iCub3 Avatar System: Enabling Remote Fully-Immersive Embodiment of Humanoid Robots

Stefano Dafarra[*,1], Ugo Pattacini[2], Giulio Romualdi[1], Lorenzo Rapetti[1],
Riccardo Grieco[1], Kourosh Darvish[1,6], Gianluca Milani[1], Enrico Valli[1],
Ines Sorrentino[1,3], Paolo Maria Viceconte[1,4],
Alessandro Scalzo[2], Silvio Traversaro[1],
Carlotta Sartore[1,3], Mohamed Elobaid[1], Nuno Guedelha[1], Connor Herron[5],
Alexander Leonessa[5], Francesco Draicchio[7], Giorgio Metta[2],
Marco Maggiali[2], Daniele Pucci[1,3]

[1] Artificial and Mechanical Intelligence, Istituto Italiano di Tecnologia, Genoa, Italy,

[2] iCub Tech Facility, Istituto Italiano di Tecnologia, Genoa, Italy

[3] School of Computer Science, University of Manchester, Manchester, UK

[4] DIAG, Sapienza Università di Roma, Rome, Italy

[5] Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA, USA

[6] Computer Science and Robotics Institute, University of Toronto, Toronto, Canada

[7] DiMEILA, Istituto Nazionale Assicurazione Infortuni sul Lavoro (INAIL), Rome, Italy

(Corresponding author contact: `stefano.dafarra@iit.it`)

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 (IIT). More precisely, the contribution of the paper is twofold: first, we present the humanoid iCub3 as a robotic avatar which integrates the latest significant improvements after about fifteen 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 face expressions with comprehensive sensory feedback including visual, auditory, haptic, weight, and touch modalities. We validate the system by implementing several avatar architecture instances, each tailored to specific requirements. First, we evaluated the optimized architecture for verbal, non-verbal, 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 Km away – thus allowing the operator to visit remotely the Italian art exhibition. Second, we evaluated the optimised architecture for 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 operates in Rimini – about 300 Km away – interacting with a recipient who entrusted the avatar 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.

Summary: We present an avatar system to embody the humanoid robot iCub3 for remote verbal, non-verbal and physical interaction.

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 (VR). The COVID-19 pandemic, for instance, highlighted the immature state of avatar technologies to facilitate effective human operations in distant loca-

tions (*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 ANA Avatar XPrize, a 10M$ 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 at creating physical avatars, one is tempted to apply the state of the art on telexistence (*4*). A telexistence system allows transferring, and possibly augmenting, the skills of the human operator to a robotic avatar. Intuitiveness is a key feature of the system, trading off the autonomy of the robotic avatar with 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. 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 tele-perception; the communication layer, enabling communications between the avatar and the operator system.

Physical avatars are often implemented with robots having locomotion capabilities. 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 feature increases its acceptability, its social interaction performances, and the clarity of its intentions (*11*). Also, when compared to wheeled or multi-

3

legged 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 non-verbal 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 due to the inherent instability of the robotic system. This complexity can be handled by letting the robot autonomously control its stability while achieving the desired tasks commanded by the remote operator, which 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.

Whilst 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 tele-perception 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 telexistence 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

4

(ROS) (*28*), and 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 are thought of as teleoperated machines, and social implications of this control mode received attention from the scientific community (*32*). These studies aim 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 focus on the retargeting of the operator's motions in full-size humanoids (*33*), but they do not consider the locomotion aspects. In some cases, the operator can also be in charge of the balancing of the robot through lower-body (*34*) or full-body exoskeletons (*35, 36*). In other cases, exoskeletons can provide 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 the vast majority of the cases, the robot is 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. In fact, we let the operator control the robot's facial expressions too, while receiving haptic feedback when the robot is touched. Moreover, we test 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*) or custom-made exoskeletons (*9,20,22*) have been adopted. Our avatar system was the only one to complete tasks in the final stage utilizing bipedal locomotion with a lightweight set of operator devices, comprising both commercial and custom-made wearables.

The contribution of the paper is twofold. First, we present the iCub3 humanoid robot. After about fifteen years of development of the iCub platform, the 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 non-invasive 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 validate the avatar system using the iCub3 as an avatar. The validations consisted of four implementations of the avatar system, each designed to meet different objectives. First, 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 $290\,\mathrm{km}$ distance. The objective here was to test the embodiment and the remote physical interaction via the iCub3 avatar system. The demonstration is visible in the Supplementary Movie and online in a longer format (*41*). Second, a remote participation to the live show We Make Future (*42*), where the operator was in Genoa and the avatar in Rimini, at about $300\,\mathrm{km}$ 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*), while the iCub3 demonstration is presented also in the Supplementary Movie. Third, 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, 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 that will be presented in the Methods. In particular, the validation scenarios listed in the Contribution paragraph of the Introduction define different requirements and shape different objectives – see 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 $290\,\mathrm{km}$ 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 possess verbal, non-verbal, and physical interaction 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 November the 8th, 2021.

