------------------------------|
|                   | Local    | Remote   | Expert                  | Naive                     | Short       | Long         | Side motions | Verbal | Non-verbal | Physical | Coarse | Precise | Tracking system | Body haptic feedback | Locomotion retargeting |
| Italian Pavilion  | ✓        | ✓        | ✓                       | ✓                         | ✓           | ✓            | ✓                         | Physical and nonverbal interaction | iFeel only | Touch feedback | Virtualizer |
| We Make Future    | ✓        | ✓        | ✓                       | ✓                         | ✓           | ✓            | ✓                         | Physical collaboration and public engagement | iFeel + trackers | Weight feedback | Virtualizer |
| XPrize Semifinals | ✓        | ✓        | ✓                       | ✓                         | ✓           | ✓            | ✓                         | Fine manipulation and shared situational awareness | iFeel + trackers | Weight feedback | Virtualizer |
| XPrize Finals     | ✓        | ✓        | ✓                       | ✓                         | ✓           | ✓            | ✓                         | Mission-oriented locomanipulation | iFeel + trackers | Weight feedback | iFeel walking |
capabilities with a person at the remote location, referred to here as the recipient. This demonstration was made possible through a collaboration between IIT and the Italian Ministry of Culture, and the test was conducted on 8 November 2021.

At the time of the demonstration, the logging systems described in Materials and Methods were not available yet; hence, we have no numerical data to present. This demonstration taught us the importance of such systems. The latency introduced by the communication channel only has been constantly monitored, remaining stably below 25 ms. This reduced latency did not affect the operator’s experience. In addition, the delay did not hinder the robot’s stability because its control system ensured balance independently from the network configurations. A video of the demonstration is available as part of movie S1. A more detailed version is available online (41), to which we refer in the following.

The first part of the video, up to time 0:55, is dedicated to the preparation of the operator, who wore the devices mentioned in Materials and Methods. At time 1:25, and later at 1:51, the operator exploited the robot locomotion capabilities. In particular, by walking inside the Cyberith Virtualizer platform, the operator was able to walk around the venue, as shown in Fig. 1 (A and B). At 1:26, the operator then interacted through the avatar with the recipient. In this context, the visual and auditory feedback were fundamental for a proficient verbal interaction. The face expression retargeting, demonstrated in Fig. 2 (A and B), enabled the nonverbal interaction, allowing the operator to smile to the recipient or to close the eyes in case of bright light, as demonstrated in Fig. 1C and in minute 2:19 of the detailed video. At times 1:58 and 2:08, the operator exploited the control over the robot body to express body language and to point at some installations while interacting with the recipient. The touch feedback was fundamental when the operator interacted with the venue at time 2:43 (Fig. 1D). The manipulation and fine control of each robot finger allowed the operator to touch the installation with delicacy while perceiving haptic feedback.

Finally, at time 2:52, we showcased the importance of the body haptic feedback for immersive interaction. As shown in Fig. 1 (E and F), the recipient reached the robot from outside its field of view. She then touched the robot’s arm. The robot skin perceived the touch and triggered the body haptic feedback. Hence, the operator perceived the remote touch and turned toward the recipient direction. The remote visit ended with the operator and the recipient sharing a hug, highlighting the emotional implications of such a rich interaction.

Remote teleoperation: iCub3 on the stage of the We Make Future festival

On 16 June 2022, iCub3 made a guest appearance on the stage of the We Make Future festival (42) in Rimini, as depicted in Fig. 3A. The robot was teleoperated from Genoa, situated approximately 300 km from the venue in Rimini, with a network delay comparable to that described in the previous subsection. The demonstration, featured in movie S1 and accessible online (43), aimed to validate an architecture that provided the operator with verbal, nonverbal, and physical interaction capabilities with another person in the remote location. Additionally, the avatar was tasked with engaging the public, as illustrated in Fig. 3E.

The implemented instance of the avatar system was similar to the one adopted for the remote visit at the Biennale di Venezia, except for the use of the VIVE trackers in conjunction with the iFeel suit for improved Cartesian control of the robot’s hands. Moreover, the iFeel haptic nodes were used to inform the operator about the weight held by the robot. During the demo, the robot was supposed to walk while carrying a weight (Fig. 3B). Figure 3C shows the center of mass tracking while walking with a box weighing about 0.5 kg. The controller was unaware of the additional weight, and it considered the weight an external disturbance. While walking, the robot controller favored the tracking of the walking-related trajectories compared with the retargeting trajectories. Thus, in case of conflicts, the robot balance was preserved at the expense of the retargeting performances. More details are in Materials and Methods. The external force

---

*Fig. 1. iCub3 explores the Biennale di Venezia. Snapshots of the video (41) demonstrating the remote teleoperation of iCub3 at the Italian Pavilion of the Biennale di Venezia. (A and B) The operator navigates the remote venue via iCub3. (C) The operator controls the iCub3’s eyelids in response to bright light. (D) The operator remotely grasps a piece of tissue through iCub3. (E) The robot is touched on the arm. The robot skin, whose activation is represented in (F), triggers the body haptic feedback on the operator.*
The scenarios were designed to assess the system’s capabilities, encompassing visual and auditory perception, gaze control, gestures, haptics, manipulation, grasping, and mobility. The team’s overall score was also based on the quality of interaction with the recipient solely through the avatar.