At the time of the demonstration, the logging systems described in the Methods were not

available yet, hence we have no numerical data to present. As a matter of fact, this demonstration taught us the importance of such systems. The latency introduced by the communication channel only has been constantly monitored, remaining stably below $25\,\mathrm{ms}$. This reduced latency did not affect the operator experience. In addition, the delay did not hinder the robot's stability since its control system ensured balance independently from the network configurations. A video of the demonstration is available as part of the Supplementary Movie. 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 the Methods section. 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 expressions retargeting, demonstrated in Fig. 2(A and B), enabled the non-verbal 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 time 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 June 16th, 2022, iCub3 made a guest appearance on the stage of the "We Make Future" Festival (*42*) in Rimini, as depicted in Fig. 3(A). The robot was teleoperated from Genoa, situated approximately $300\,\mathrm{km}$ from the venue in Rimini, with a network delay comparable to that described in the previous subsection. The demonstration, featured in the Supplementary Movie and accessible online (*43*), aimed to validate an architecture that provided the operator with verbal, non-verbal, and physical interaction capabilities with another person in the remote location. Additionally, the avatar was tasked with engaging the public, as illustrated in Fig. 3(E).

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. In fact, during the demo, the robot was supposed to walk while carrying a weight, see Fig. 3(B). Figure 3(C) shows the center of mass tracking while walking with a box weighing about $0.5\,\mathrm{kg}$. The controller was unaware of the additional weight and it considered it an external disturbance. While walking, the robot controller favored the tracking of the walking-related trajectories compared to the retargeting trajectories. Thus, in case of conflicts, the robot balance was preserved at the expense of the retargeting performances. More details are in the Methods section. The external force induced by the weight of the box was measured by the force-torque (F/T) sensors installed on the robot arms. Figure 3(D) displays the vertical component of the measured forces exerted

9

on the robot arms. At $t = 0\,\mathrm{s}$, we can notice an initial offset measured by the F/Ts. Then, around $t = 10\,\mathrm{s}$, the box was handed to the robot. At about $t = 17\,\mathrm{s}$, the robot started walking and the impacts with the ground induced some disturbances in the measured linear force. Finally, after the robot stopped walking, at $t = 27\,\mathrm{s}$ the recipient took the box back from the robot.

ANA Avatar XPrize Semifinals

In the semifinals of the ANA Avatar XPrize competition, the iCub3 Avatar system underwent a series of tasks while being operated by two competition judges. Following a half-hour training session, each judge was required to execute three scenarios, each consisting of six atomic tasks. The scenarios were designed to assess the system's capabilities, encompassing visual and auditory perception, gaze control, gestures, haptics capabilities, manipulation, grasping, and mobility. The team's overall score during the semifinals hinged on the proficiency exhibited in these aspects. Additional judging factors included the quality of interaction between the operator and another human being, the recipient. Hence, body language, emotional expression, and shared situational awareness were also taken into consideration. Moreover, being the system teleoperated by a novice operator, the intuitiveness and the ease of use of the teleoperation system were implicitly factored into the evaluation. Some tasks were replicated among different scenarios, while others specifically necessitated the operator to communicate to the recipient solely through the avatar. In the following, we present a meaningful excerpt of the semifinal tasks, defining how the iCub3 Avatar system was used to approach them.

In total, 38 teams from 16 countries qualified for the semifinal stage (*44*). The total score assigned to the iCub3 Avatar system was 95 over a total of 100 points (*39*), which was worth second place overall (*45*). The semifinals took place on the 21st of March 2022, with the XPrize judges visiting the IIT labs in Genoa.

Puzzle task

In order 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. 4(A). Figures 4(E and F) plot the cartesian error for the left and right hand. In particular, the desired position has been reconstructed from the joint values obtained from the manipulation interface described in the Methods, whereas the measured position is 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 $t = 255\,\mathrm{s}$, a small offset is visible on the z and y axes. This offset can be blamed on the balancing controller presented in the Methods. In fact, wide motions of the arms and torso can affect the CoM position, causing the balancing controller to intervene and potentially reduce the Cartesian tracking performances. Figure 4(G) presents a magnified version of the left hand Cartesian tracking. There is an error of about $5\,\mathrm{cm}$ 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 is a retargeting lag in the order of $0.5\,\mathrm{s}$. The Cartesian tracking does not necessarily indicate 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.

11

Weight task

Another semifinal task consisted in the operator detecting the weight of a vase through the avatar. The vase was placed on a table in front of the robot. Due to the limited dimension of the iCub 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 4(H) 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 $1\,\mathrm{kg}$, and the F/Ts partially overestimated it. The task was performed using two hands, thus introducing some internal forces that were measured by the F/Ts. When placing the vase on the table, the robot also gently nudged it against the tabletop, resulting in a positive force measured by the F/Ts.

Texture task

The avatar was supposed to let the operator feel the texture embossed on the surface of the same vase of the previous paragraph. In Fig. 4(B), the operator was keeping the vase still with the robot's left hand, while he scanned 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 distractive for power grasps. The activation of the skin and the consequent vibration feedback was not part of the logged data and, unfortunately, we have no numerical data to show for this task.

Locomotion task

The Semifinals tested the system locomotion capabilities, Fig. 4(C). 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 4(D) shows the CoM tracking of the balancing controller described in the 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 1-5 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 minutes. The Avatar system was tested in a single scenario themed on the 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 minutes. 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 is included in the Supplementary

Video and is available online (*46*). The following section outlines our approach in deploying the iCub3 avatar system for the various tasks, even those that were not successfully completed during the competition.

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 $100\,\mathrm{MB/sec}$, whereas it reached $150\,\mathrm{MB/sec}$ 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 $30\,\mathrm{N}$ 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 the 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 to the semifinals, the Avatar locomotion had significance during the XPrize finals. Due to the time necessary to set up the Cyberith Virtualizer™ , and the necessity of walking sideways, we adopted the iFeel walking solution, described in the Methods.

To improve the robustness of the walking motion, we increased the walking controller frequency to $500\,\mathrm{Hz}$. iCub3 was the only robot in the finals successfully exploiting bipedal locomotion, Fig. 5(A). Whilst 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. 5(C), 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 prior to 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 to 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. 5(E).

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 $2.5\,\text{kg}$, and the force necessary to activate the trigger, about $15\,\text{N}$. The iCub wrist was not strong enough to fully sustain the drill weight. Nonetheless, we noticed that when trying to raise it, one of the wrist joint would have 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 $5\,\text{mm}$ longer. Finally, we tied the index and middle fingers together to use them jointly. Figure 5(D) shows iCub3 activating the drill.

Texture task

The texture task required to identify a rough textured rock. To this end, we took advantage of the artificial skin covering the robot hand palms. Since 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. 5(F).

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 [0,255]. The higher the value, the higher the measured pressure. We interpreted these measurements as a 9x11-pixels grayscale image, where each pixel, excluding padding, corresponded to a tactile sensor. Fig. S3(A and B) shows sample images retrieved from the palm in contact with a plain and a rough rock, respectively. 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 less 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*). On our test dataset which included around 40 contacts, the overall trained model accuracy was 78% - see Fig. S3(C).

Discussion

We present 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 non-verbally with a recipient through the avatar. We also demonstrate the iCub3 avatar system capabilities by participating to 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. In the following, we provide a series of insights and design recommendations. Nonetheless, the XPrize finals proved to be a severe testing ground for our system, allowing us to identify a series of shortcomings.

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 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 our lessons contribute valuable experiential knowledge to the field, potentially useful to many researchers working on humanoid robot avatars.

Tradeoff between the operator's physical effort and the transparency   Operator movements can be mapped seamlessly onto the robot. Nonetheless, operators may need to put effort in order 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 for 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. In fact, 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 as "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 shall encourage the usage of wheeled robots against those implementing legged locomotion. At the same time, in case of narrow spaces or irregular terrain, legged locomotion should be exploited by legged robots. In this respect, the operator should have the possibility 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 the balance while adapting to the environment.

From the 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 robots with faces able to follow the recipient's gaze increase their 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 due to the maximum force exerted by each finger, as it was 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 to 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, since 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.

Tradeoff 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. Not all possible operators might feel comfortable with a given wearable device. In more extreme cases like people with disabilities, some devices might not be used at all. Therefore, the flexibility of the types of operator devices allows for accommodating a wider range of potential users.

Tradeoff 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 employed 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 FT 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 adopt an agile methodology, common in project management, but particularly shaped for robotics research. In particular, we divide the work into biweekly sprints. Each week, the team members join in update meetings to discuss the progress and eventual difficulties. When close to an important event, the frequency of the updates increases by implementing standup meetings. The team components are encouraged to detail their progress in `GitHub` issues, thus providing implicit documentation. This proved to be fundamental to get prepared for the events detailed in the Results section.

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 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, due to 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. The haptic feedback can usually be acquired and sent at a higher frequency with less lag compared to the visual feedback, but, at the same time, this desynchronization between feedbacks increases the cognitive load.

Communication layer   During the XPrize finals, our team was particularly affected by Wi-Fi issues. Apart from the natural interferences that occur in an event with a multitude of electronic

devices and wireless networks, we identified that the WiFi 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 feed are strongly delayed and with numerous drops. The smoothness of the image feed then increases later in the course, since the wireless signal is stronger. This also indicates that using more complex compression techniques for audio and video is important. In our case, the audio is not compressed, whereas the camera images are compressed at a constant rate. Given that the signal strength was varying 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 of all, it is inherently "unstable", consequently any malfunction or unexpected disturbance might cause the robot to fall. In the iCub3 particular case, the amount of exteroceptive sensors is reduced to the cameras and the Intel Realsense™ in the torso, limiting the possibility of the operator realizing if there are obstacles close to the robot. This might result in a problem in case of 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 non-continuous, 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 like LIDARs.

The hands represent another point of discussion. The iCub3 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 utilizing a heavy object, like a drill. The complexity of the hand made it very difficult to apply last-minute changes, where 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. In fact, we use 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 a logical, hardware, and software level.

Material and Methods

The outcomes presented in the Results are 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 interface, whereas a physical network establishes a logical link between the components of these two logical interfaces – see Fig. 7.

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 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 a 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 last fifteen 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 one meter – nor the robot actuation and transmission mechanisms – which never evolved for the robot to increase its dynamism substantially – nor its force sensing capabilities – which are derived from Force/Torque 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 humanoid that will be the starting point when conceptualising the next generations of the iCub platform.

Mechanics

The iCub3 humanoid robot is $125\,\mathrm{cm}$ tall, and weighs $52\,\mathrm{kg}$. Its mechanical structure is mainly composed by 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 $63\,\mathrm{cm}$ long, while the arms are $56\,\mathrm{cm}$ long from the shoulder to the fingertips. With the arms along the body, the robot is $43\,\mathrm{cm}$ wide. Each foot is composed of two separate rectangular sections, with a total length of about $25\,\mathrm{cm}$ and $10\,\mathrm{cm}$ wide.

The iCub3 robot possesses 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: 4 joints in the head controlling the eyelids and the eyes, 3 joints in the neck, 7 joints in each arm, 9 joints in each hand, 3 joints in the torso, 6 joints in each leg. The iCub3 hands are equipped

with tendon driven joints, moved by 9 motors, allowing to control separately the thumb, the index, and the middle finger, while the ring and the pinkie fingers move jointly (*53*).