### Puzzle task

To test the manipulation and grasping capabilities of the system, one of the semifinal tasks required the operator to collaborate with the recipient, via the iCub3 avatar, on a toddler-like puzzle, shown in fig. S1. This task required high accuracy in the placement of the robot hand. To improve the Cartesian tracking of the operator’s hand movements, we resorted to the VIVE trackers in conjunction with the iFeel nodes. At the same time, the fine control of the robot fingers allowed the operator to firmly grasp and place the puzzle pieces (Fig. 4A). Figure 4 (E and F) plots the Cartesian error for the left and right hands. In particular, the desired position was reconstructed from the joint values obtained from the manipulation interface described in Materials and Methods, whereas the measured position was reconstructed from the joint values measured on the robot. Both quantities are expressed with respect to a frame attached to the robot pelvis link, with the z axis pointing upward and the x axis pointing forward. The measured position largely follows the desired position, with some exceptions. For example, at around \( t = 255 \) s, a small offset is visible on the z and y axes. This offset can be blamed on the balancing controller presented in Materials and Methods. Wide motions of the arms and torso can affect the CoM position, causing the balancing controller to intervene and potentially reduce the Cartesian tracking performance. Figure 4G presents a magnified version of the left-hand Cartesian tracking. There is an error of about 5 cm in the z direction. The operator was requesting the robot hand to be in a lower position, but this relatively large error may indicate that the robot hand was already touching the tabletop. From the same figure, there was a retargeting lag of approximately 0.5 s. The Cartesian tracking does not necessarily illustrate the performance of the system in completing the task but demonstrates that the motion of the operator was tracked on the robot. Moreover, for this specific task, the visual feedback was the most useful for the operator, who was able to compensate for lag and tracking errors by looking directly at the robot hand and performing movements at low speed.

### Weight task

Another semifinal task consisted of the operator detecting the weight of a vase through the avatar. The vase was placed on a table in front of the robot. Because of the limited size of the iCub’s hands, we instructed the operator to use both robot hands to determine the weight of the object. Similarly to the remote teleoperation experiment at the We Make Future festival, we exploited the iFeel haptic nodes to provide the operator an indication of the weight of the vase via haptic feedback. Moreover, we displayed in the headset the numeric value of the weight estimated via the arms’ F/T sensors. Figure 4H displays the normal force estimated by both arms when lifting the vase and putting it back in place. The weight of the vase was about 1 kg, and the F/T sensors partially overestimated it. The task was performed using two hands, thus introducing some internal forces that were measured by the F/T sensors.
vase on the table, the robot also gently nudged it against the tabletop, resulting in a positive force measured by the F/T sensors.

**Texture task**
The avatar was supposed to let the operator feel the texture embossed on the surface of the same vase. In Fig. 4B, the operator kept the vase still with the robot's left hand while scanning the surface with the right fingers to detect the embossed texture. For this task, we exploited the sensorized skin installed on the robot's fingers. The activation of the sensing elements triggered a vibration on the corresponding operator's finger via the haptic glove. The mapping function was tailored...
to be sensitive to light touches without being too distracting for power grasps. The activation of the skin and the consequent vibration feedback were not part of the logged data, and unfortunately, we have no numerical data to show for this task.

**Locomotion task**

The semifinals tested the systems' locomotion capabilities (Fig. 4C). The robot had to move away from the table and walk a couple of meters to reach a designated area, indicated by tape on the ground.
We exploited the Cyberith Virtualizer to trigger the robot’s motion. In particular, we instructed the operator to first turn around. At this point, the robot autonomously defined a set of steps to turn in place without moving forward (thus avoiding the table). Once fully turned, the operator started walking forward inside the Virtualizer to reach the designated area. Figure 4D shows the CoM tracking of the balancing controller described in Materials and Methods while walking away from the table and to the goal position.

ANA Avatar XPrize finals
We used the iCub3 avatar system in the ANA Avatar XPrize finals, conducted at the Long Beach Convention Center in Los Angeles, California on 1 to 5 November 2022. This conclusive phase of the competition featured the participation of 17 teams. An overview of the competition test course is illustrated in fig. S2.

In contrast to the semifinals, the competition’s focus shifted, emphasizing heavy-duty tasks over the interaction between the avatar and the recipient. Similar to the semifinals, the operator maneuvering the avatar system was an XPrize judge. However, dressing and training time were constrained to a total of 45 min. The avatar system was tested in a single scenario themed around exploration of another planet. The tasks are summarized in Table 2.

Points were awarded to the avatar system upon the completion of each task, with the requirement to accomplish all tasks within a total time of less than 25 min. Failure to complete a single task resulted in the termination of the trial. During our scored trial, the robot collided with one of the door’s pillars, leading to a fall that precluded further participation in the competition. We ranked 14th (39). The video of the trial can be found in the supporting information.

One additional complication was the Wi-Fi connectivity. The wireless connection to the robot was provided by the organizers, and the maximum bandwidth provided at the edges of the competition course was below 100 MB/s, whereas it reached 150 MB/s toward the center of the course. We minimized the network usage to avoid delays in the visual-manipulation pipeline, and we were unable to record logging data during the competition. Hence, we lack numerical data on the tasks performed during the finals.

Switch task
The switch used during the XPrize finals was a widely available commercial product. The handle required about 30 N to be moved. Such force could have had destructive effects on the robot hand/wrist mechanism, which was designated to hold light objects. Hence, we equipped the robot with a small plastic cylinder installed directly on the forearm assembly. On site, the internal handle resistance had been almost completely removed by the organizers. Nonetheless, the operator exploited the cylinder to activate the switch (Fig. 5B).

Locomotion tasks
Compared with the semifinals, the avatar locomotion had importance during the XPrize finals. Because of the time necessary to set up the Cyberith Virtualizer, and the necessity of walking sideways, we adopted the iFeel walking solution, described in Materials and Methods.

To improve the robustness of the walking motion, we increased the walking controller frequency to 500 Hz. iCub3 was the only robot in the finals successfully exploiting bipedal locomotion (Fig. 5A). While trying to pass through the door, the operator underestimated the dimensions of the robot, passing excessively close to one of the pillars. The robot arms were controlled with a rigid position controller (to allow fine control while manipulating), and when one arm hit the pillar, the resulting reaction force destabilized the robot, causing the fall (Fig. 5C) and thus ending our trial. The next subsections present our approach to the tasks we were not able to complete during the scored trial, with insights from the tests before the finals.

Bottles task
The estimation of the heavy canister exploited the same infrastructure of the weight task of the semifinals. On the other hand, compared with the previous case, we tested a configuration where the robot did not hold the object with two hands, but with a single one. Moreover, the weight estimation coming from the arms was printed separately with the previous case, we tested a configuration where the robot did not hold the object with two hands, but with a single one. Moreover, the weight estimation coming from the arms was printed separately on the headset. In this way, the operator could have easily compared the weight of two canisters while holding them in hand like in Fig. 5E.