Actuation

The 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 $110\,\mathrm{W}$, with a rated torque of $0.18\,\mathrm{N\,m}$, while the continuous stall torque is $0.22\,\mathrm{N\,m}$. The hip pitch, knee, and ankle pitch joints are driven by another type of brushless motor, whose rated power is $179\,\mathrm{W}$, the rated torque is $0.43\,\mathrm{N\,m}$ and the continuous stall torque is $0.48\,\mathrm{N\,m}$.

Power, Connectivity, Computation, and Electronics

The iCub3 robot is powered either by an external supplier or by a custom-made battery of $600\,\mathrm{W\,h}$. The connection to the robot can be established through an Ethernet cable or wirelessly via a standard 5GHz Wi-Fi network. The robot head is equipped with a 11th generation Intel® Core i7@$1.8\,\mathrm{GHz}$ computer with $16\,\mathrm{GB}$ of RAM and 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 $1\,\mathrm{kHz}$ and communicate via CAN protocol with the motor driver board (2FOC), which generates PWM signals at $20\,\mathrm{kHz}$;. The second are the MC4Plus

boards, controlling the DC motors.

Sensors

A particular feature of iCub3 is the vast array of sensors available. iCub3 possesses 8 six-axes force/torque (F/T) sensors (*52*) with integrated 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 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 possesses tactile sensors as an artificial skin (*56*) on the upper arm and the hands.

The head possesses 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 a NVIDIA® Jetson Xavier NX Module. The cameras are placed within the eyes bulb and can be controlled to 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 LEDs define 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 $21\,\mathrm{cm}$ taller, and weighs $19\,\mathrm{kg}$ more. Fig. 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

26

serial direct mechanisms, while 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, $10\,050\,\text{mA}\,\text{h}$ versus $9300\,\text{mA}\,\text{h}$, 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 the electronics point of view, both platforms share the same 2FOC/EMS/MC4Plus architecture, although iCub3 has higher resolution joint encoders, using 18 bits compared to 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, while 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 inner the layer, the shorter the time horizon used to evaluate the output. In addition, lower layers usually employ 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 inter-linked local area networks (LAN), 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, UDP, 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 sub-networks connected, we use OpenVPN (*58*). A simplified diagram of the robot and operator network is depicted in Fig. 7. The latency of introduced by the VPN can go from $5\,\mathrm{ms}$ 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 adopt 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 interfaces are responsible to process the operator motion to control the robot upper-body. The reference trajectories, fed to the robot controller presented in the "Robot

control" section, are computed using a multi-modal 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 inertial measurement unit (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 interfaces inverse kinematics algorithm" of the Supplementary Material. 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, while the body motion is controlled exclusively by using the orientation and velocity measurements provided by the nodes, whose data is acquired at $70\,\mathrm{Hz}$. This configuration and the corresponding mapping are presented in Figure S5A.

iFeel plus trackers   Since the IMUs 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 are 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 $90\,\mathrm{Hz}$, Figure S5B.

In addition, the operator's gaze and eye openness are tracked using the VIVE headset and facial tracker, allowing to control directly the robot eyelids and gaze (*65*). The SenseGlove haptic glove completes the set of devices of the manipulation interface. It is an exoskeleton-

29

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 takes care of detecting the operator's walking intention and commands the robot locomotion. We implement 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 motion direction is estimated via a moving ring attached to the harness secured to the operator's waist. The base plate can also be inclined of 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 Material, 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 thanks to the iFeel shoes, shown 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 to Virtualizer solution, the iFeel walking solution does not constrain the operator at a fixed point. This might disorient some novice operators, as the immersivity can affect their sense of equilibrium. Moreover, the operator could step away from the tracked area. In order to avoid this issue, a message is printed on the headset to suggest the operator to move back to the original location.

Voice and Face Expressions interfaces

The voice and face expressions 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 the non-verbal interaction. Thanks to the headset facial tracker, the operator's face expressions are replayed by the robot LEDs, Fig. 2(A and B).

The feedback interfaces

The feedback interfaces report the robot sensors measurement to the operator. In the iCub3 teleoperation system we have the following feedback interfaces: visual, auditory, haptic, touch. The headset is fundamental for the visual and auditory feedback. The images captured by the robot cameras are displayed inside the headset. At the same time, the audio captured by the

31

robot microphones is directly played on the headset's headphones. The SenseGlove™ haptic gloves provide touch feedback by means of vibration motors in each fingertip, on the back of the hand, and through a set of brakes that produce up to $20\,\text{N}$ of passive force per finger. The iFeel haptic nodes are fundamental for the body haptic feedback. They are used in two different ways: touch feedback and weight feedback.

Touch feedback The iFeel haptic nodes reproduce a touch occurring on the robot arm. The sensorized skin mounted on the robot arms detects a contact that is reproduced on the operator's arms through a vibration.

Weight feedback We use the iFeel haptic nodes also to retarget the effort endured by the robot arms. In particular, the haptic nodes modulate the vibration differently according to the amount of vertical force exerted on each robot arm.

Logging systems

We implemented two logging systems for two different purposes: online monitoring and offline processing. The online logging mechanism exploits the `openmct` (*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*). Figure S6(A) shows an example visualization of the GUI, with a live plot of the battery status of charge, and the communication delay to the robot PC.