Drill task
The grasping and activation of the drill would have been a difficult task for the iCub3 wrist/hand mechanism. The main problems were the weight of the drill, about 2.5 kg, and the force necessary to activate the trigger, about 15 N. The iCub wrist was not strong enough to fully sustain the drill weight. Nonetheless, we noticed that when trying to raise the drill, one of the wrist joints reached its mechanical limit. As a consequence, the weight of the tool was sustained by the forearm at the cost of reduced control over the orientation of the tool. At the same time, the iCub3 index alone was not strong enough to pull the trigger. To circumvent this issue, we installed a new gearbox on both the index and middle finger motors. The new gearbox had a reduction ratio four times higher. Moreover, the index finger appeared to be too short to operate the trigger successfully. As a consequence, we replaced the index finger with another middle finger, which was 5 mm longer. Finally, we tied the index and middle fingers together to use them jointly. Figure 5D shows iCub3 activating the drill.

Table 2. List of the ANA Avatar XPrize Finals tasks. Each task had to be completed sequentially. Failing to complete one task caused the end of the scored trial.

| Tasks | Description |
|-------|-------------|
| The avatar walks about 5 m to a designated spot, allowing the operator to communicate with the mission commander, who explains the mission. |
| The avatar walks about 30 m to the next task, where it has to identify one heavy canister according to its weight and place it in a designated spot. |
| The avatar walks about 10 m between obstacles and up to a table with a drill. The avatar activates the drill and unscrews a pin holding an opening with a small curtain. |
| The avatar reaches through the curtain to identify a rough textured rock and retrieve it. |
**Texture task**

The texture task required identification of a rough textured rock. To this end, we took advantage of the artificial skin covering the robot hand palms. Because the rocks were light and not fastened to the table, they could have easily slipped away. Therefore, our approach was to make contact with the rocks from the top using the sensorized palm as shown in Fig. 5F.

When contact was detected, a vibration pattern resembling either plain or rough texture was triggered on the operator’s hand. For the selection of the vibration pattern, we relied on a neural network trained to classify the type of contact (rough or plain) from the tactile sensors’ activations. In particular, each sensorized palm included 48 tactile sensors providing measurements in the numeric range from 0 to 255. The higher the value, the higher the measured pressure. We interpreted these measurements as a 9 pixel by 11 pixel–grayscale image, where each pixel, excluding padding, corresponded to a tactile sensor. Sample images retrieved from the palm in contact with a plain and a rough rock are shown in fig. S3 (A and B). Such images represent the input for our binary classifier, a customized version of the well-known AlexNet architecture (47) scaled in size by a factor of 32 to meet real-time inference constraints and equipped with smaller convolutional filters and fewer max-pooling layers to cope with our low-dimensional input. We trained our classifier for 25 epochs on a training dataset consisting of around 150 contacts per class, using batches of 32 samples and the Adam optimizer (48). In our test dataset, which included around 40 contacts, the overall trained model accuracy was 78% (see fig. S3C).

**DISCUSSION**

We presented a set of validations where an operator teleoperated the humanoid robot iCub3 to visit a remote exhibition or performed a live exhibition on a stage. The operator was able to walk while interacting physically, verbally, and nonverbally with a recipient through the avatar. We also demonstrated the iCub3 avatar systems’ capabilities by participating in the ANA Avatar XPrize international competition. In this context, the system proved to be very immersive and easy to use, given the placement at the semifinals. Nonetheless, the XPrize finals proved to be a severe testing ground for our system, allowing us to identify a series of shortcomings. In the following, we provide a series of insights and design recommendations.

**Design recommendations for system usability and insights**

In this section, we outline the key takeaways from the design process of the iCub3 avatar system. These lessons concern our context and the specific challenges that we encountered. We acknowledge that avatar systems are inherently diverse, and what worked well in our scenario may not generalize to other applications. Nonetheless, we believe that our lessons contribute valuable experiential knowledge to the field, potentially useful to many researchers working on humanoid robot avatars.

**Trade-off between the operator’s physical effort and transparency**

Operator movements can be mapped seamlessly onto the robot. Nonetheless, operators may need to put in effort to move their body against gravity or to maintain balance. All the more, if the operator has to move to trigger the robot locomotion, the operator’s energy expenditure may not be sustainable if the robot has to walk long distances. The use of supportive devices and equipment (like lightweight lower-body exoskeletons, chairs, or similar) can circumvent this issue at the expense of the embodiment. These devices also limit the operator’s motion, constraining their sense of presence in the remote location. In brief, light and wearable devices together with the
possibility of moving freely provide high transparency and immersion, at the cost of higher physical stress for the operator.

**Avatar design**

Humanoid robots are often considered “general purpose,” meaning that their human likeness can be useful in an unstructured environment at a human scale. In contrast, environments characterized by wide spaces and smooth flat ground encourage the use of wheeled robots against those implementing legged locomotion. At the same time, in cases of narrow spaces or irregular terrain, legged locomotion can be exploited by legged robots. In this respect, the operator should have the ability to define with more precision where the feet need to be placed. In our case, a possible approach could be to extend the iFeel walking solution presented in the “Manipulation interfaces” section. For example, a particular operator’s movement with one foot may be interpreted as a “manual mode” trigger, enabling direct control over the corresponding foot position. The robustness required to operate in a given environment is another element to consider. In the case of a humanoid robot, it is necessary to consider the possibility of a fall, thus implementing strategies that can reduce the resulting effects. Moreover, the robot should have some degree of autonomy to keep its balance while adapting to the environment.