The data streamed by the robot, together with some additional data coming from the walking controller, is 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 IHMC's SCS (*70*), `robot-log-visualizer` allows visualizing the data by simply clicking on the data of interest in the left

panel, as shown in Fig. S6(B). On the right panel, we have a 3D representation of the robot. If available, it is also possible to display a synchronized camera stream.

Supplementary Materials and Methods

The Supplementary Materials contain the mathematical derivations of the "Robot control" and the "Manipulation interfaces" in the sections named "Robot control layered architecture" and "Manipulation interfaces inverse kinematics algorithm", respectively. It also contains Figs. S1 to S6.

References

1. D. Leidner, "The covid-19 pandemic: an accelerator for the robotics industry?" IEEE Robotics & Automation Magazine, vol. 28, no. 1, pp. 116–116, 2021.

2. A. S. Pillai and G. Guazzaroni, Extended Reality Usage During COVID 19 Pandemic. Springer, 2022.

3. [Online]. Available: https://www.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," Advanced Robotics, vol. 35, no. 11, pp. 650–656, 2021.

6. Y. Horikawa, T. Miyashita, A. Utsumi, S. Nishimura, and S. Koizumi, "Cybernetic avatar platform for supporting social activities of all people," in 2023 IEEE/SICE International Symposium on System Integration (SII). IEEE, 2023, pp. 1–4.

7. T. Klamt, M. Schwarz, C. Lenz, L. Baccelliere, D. Buongiorno, T. Cichon, A. DiGuardo, D. Droeschel, M. Gabardi, M. Kamedula et al., "Remote mobile manipulation with the centauro robot: Full-body telepresence and autonomous operator assistance," Journal of Field Robotics, vol. 37, no. 5, pp. 889–919, 2020.

8. C. Lenz and S. Behnke, "Bimanual telemanipulation with force and haptic feedback through an anthropomorphic avatar system," Robotics and Autonomous Systems, vol. 161, p. 104338, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889022002275

9. M. Schwarz, C. Lenz, A. Rochow, M. Schreiber, and S. Behnke, "Nimbro avatar: Interactive immersive telepresence with force-feedback telemanipulation," in 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2021, pp. 5312–5319.

10. G. Lentini, A. Settimi, D. Caporale, M. Garabini, G. Grioli, L. Pallottino, M. G. Catalano, and A. Bicchi, "Alter-ego: a mobile robot with a functionally anthropomorphic upper body designed for physical interaction," IEEE Robotics & Automation Magazine, vol. 26, no. 4, pp. 94–107, 2019.

11. A. D. Dragan, K. C. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 2013, pp. 301–308.

12. K. Darvish, L. Penco, J. Ramos, R. Cisneros, J. Pratt, E. Yoshida, S. Ivaldi, and D. Pucci, "Teleoperation of humanoid robots: A survey," IEEE Transactions on Robotics, vol. 39, no. 3, pp. 1706–1727, 2023.

13. K. Darvish, Y. Tirupachuri, G. Romualdi, L. Rapetti, D. Ferigo, F. J. A. Chavez, and D. Pucci, "Whole-body geometric retargeting for humanoid robots," in 2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids). IEEE, 2019, pp. 679–686.

14. L. Penco, N. Scianca, V. Modugno, L. Lanari, G. Oriolo, and S. Ivaldi, "A multimode tele-operation framework for humanoid loco-manipulation: An application for the icub robot," IEEE Robotics & Automation Magazine, vol. 26, no. 4, pp. 73–82, 2019.

15. F. Abi-Farrajl, B. Henze, A. Werner, M. Panzirsch, C. Ott, and M. A. Roa, "Humanoid tele-operation using task-relevant haptic feedback," in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018, pp. 5010–5017.

16. M. Elobaid, Y. Hu, G. Romualdi, S. Dafarra, J. Babic, and D. Pucci, "Telexistence and teleoperation for walking humanoid robots," in Proceedings of SAI Intelligent Systems Conference. Springer, 2019, pp. 1106–1121.

17. D. Kim, B.-J. You, and 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. Eichermueller, G. Barbieri, E. Fowler, T. Hartikainen, L. Moss, and R. Walker, "Tactile telerobots for dull, dirty, dangerous, and inaccessible tasks," in 2020 IEEE International conference on robotics and automation (ICRA). IEEE, 2020, pp. 11 305–11 310.

19. A. Kaplish and 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, pp. 723–730.

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 et al., "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, pp. 2067–2072.

21. A. Rochow, M. Schwarz, M. Schreiber, and S. Behnke, "Vr facial animation for immersive telepresence avatars," in 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2022, pp. 2167–2174.

22. R. Luo, C. Wang, E. Schwarm, C. Keil, E. Mendoza, P. Kaveti, S. Alt, H. Singh, T. Padir, and 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, pp. 7726–7733.

23. U. Martinez-Hernandez, L. W. Boorman, and 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, and J. Park, "Stereoscopic low-latency vision system via ethernet network for humanoid teleoperation," in 2022 19th International Conference on Ubiquitous Robots (UR). IEEE, 2022, pp. 313–317.

25. J. C. Vaz, A. Dave, N. Kassai, N. Kosanovic, and P. Y. Oh, "Immersive auditory-visual real-time avatar system of ana avatar xprize finalist avatar-hubo," in 2022 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO). IEEE, 2022, pp. 1–6.

26. E. Schwarm, K. M. Gravesmill, and 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, pp. 7562–7568.

27. P. Schmaus, D. Leidner, T. Krüger, A. Schiele, B. Pleintinger, R. Bayer, and N. Y. Lii, "Preliminary insights from the meteron supvis justin space-robotics experiment," IEEE Robotics and Automation Letters, vol. 3, no. 4, pp. 3836–3843, 2018.

28. M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, A. Y. Ng et al., "Ros: an open-source robot operating system," in ICRA workshop on open source software. Kobe, Japan, 2009, p. 5.

29. P. Fitzpatrick, E. Ceseracciu, D. E. Domenichelli, A. Paikan, G. Metta, and L. Natale, "A middle way for robotics middleware," J. Softw. Eng. Robot, vol. 5, no. 2, pp. 42–49, 2014.

30. T. Tanioka, "Nursing and rehabilitative care of the elderly using humanoid robots," The Journal of Medical Investigation, vol. 66, no. 1.2, pp. 19–23, 2019.

31. F. Alnajjar, M. L. Cappuccio, O. Mubin, R. Arshad, and S. Shahid, "Humanoid robots and autistic children: A review on technological tools to assess social attention and engagement," International Journal of Humanoid Robotics, vol. 17, no. 06, p. 2030001, 2020.

32. A. Dave, J. C. Vaz, J. Kim, N. Kosanovic, N. Kassai, and P. Y. Oh, "Avatar-darwin a social humanoid with telepresence abilities aimed at embodied avatar systems," in 2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids).   IEEE, 2022, pp. 47–52.

33. A. Di Fava, K. Bouyarmane, K. Chappellet, E. Ruffaldi, and A. Kheddar, "Multi-contact motion retargeting from human to humanoid robot," in 2016 IEEE-RAS 16th international conference on humanoid robots (humanoids).   IEEE, 2016, pp. 1081–1086.

34. J. Ramos and S. Kim, "Humanoid dynamic synchronization through whole-body bilateral feedback teleoperation," IEEE Transactions on Robotics, vol. 34, no. 4, pp. 953–965, 2018.

35. Y. Ishiguro, K. Kojima, F. Sugai, S. Nozawa, Y. Kakiuchi, K. Okada, and M. Inaba, "High speed whole body dynamic motion experiment with real time master-slave humanoid robot system," in 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018, pp. 5835–5841.

36. Y. Ishiguro, T. Makabe, Y. Nagamatsu, Y. Kojio, K. Kojima, F. Sugai, Y. Kakiuchi, K. Okada, and M. Inaba, "Bilateral humanoid teleoperation system using whole-body exoskeleton cockpit tablis," IEEE Robotics and Automation Letters, vol. 5, no. 4, pp. 6419–6426, 2020.

37. A. Brygo, I. Sarakoglou, N. Garcia-Hernandez, and N. Tsagarakis, "Humanoid robot teleoperation with vibrotactile based balancing feedback," in Haptics: Neuroscience, Devices, Modeling, and Applications: 9th International Conference, EuroHaptics 2014, Versailles, France, June 24-26, 2014, Proceedings, Part II 9. Springer, 2014, pp. 266–275.

38. N. Y.-S. Lii, P. Schmaus, D. Leidner, T. Krueger, J. Grenouilleau, A. Pereira, A. Giuliano, A. S. Bauer, A. Köpken, F. S. Lay et al., "Introduction to surface avatar: the first heterogeneous robotic team to be commanded with scalable autonomy from the iss," in Proceedings of the International Astronautical Congress, IAC. International Astronautical Federation, IAF, 2022.

39. K. Hauser, E. Watson, J. Bae, J. Bankston, S. Behnke, B. Borgia, M. G. Catalano, S. Dafarra, J. B. Van Erp, T. Ferris, J. Fishel, G. Hoffman, S. Ivaldi, F. Kanehiro, A. Kheddar, G. Lannuzel, J. F. Morie, P. Naughton, S. NGuyen, P. Oh, T. Padir, J. Pippine, J. Park,

J. Vaz, D. Pucci, P. Whitney, P. Wu, and D. Locke, "Analysis and perspectives on the ana avatar xprize competition," International Journal of Social Robotics (Submitted), 2023.

40. [Online]. Available: https://www.labiennale.org/en

41. [Online]. Available: https://youtu.be/r6bFwUPStOA

42. [Online]. Available: https://en.wemakefuture.it/

43. [Online]. Available: https://youtu.be/FkBXzLaO7W0?t=4027

44. [Online]. Available: https://www.xprize.org/prizes/avatar/articles/38-semifinalist-teams-from-16-countries-aim-to-create-an-avatar-system

45. [Online]. Available: https://ieeetv.ieee.org/channels/ieee-future-directions/2022-ieee-telepresence-symposium-ana-avatar-xprize-overview-finals-recap-david-locke

46. [Online]. Available: https://youtu.be/lOnV1Go6Op0?t=4524

47. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, ser. NIPS'12, 2012, p. 1097–1105.

48. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

49. C. Willemse, S. Marchesi, and A. Wykowska, "Robot faces that follow gaze facilitate attentional engagement and increase their likeability," Frontiers in psychology, vol. 9, p. 70, 2018.

50. L. Natale, C. Bartolozzi, D. Pucci, A. Wykowska, and G. Metta, "icub: The not-yet-finished story of building a robot child," Science Robotics, vol. 2, no. 13, 2017.