From an acceptability point of view, iCub3 appeared to have high scores when engaging with the recipient, mostly because of its humanoid shape with relatively small dimensions and its physical resemblance to a child. This qualitative observation is supported by recent studies showing that having a face able to follow the recipient’s gaze increases a robot’s likeability (49). However, the use of a robotic head does not allow the recipient to immediately recognize the operator, which may impair the overall goal of making a robot avatar. Similarly, the use of robotic hands with anthropomorphic sizes increases the robot’s human likeness. However, the resulting mechanical complexity may limit the applicability to tasks because of the maximum force exerted by each finger, as occurred for our iCub3, where hands could only support relatively light external perturbations. In brief, human-like features seem to increase acceptability and engagement for the recipient but might be a limiting factor in case of heavy-duty tasks in simple environments.

Compared with its previous versions, iCub3 proved to be a much more robust robot. In particular, the absence of tendons in the legs and shoulders consistently reduced the maintenance time, because tendons tend to break over time. Moreover, iCub3 calibrates its joint position sensors at every startup by moving each joint to the hard stops. This ensures the repeatability of the robot’s motions. **Trade-off between modularity and ease of use**

Having a degree of modularity at the software level helps in developing and integrating different technologies into the robot. This allows one to enable and disable features, thus having a teleoperation system that meets multiple requirements. At the same time, there could be many operational units running in parallel, each one fallible in different ways. This might increase the complexity of having everything up and running. On the contrary, a monolithic system with a single “on-off” button, where everything is interconnected, might be easier to start but more difficult to recover in case of failures in one of its subsystems. The initial component development should be as separate as possible, working then on orchestration tools to start all the different parts in the correct order. A second layer to automatically recover in case of failure can render the system more robust.

At the hardware level, we can extend the modularity concept in terms of acceptability as well. All possible operators may not feel comfortable with a given wearable device. In other cases, such as people with disabilities, some devices might not be used at all. Therefore, the flexibility of the types of operator devices accommodates a wider range of potential users.

**Trade-off between off-the-shelf technology and in-house development**

When designing the avatar system, we advocate the good engineering practice of exploiting existing technologies, thus limiting the integration cost to the development of the layers to establish the connection with the existing architecture. This cost is often proportional to the flexibility of the architecture, as mentioned in the point above. In our system, this has been the case for the VR headsets, for example, where we used commercial devices only. Nonetheless, in some cases, it has been proven useful to “reinvent the wheel.” When a particular technology is aligned with one’s research direction, an attempt to develop a similar technology from a fundamental level can be useful and insightful, although very time consuming. The end result might allow large customizability and extensibility. As an example, the in-house development of F/T sensors allowed us to integrate them into shoes, an application that stemmed from the initial robotic application.

**Use of agile for team management**

When dealing with the organization of demos and competitions, it is fundamental to organize the work of the different team components. We adopted an agile methodology, common in project management but particularly shaped for robotics research. In particular, we divided the work into biweekly sprints. Each week, the team members joined in update meetings to discuss the progress and eventual difficulties. When close to an important event, the frequency of the updates increased by implementing standup meetings. The team components were encouraged to detail their progress in GitHub issues, thus providing implicit documentation. This proved to be fundamental to prepare for the events detailed in Results.

**Shortcomings identified during the XPrize finals**

In the spirit of full transparency and continuous improvement, it is essential to acknowledge and discuss the limitations of our system. We believe that identifying and understanding these shortcomings contributes to the portrayal of our work and can serve as a foundation for possible enhancements. In the following sections, we delineate key areas where our system exhibited room for improvement, thus identifying the challenges inherent in its current state.

**Operator system**

The avatar system allows the user to have direct control of different behaviors of the avatar at the same time, thus requiring the operator to undergo a constant and heavy cognitive load. One key difficulty is related to the sense of depth and the estimation of the actual avatar occupancy in the space.

Controlling the robot walking by stepping in place seems to improve the immersivity of the system to the point where the operator starts wandering unintentionally. During a trial run of the XPrize finals, the operator also felt like he was losing his balance, especially when he was able to see his physical body through the robot cameras. Although this condition is rare, it raises some concerns related to the use of supporting equipment for the operators at the cost of some degree of immersivity.

Another point to consider is lag of the video stream. The camera feed represents the principal source of feedback used by the operator.

Dafarra et al., Sci. Robot. 9, eadh3834 (2024) 24 January 2024
to control the position of the robot arms in space. The delay in the visual feedback leads the operator to be more cautious, thus requiring more time to perform a task, which in our experience also resulted in an increased cognitive load. When performing a fine manipulation task, the operator would often look at the robot hand through the vision system and perform small adjustments accordingly. However, because of the lag in the vision system, the corresponding robot motion might “overshoot” the operator’s intentions. This issue might be addressed by using multiple types of feedback. If the operator has to grasp an object, “feeling” the object in the hand before actually seeing it might make the operator understand that the task has been accomplished. Haptic feedback can usually be acquired and sent at a higher frequency with less lag compared with visual feedback, but, at the same time, this desynchronization between feedback channels increases the cognitive load.

**Communication layer**
During the XPrize finals, our team was particularly affected by Wi-Fi issues. Apart from the natural interference that occurs in an event with a multitude of electronic devices and wireless networks, we identified that the Wi-Fi antennas installed on the robot were too small. The effect of the network difficulties is also visible in the video of the scored trial (46), where the audio and the camera feeds are strongly delayed with numerous drops. The smoothness of the image feed then increased later in the course, because the wireless signal was stronger. This also indicates that using more complex compression techniques for audio and video is important. In our case, the audio was not compressed, whereas the camera images were compressed at a constant rate. Given that the signal strength varied across the field, having a variable rate compression algorithm could have helped.

**Avatar**
The use of a humanoid robot as an avatar poses many challenges. First, it is inherently unstable; consequently, any malfunction or unexpected disturbance might cause the robot to fall. In the case of iCub3, the number of exteroceptive sensors was reduced to the cameras and the Intel Realsense in the torso, limiting the possibility of the operator realizing whether there are obstacles close to the robot. This might result in a problem in a crowded or delicate environment. Some degree of shared autonomy could help the operator avoid hitting obstacles in this context. At the same time, exploiting more compliance and step recovery strategies could have helped us in recovery from the unexpected push.

The robot’s motion can be considered noncontinuous, dictated by the location of the footsteps. The operator cannot choose freely where to place the footsteps, whereas the robot has to necessarily alternate side motions while proceeding forward. Hence, the motion of the robot might appear unpredictable. Thus, an autonomous collision avoidance system could reduce the cognitive load required from the operator, but this would necessitate dedicated sensors, such as LIDARs (light detection and ranging sensors).

The hands represent another point of discussion. The iCub3’s hands are a complex mechanical system designed to finely manipulate light objects. This characteristic is also a consequence of the small space available to place the motors to control all the fingers. Consequently, we had difficulties when faced with the task of using a heavy object, like a drill. The complexity of the hand made it very difficult to apply last-minute changes, because a considerable amount of time is required to design and machine new parts. A more modular approach could have helped us fine-tune the hand according to the required tasks.

Despite these issues, the XPrize finals allowed us to test our system to its limits, and we were the only team able to exploit bipedal locomotion to complete a task. The finals also pushed us to exploit more of the avatar technologies on the operator side. We used the same F/T sensors in the robot’s feet and on the operator’s shoes. The operator and the robot technologies also share similarities at the code level. For example, the inverse kinematics approach is similar on both sides. Hence, the iCub3 avatar system represents an organic ensemble where the components are connected at logical, hardware, and software levels.

**MATERIALS AND METHODS**
The outcomes presented in Results were obtained by implementing several instances of the avatar system architecture detailed in this section. In particular, the avatar architecture is composed of two main interfaces, namely, the teleoperation and the teleperception interfaces, whereas a physical network establishes a logical link between the components of these two logical interfaces (see below).

The teleoperation interface is composed of two components, retargeting and control. The former collects the operator’s actions, intentions, and expressions via a set of devices that the operator wears. These inputs are transmitted in the form of references to the avatar control.

The second interface, the teleperception, is composed of the measurements and feedback components. The measurements retrieved by the robot are transmitted to the operator as feedback, providing a first-person perspective of the surroundings sensed by the robot.

**The avatar: iCub3**
The longstanding iCub platform has been evolving along several directions over the past 15 years (50). However, all its versions, which range from v1.0 to v2.9 (51), are based on a humanoid robot having mostly the same morphology, size, joint topology, actuation, and transmission mechanisms. In other words, the evolution of iCub mechanics never focused on the robot height, which was kept constant at about 1 m; the robot actuation and transmission mechanisms, which never evolved for the robot to increase its dynamism substantially; or its force-sensing capabilities, which are derived from F/T sensors of 45 mm diameter installed in the robot (52). The iCub3 humanoid robot shown in Fig. 6A is the outcome of a design effort that takes a step in all these directions. The robot represents a concept of a humanoid that will be the starting point when conceptualizing the next generation of the iCub platform.

**Mechanics**
The iCub3 humanoid robot is 125 cm tall and weighs 52 kg. Its mechanical structure is mainly composed of an aluminum alloy. The robot also presents plastic covers that partially cover the electronics. The weight is distributed as follows: 45% of the weight is in the legs, 20% in the arms, and 35% in the torso and head. Each robot leg is approximately 63 cm long, whereas the arms are 56 cm long from the shoulder to the fingertips. With the arms along the body, the robot is 43 cm wide. Each foot is composed of two separate rectangular sections, with a total length of about 25 cm and a width of 10 cm. The iCub3 robot has in total 54 degrees of freedom, including those in the hands and in the eyes, and they are all used in the avatar system. They are distributed as follows: four joints in the head controlling the eyelids and the eyes, three joints in the neck, seven
joints in each arm, nine joints in each hand, three joints in the torso, and six joints in each leg. The iCub3’s hands are equipped with tendon-driven joints, moved by nine motors, allowing the operator to control the thumb, the index, and the middle finger separately, whereas the ring and the pinkie fingers move jointly (53).

**Actuation**

iCub3 is equipped with both DC and brushless three-phase motors. The DC motors actuate the joints controlling the eyes, the eyelids, the neck, the wrists, and the hands. They are equipped with a Harmonic Drive gearbox with 1:100 reduction ratio. The torso, the arms, and the legs are controlled by three-phase brushless motors, also coupled with a 1:100 Harmonic Drive gearbox, with the exception of the hip and ankle roll joints, which have a 1:160 gearbox. The motor characteristics are as follows. The rated power is 110 W, with a rated torque of 0.18 N m, whereas the continuous stall torque is 0.22 N m. The hip pitch, knee, and ankle pitch joints are driven by another type of brushless motor, whose rated power is 179 W, rated torque is 0.43 N m, and continuous stall torque is 0.48 N m.

**Power, connectivity, computation, and electronics**

The iCub3 robot is powered either by an external supplier or by a custom-made battery of 600 W h. The connection to the robot can be established through an Ethernet cable or wirelessly via a standard 5-GHz Wi-Fi network. The robot head is equipped with an 11th generation Intel Core i7@1.8 GHz computer with 16 GB of random-access memory (RAM) running Ubuntu. This central unit represents the interface between the robot and the other laptops in the robot network (Fig. 7). The iCub3 central unit communicates with a series of boards distributed on the robot body and connected via an Ethernet bus (54). There are two main types of boards connected to the bus, both 32-bit Arm Cortex micro-controllers. The first are the Ethernet Motor Supervisor (EMS) boards, controlling the three phase motors with different control strategies. More details are available online (55). They run at 1 kHz and communicate via CAN protocol with the motor driver board (2FOC), which generates pulse-width modulation (PWM) signals at 20 kHz. The second are the MC4Plus boards, controlling the DC motors.

**Sensors**

A particular feature of iCub3 is the vast array of sensors available. iCub3 has eight six-axis F/T sensors (52) with integrated inertial measurement units (IMUs). More specifically, there are two different types, F/T-45 and F/T-58, where the number indicates the outer diameter of the sensor. The robot has six F/T-45 sensors. Two of them are mounted at the shoulders, and two are mounted on each foot, connecting the two sections of the feet to the ankle assembly. Two F/T-58 are located in the middle of the robot thighs. iCub3 has tactile sensors as an artificial skin (56) on the upper arm and the hands.