51. [Online]. Available: https://icub-tech-iit.github.io/documentation/icub_versions

52. M. Fumagalli, S. Ivaldi, M. Randazzo, L. Natale, G. Metta, G. Sandini, and F. Nori, "Force feedback exploiting tactile and proximal force/torque sensing," Autonomous Robots, vol. 33, no. 4, pp. 381–398, 2012. [Online]. Available: http://dx.doi.org/10.1007/s10514-012-9291-2

53. A. Schmitz, U. Pattacini, F. Nori, L. Natale, G. Metta, and G. Sandini, "Design, realization and sensorization of the dexterous icub hand," in 2010 10th IEEE-RAS International Conference on Humanoid Robots.   IEEE, 2010, pp. 186–191.

54. [Online]. Available: https://icub-tech-iit.github.io/documentation/icub_wiring/icub3_x/

55. [Online]. Available: https://icub-tech-iit.github.io/documentation/icub_force_control/icub-force-control/

56. G. Cannata, M. Maggiali, G. Metta, and G. Sandini, "An embedded artificial skin for humanoid robots," in Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on, Aug 2008, pp. 434–438.

57. G. Romualdi, S. Dafarra, Y. Hu, and D. Pucci, "A benchmarking of dcm based architectures for position and velocity controlled walking of humanoid robots," in 2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids).   IEEE, 2018, pp. 1–9.

58. [Online]. Available: https://openvpn.net/

59. [Online]. Available: https://www.vive.com/eu/product/vive-pro-eye/overview/

60. [Online]. Available: https://www.vive.com/eu/accessory/facial-tracker/

61. [Online]. Available: https://www.vive.com/eu/accessory/tracker3/

62. [Online]. Available: https://www.senseglove.com/

63. [Online]. Available: https://www.cyberith.com/virtualizer-elite/

64. [Online]. Available: https://ifeeltech.eu/

65. R. Cambuzat, F. Elisei, G. Bailly, O. Simonin, and A. Spalanzani, "Immersive teleoperation of the eye gaze of social robots - assessing gaze-contingent control of vergence, yaw and pitch of robotic eyes," in ISR 2018; 50th International Symposium on Robotics, 2018, pp. 1–8.

66. [Online]. Available: https://nasa.github.io/openmct/

67. [Online]. Available: https://github.com/ami-iit/yarp-openmct

68. [Online]. Available: https://github.com/robotology/robometry

69. [Online]. Available: https://github.com/ami-iit/robot-log-visualizer

70. J. E. Pratt, B. Krupp, V. Ragusila, J. Rebula, T. Koolen, N. van Nieuwenhuizen, C. Shake, T. Craig, J. Taylor, G. Watkins et al., "The yobotics-ihmc lower body humanoid robot," in 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2009, pp. 410–411.

71. D. Flavigne, J. Pettrée, K. Mombaur, J.-P. P. Laumond, others, T. V. A. Truong, D. Flavigne, J. Pettré, K. Mombaur, and J.-P. P. Laumond, "Reactive synthesizing of human locomotion combining nonholonomic and holonomic behaviors," in Biomedical Robotics and

Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on. IEEE, 2010, pp. 632–637.

72. S. Dafarra, G. Nava, M. Charbonneau, N. Guedelha, F. Andradel, S. Traversaro, L. Fiorio, F. Romano, F. Nori, G. Metta, and D. Pucci, "A Control Architecture with Online Predictive Planning for Position and Torque Controlled Walking of Humanoid Robots," in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 10 2018, pp. 1–9. [Online]. Available: https://ieeexplore.ieee.org/document/8594277/

73. D. Pucci, L. Marchetti, and P. Morin, "Nonlinear control of unicycle-like robots for person following," in Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, 2013, pp. 3406–3411.

74. M. Vukobratović and B. Borovac, "Zero-moment point—thirty five years of its life," International journal of humanoid robotics, vol. 1, no. 01, pp. 157–173, 2004.

75. J. Englsberger, T. Koolen, S. Bertrand, J. Pratt, C. Ott, and A. Albu-Schaffer, "Trajectory generation for continuous leg forces during double support and heel-to-toe shift based on divergent component of motion," in 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 9 2014, pp. 4022–4029. [Online]. Available: http://ieeexplore.ieee.org/document/6943128/

76. S. Kajita, F. Kanehiro, K. Kaneko, K. Yokoi, and H. Hirukawa, "The 3D linear inverted pendulum mode: a simple modeling for a biped walking pattern generation," in Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No.01CH37180). IEEE, 2001, pp. 239–246. [Online]. Available: http://ieeexplore.ieee.org/document/973365/

77. Y. Choi, D. Kim, Y. Oh, and B.-j. J. You, "On the Walking Control for Humanoid Robot Based on Kinematic Resolution of CoM Jacobian With Embedd ed Motion," Proceedings of the 2006 IEEE International Conference on Robotics and Automation, vol. 23, no. 6, pp. 1285–1293, 2007.

78. D. E. Orin and A. Goswami, "Centroidal momentum matrix of a humanoid robot: Structure and properties," Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on, pp. 653 – 659, 2008.

79. [Online]. Available: https://github.com/ami-iit/bipedal-locomotion-framework/tree/v0.11.1/src/IK

80. [Online]. Available: https://github.com/robotology/osqp-eigen

81. B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd, "OSQP: An Operator Splitting Solver for Quadratic Programs," 2018 UKACC 12th International Conference on Control, CONTROL 2018, p. 339, 10 2018.