The head has two Basler daA3840-30mc cameras capturing images at 30 frames per second, with a 4K resolution. The resolution and the framerate are trimmable to reduce the network load. The images coming from the two sensors are processed by an NVIDIA Jetson Xavier NX module. The cameras are placed within the eye bulb and can be controlled to have a specified vergence, version, and tilt angle. Both eyes are equipped with eyelids, controlled jointly by a single DC motor. The robot head also includes a microphone on both ears and a speaker behind the face cover. Finally, a set of light-emitting diodes (LEDs) defines the robot face expression.

At the joint level, the iCub3 robot uses a series of encoders. First, an optical
encoder mounted on the motor axis estimates the motor magnetic flux. Second, the EMS boards exploit an off-axis absolute magnetic encoder mounted on each joint, after the gearbox, to estimate each joint position and velocity.

**Comparison with the classical iCub platform**

With respect to a classical iCub platform (50), the iCub3 humanoid robot is 21 cm taller and weighs 19 kg more. Figure 6B shows the different dimensions of the two platforms. The increased weight requires more powerful motors on the legs. Moreover, the torso and shoulder joints are serial direct mechanisms, whereas classical iCub robots have coupled tendon-driven mechanisms. This allows higher range of motion and greater mechanical robustness.

In addition, iCub3 has a higher capacity battery, 10,050 mA h versus 9300 mA h, and this is part of the torso assembly instead of being included in a rigidly attached backpack. The mechanics of the iCub head and hands have been retained from the classical iCub. From an electronics point of view, both platforms share the same 2FOC/EMS/MC4Plus architecture, although iCub3 has higher resolution joint encoders, using 18 bits compared with the 12 of the classical iCub architecture. The PC mounted inside the iCub3 head is more powerful and can also leverage the GPU capabilities of the Jetson Xavier board. The iCub3 platform has an additional Intel Realsense D435i depth camera placed in the front part of the torso, whereas the eye cameras have better resolution. In addition, the F/T-58 sensors are only used in the iCub3 robot.

**Robot control**

The robot motion is controlled by adopting a layered control architecture (57). Each layer generates references for the layer below by processing inputs from the robot, the environment, and the output of the previous layer. The more interior the layer, the shorter the time horizon used to evaluate the output. In addition, lower layers usually use more complex models to evaluate output, but a shorter time horizon often results in faster computations to obtain these outputs. The mathematical details of this layered architecture are provided in the "Robot control layered architecture" section in the Supplementary Materials.
The communication layer
Both the robot and the operator system require a cluster of different PCs connected in two interlinked local area networks (LANs) running multiple applications at once on different operating systems. The communication between the different applications is done through YARP (29). YARP supports building a robot control system as a collection of programs communicating in a peer-to-peer way, with an extensible family of connection types, like TCP (transmission control protocol), UDP (user datagram protocol), or other carriers tailored for the streaming of images.

For real-time operation, network overhead has to be minimized, so YARP is designed to operate on an isolated network or behind a firewall. However, the operator and the robot might be in two different, far places. To have the two subnetworks connected, we used OpenVPN (58). A simplified diagram of the robot and operator network is depicted in Fig. 7. The latency introduced by the VPN can go from 5 ms in a local configuration to several hundreds of milliseconds in case of bad internet connection.

The operator system
In the iCub3 avatar system, presented in Fig. 7, the operator exploits a series of devices. From the HTC VIVE family, we adopted the Pro Eye headset (59) with the facial tracker (60) and a set of trackers (61). The operator also uses the SenseGlove DK1 haptic gloves (62) and the Cyberith Virtualizer Elite 2 omnidirectional treadmill (63). Finally, the IIT custom-developed iFeel (64) sensorized haptic suit and shoes complete the set of wearable devices. The operator devices constitute the retargeting and feedback interfaces defined in Fig. 7.

The retargeting interfaces contain the set of commands that the operator exploits (on the robot) to achieve a specified task in the remote environment. In the iCub3 avatar system, we can distinguish the following retargeting interfaces: manipulation, locomotion, voice, and face expressions.

Manipulation interfaces
The manipulation process the operator motion to control the robot’s upper body. The reference trajectories, fed to the robot controller presented in the “Robot control” section, are computed using a multimodal sensor-fusion algorithm able to combine sensory information from the HTC VIVE headset and trackers, SenseGlove haptic gloves, and iFeel nodes. The headset and the trackers provide position-and-orientation measurements, which are scaled depending on the operator-avatar length ratio and used as a reference for the head and hands motion. The iFeel nodes contain an integrated IMU that measures the gravity vector, orientation, and angular velocity of the associated limbs. Similarly, an IMU is integrated into the haptic gloves, providing orientation and angular velocity of the hands.

The retargeting algorithm is modular and can be scaled depending on the available measurements. It is detailed in the section “Manipulation interface inverse kinematics algorithm” of the Supplementary Materials. Figure S5 presents two different sensor configurations used for upper-body motion retargeting on iCub3: iFeel only and iFeel plus trackers.

iFeel only The headset tracks the motion of the head, and the body motion is controlled exclusively by using the orientation and velocity measurements provided by the nodes, whose data are acquired at 70 Hz. This configuration and the corresponding mapping are presented in fig. S5A.

iFeel plus trackers Because the IMU estimation can be subject to divergence around the gravity axis and the robot end-effector Cartesian position is dependent on the model kinematic chain, VIVE trackers were added to the tracking system. In this configuration, gravity information provided by the nodes is used to regulate the internal movements of the robot while the trackers measure the desired Cartesian position for the hands at 90 Hz (fig. S5B).

In addition, the operator’s gaze and eye openness are tracked using the VIVE headset and facial tracker, allowing the operator to directly control the robot’s eyelids and gaze (65). The SenseGlove haptic glove completes the set of devices of the manipulation interface. It is an exoskeleton-like haptic glove allowing the translation of the motion of each of the operator’s fingers into a reference for the robot fingers.

Locomotion interfaces
The locomotion interface detects the operator’s walking intention and commands the robot locomotion. We implemented this interface in two different ways: Virtualizer and iFeel Walking.

Virtualizer The Cyberith Virtualizer Elite 2 is an omnidirectional treadmill where the operator walks by sliding. The motion is detected through optical sensors located on the device base plate. The movement direction is estimated via a moving ring attached to the harness secured to the operator’s waist. The base plate can also be inclined a fixed amount to ease the sliding motion, allowing the operator to walk naturally. The walking motion of the operator generates a reference walking direction and speed (16). These references are fed to the planning layer, described in the “Robot control layered architecture” section of the Supplementary Materials and interpreted as a reference point in eq. S3.

iFeel Walking The Virtualizer platform is bulky, limiting its transportability. Moreover, it is not possible to command a sideways motion. Hence, we developed the iFeel walking. It is composed of two logical components: intention detection and triggering. The intention detection defines the desired locomotion type. More specifically, moving one foot forward or backward enables forward and backward walking, respectively. Contrarily, moving one foot aside enables the lateral walking in the direction of the foot that moved (for example, moving the right foot to the side enables the right sidestepping). Finally, rotating the right (left) foot clockwise (counterclockwise) enables the clockwise (counterclockwise) in-place rotation. The intention is visualized in the VR headset through a set of arrows (Fig. 2D). Then, by stepping in place, the operator triggers the robot’s motion in the specified direction. The robot’s desired walking speed is modulated by the stepping frequency. Each intention is mapped to the corresponding control input on the modified unicycle dynamics of eq. S4.

The iFeel walking system requires measuring the relative position of each operator’s foot with respect to the waist and the normal force exerted in each foot to detect the stepping. The first quantity is measured via a set of VIVE trackers on the operator’s feet and waist. The second quantity, instead, is measured by the iFeel shoes, shown in Fig. 2C. The iFeel shoes estimate the interaction forces exchanged by the operator with the ground by means of two F/T-45 sensors installed on the soles.

Compared with Virtualizer solution, the iFeel walking solution does not constrain the operator at a fixed point. This might disorient some novice operators, because the immersivity can affect their sense of equilibrium. Moreover, the operator could step away from the tracked area. To avoid this issue, a message is printed on the headset to suggest that the operator move back to the original location.
The voice and face expression interfaces
The voice and face expression interfaces exploit the HTC VIVE headset microphone and the attached VIVE facial tracker. The former allows the operator to verbally interact through the robot. The latter is fundamental for nonverbal interaction. Using the headset facial tracker, the operator’s face expressions are replayed by the robot LEDs (Fig. 2, A and B).

Logging systems
We implemented two logging systems for two different purposes: online monitoring and offline processing. The online logging mechanism exploits the opencmt (66) framework to display the data measured from the robot. It connects through YARP reading the robot data streams, making it available from a normal browser, also from personal mobile devices. The code is available online (67). An example visualization of the GUI, with a live plot of the battery charge status and the communication delay to the robot PC, is shown in Fig. S6A.

The data streamed by the robot, together with some additional data coming from the walking controller, were also saved periodically in .mat files for offline analysis: The code is open source and available online (68). We implemented the so-called robot-log-visualizer (69) to quickly visualize and plot such data. Inspired by the Simulation Construction Set (SCS) tool (70) developed by the Florida Institute for Human and Machine Cognition (IHMC), robot-log-visualizer allows the user to visualize the data by simply clicking on the data of interest in the left panel, as shown in fig. S6B. On the right panel, we have a three-dimensional representation of the robot. It is also possible to display a synchronized camera stream, if available.

Supplementary Materials
This PDF file includes:
Supplementary Text
Figs. S1 to S6
References (71–82)

Other Supplementary Material for this manuscript includes the following:
Movie S1

REFERENCES AND NOTES
1. D. Leidner, The COVID-19 pandemic: An accelerator for the robotics industry? [Young Professionals]. IEEE Robot. Autom. Mag. 28, 116 (2021).
2. A. S. Pillai, G. Guazzaroni, Extended Reality Usage During COVID 19 Pandemic (Springer, 2022).
3. ANA Avatar XPrize; https://xprize.org/prizes/avatar.
4. S. Tachi, Telexistence, in Virtual Realities (Springer, 2015), pp. 229–259.
5. H. Ishiguro, The realisation of an avatar-symbiotic society where everyone can perform active roles without constraint. Adv. Robot. 35, 650–656 (2021).
6. Y. Horikawa, T. Miyashita, A. Utsumi, S. Nishimura, S. Koizumi, Cybernetic avatar platform for supporting social activities of all people, in 2023 IEEE/SICE International Symposium on System Integration (SII) (IEEE, 2023).
7. T. Kiami, M. Schwarz, C. Lenz, L. Baccellieri, D. Buongiorno, T. Chicon, A. DiGuardo, D. Droschel, M. Galbardi, M. Kamedula, N. Kashiri, A. Laurenzi, D. Leonardi, L. Menciotti, D. Pavlichenco, A. S. Periyasamy, D. Rodriguez, M. Solazali, A. Frisoli, M. Gustmann, J. Rollmann, U. Süss, N. G. Taagarski, S. Behnke, Remote mobile manipulation with the centaura robot: Full-body telepresence and autonomous operator assistance. J. Field Robot. 37, 889–919 (2020).
8. C. Lenz, S. Behnke, Bimanual telemanipulation with force and haptic feedback through an anthropomorphic avatar system. Rob. Auton. Syst. 161, 104338 (2023).
9. M. Schwarz, C. Lenz, A. Rochow, M. Schreiber, S. Behnke, Nimbro avatar: Interactive immersive telepresence with force-feedback telemanipulation, in 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE, 2021).
10. G. Lentini, A. Settini, D. Caporale, M. Garabini, G. Gori, L. Pallottino, M. G. Catalano, A. Bichi, After-ego: A mobile robot with a functionally anthropomorphic upper body designed for physical interaction. IEEE Robot. Autom. Mag. 26, 94–107 (2019).
11. A. D. Dragan, K. C. Lee, S. S. Srinivasa, Legibility and predictability of robot motion, in 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (IEEE, 2013).
12. K. Darvish, L. Penco, J. Ramos, R. Cisneros, J. Pratt, E. Yoshida, S. Ivandić, D. Pucci, Teleoperation of humanoid robots: A survey. IEEE Trans. Robot. 39, 1706–1727 (2023).
13. K. Darvish, Y. Tureapachuri, G. Romualdi, L. Rapetti, D. Ferigo, F. A. J. Chavez, D. Pucci, Whole-body geometric retargeting for humanoid robots, in 2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids) (IEEE, 2019).
14. L. Penco, N. Scianca, V. Modugno, L. Lanari, G. Oriolo, S. Ivandić, A multimode teleoperation framework for humanoid loco-manipulation: An application for the icub robot. IEEE Robot. Autom. Mag. 26, 73–82 (2019).
15. F. Abi-Farraj, B. Herze, A. Werner, M. Panzirsch, C. Ott, M. A. Roa, Humanoid teleoperation using task-relevant haptic feedback, in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE, 2018).
16. M. Elobaid, Y. Hu, G. Romualdi, S. Dafarra, J. Babic, D. Pucci, Teleexistence and teleoperation for walking humanoid robots, in Proceedings of SAII Intelligent Systems Conference (Springer, 2020), pp. 1106–1121.
17. D. Kim, B.-J. You, S.-R. Oh, Whole body motion control framework for arbitrarily and simultaneously assigned upper-body tasks and walking motion, in Modeling, Simulation and Optimization of Bipedal Walking (Springer, 2013), pp. 87–98.
18. J. A. Fishel, T. Oliver, M. Eichemrueller, G. Barbieri, E. Fowler, T. Hartikainen, L. Moss, R. Walker, Tactile telerobots for dull, dirty, dangerous, and inaccessible tasks, in 2020 IEEE International Conference on Robotics and Automation (ICRA) (IEEE, 2020).
19. A. Kaplish, K. Yamane, Motion retargeting and control for teleoperated physical human-robot interaction, in 2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids) (IEEE, 2019).
20. J. B. Van Erp, C. Sallaberry, C. Brekelmans, D. Dresscher, F. Ter Haar, G. Englebienne, J. Van Bruggen, J. De Greeff, L. F. S. Pereira, A. Toet, N. Hoeba, R. Lieftink, S. Falcone, T. Brug, What comes after telepresence? Embodiment, social presence and transporting one’s functional and social self, in 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (IEEE, 2022).
21. A. Rochow, M. Schwarz, M. Schreiber, S. Behnke, VR facial animation for immersive telepresence avatars, in 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE, 2022).
22. R. Luo, C. Wang, E. Schwarm, C. Keil, E. Mendoza, P. Kaveti, S. Alt, H. Singh, T. Padir, J. P. Whitney, Towards robot avatars: Systems and methods for teleinteraction at avatar xprize semi-finals, in 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE, 2022).
23. U. Martinez-Hernandez, L. W. Boorman, T. J. Prescott, Telepresence: Immersion with the icub humanoid robot and the oculus rift, in Conference on Biomimetic and Biohybrid Systems (Springer, 2015), pp. 461–464.
24. J. Shin, J. Ahn, J. Park, Stereoscopic low-latency vision system via ethernet network for humanoid teleoperation, in 2022 19th International Conference on Ubiquitous Robots (UR) (IEEE, 2022).
25. J. C. Vaz, A. Dave, N. Kassai, N. Kosanovic, P. Y. Oh, Immersive auditory-visual real-time avatar system of an avatar xprize finalist avatar-hub, in 2022 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO) (IEEE, 2022).
26. E. Schwarm, K. M. Gravesmill, J. P. Whitney, A floating-piston hydrostatic linear actuator and remote-direct-drive 2-dof gripper, in 2019 International Conference on Robotics and Automation (ICRA) (IEEE, 2019).
Dafarra et al. | Sci. Robot. 9, e9d3834 (2024) | 24 January 2024

Acknowledgments: This system paper was made possible by the effort of many people that contributed at different stages of the project. In particular, we would like to acknowledge the help of M. Salvi, D. Tomé, P. Ramadoss, Y. Tirupachuri, D. Ferigo, R. Camorani, and G. Cervetti in making this paper a reality. Funding: The paper was supported by the Italian National Institute for Insurance against Accidents at Work (INAIL) ergoCube Project. Author contributions: S.D. led the integration and validation activities. U.P. led the activities for the preparation and enhancements of the icub3 robot. G.R. developed the robotic locomotion.
algorithms. L.R. developed the operator retargeting algorithms. R.G. developed the weight retargeting application and contributed to the iFeel walking algorithm. K.D. contributed to the telexistence system algorithms/integration and developed the anthropomorphic telemanipulation framework. G.M. contributed to the development of the iFeel hardware. E.V. supervised the development of the iFeel hardware. I.S. improved the robot’s low-level control and estimation of external forces. R.M.V. developed the texture classification algorithm. A.S. developed the iCub3 low-level control. S.T. contributed to the architecture software stack and developed the estimation of the iCub3 joint torques. C.S. developed the iFeel walking algorithm. M.E. developed the first teleoperation architecture. N.G. developed the online logging visualization tool. C.H. developed the first version of the iFeel walking. A.L. supervised the development of the first version of the iFeel walking. F.D. contributed to design iteration of the iFeel wearables. G.M. conceived the original idea of the iCub3 robot. M.M. led the development of the iCub3 robot. D.P. supervised all the activities. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to support the conclusions of this manuscript are included in the main text or Supplementary Materials. The scripts to generate Figs. 3 (C and D) and 4 (D to H) and the corresponding data are available on the online repository Zenodo: https://zenodo.org/doi/10.5281/zenodo.10412849.

Submitted 28 February 2023
Accepted 21 December 2023
Published 24 January 2024
10.1126/scirobotics.adh3834