82. L. Rapetti, Y. Tirupachuri, K. Darvish, S. Dafarra, G. Nava, C. Latella, and D. Pucci, "Model-based real-time motion tracking using dynamical inverse kinematics," Algorithms, vol. 13, no. 10, p. 266, 2020.

Acknowledgments

Data and materials availability: All data needed to support the conclusions of this manuscript are included in the main text and Supplementary Materials. The scripts to generate Figures 3(C-D) and 4(D-H) are provided as separate zip file.

Figures and Tables

Figure 1: **iCub3 explores the Biennale di Venezia**. Snapshots of the video (*41*) demonstrating the remote teleoperation of iCub3 at the Italian Pavillion of the Biennale di Venezia. The operator navigates the remote venue via iCub3 (**A**, **B**). The operator controls the iCub3 eyelids in response to bright light (**C**). The operator remotely grasps a piece of tissue through iCub3 (**D**). The robot is touched on the arm (**E**). The robot skin, whose activation is represented in (**F**), triggers the body haptic feedback on the operator.

Figure 2: **Face expressions retargeting and iFeel walking**. Examples of the retargeting interfaces. With the facial tracker (**A**), the operator can directly control the emotions displayed by the robot (**B**). The iFeel shoes (**C**). They measure the force and torque exchanged by the operator with the ground. When paired with a set of trackers, it is also possible to detect their position. An example of the intention mechanism used for the locomotion interface (**D**).

Figure 3: **iCub3 on the stage of the "We Make Future" Festival**. iCub3 is on the stage of a tech fair, while being teleoperated from remote (*43*). iCub3 interacts with the audience (**A**). iCub3 holding a box (**B**). Plot of the center of mass tracking of the robot while walking with a box (**C**). Plot of the vertical force measured on the robot arms while walking. A recipient hands the robot the box at around $10\,\mathrm{s}$. Then, the box is taken back by the recipient after the robot stopped walking (**D**). iCub3 interacts with the recipient (**E**).

Figure 4: **iCub3 at the XPrize semifinals**. Pictures and plots of the iCub3 avatar system performance at the Xprize semifinals. iCub3 manipulating one puzzle piece (**A**). iCub3 checking the texture of the vase (**B**). iCub3 walking during the XPrize semifinals (**C**).CoM tracking during the walking task (**D**). The left and right hand Cartesian errors during the puzzle task (**E**) and (**F**). Magnified version of (**E**) to highlight the Cartesian tracking and lags (**G**). Normal force measured by the hands while lifting the vases and putting it back in place (**H**).

49

Figure 5: **iCub3 at the XPrize finals**. Pictures of the iCub3 avatar system performance at the XPrize finals. The first three pictures have been shot by the authors while on the finals course. The last three pictures have been taken during tests on the lab prior to the finals. iCub3 walking on the course, themed on the exploration of another planet (**A**). iCub3 activating the switch using a plastic cylinder installed on the wrist (**B**). A portion of a video of the iCub3 finals trial, while hitting the door. The Operator's view is visible in the bottom right corner (**C**). iCub3 activating the drill (**D**). iCub3 holding the XPrize finals canisters (**E**). iCub3 making contact with a rough textured rock via the hand palm skin (**F**).

Figure 6: **iCub3 upperbody and comparison with the previous iCub version**. iCub3 differs from its predecessors, being taller and heavier. iCub3 upperbody (**A**). Compared to the previous versions, the iCub3 shoulders and torso do not use any tendon-driven mechanism. The iCub3 robot compared to the iCub versions 1.0-2.5 (**B**).

Figure 7: **The full architecture**. The avatar architecture, comprising the operator, the delayed network, and the avatar. The operator skills are retargeted to the robot through the control architecture, and the operator receives feedback due to the robot sensor measurements.

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 someone that has deep knowledge of the avatar system, while a Naive operator has to be trained before the beginning of the validation – the training time is fixed at about thirty minutes. We also categorize the validations according to the skills required on 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 the Methods section.

Validation	Location		Operator		Locomotion Capabilities			Interaction with Recipient			Manipulation		Objectives	Tracking System	Body Haptic Feedback	Locomotion Retargeting
	Local	Remote	Expert	Naive	Short Distance	Long Distance	Side Motions	Verbal	Non-Verbal	Physical	Coarse	Precise				
Italian Pavillion		✓	✓				✓	✓	✓	✓	✓		Physical and non-verbal 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 loco-manipulation	iFeel + Trackers	Weight Feedback	iFeel Walking

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
The Avatar walks about 5 meters to a designated spot, allowing the Operator to communicate with the Mission Commander, who explains the mission.
The Avatar walks about 5 meters and activates a switch that opens the station door.
The Avatar walks about 30 meters 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 meters between obstacles, 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.

iCub3 Avatar System:
Enabling Remote Fully-Immersive Embodiment of Humanoid Robots

Supplementary Material

Stefano Dafarra[*,1], Ugo Pattacini[2], Giulio Romualdi[1], Lorenzo Rapetti[1],
Riccardo Grieco[1,6], Kourosh Darvish[1,6], Gianluca Milani[1], Enrico Valli[1],
Ines Sorrentino[1,3], Paolo Maria Viceconte[1,4],
Alessandro Scalzo[2], Silvio Traversaro[1],
Carlotta Sartore[1,3], Mohamed Elobaid[1], Nuno Guedelha[1], Connor Herron[5],
Alexander Leonessa[5], Francesco Draicchio[7], Giorgio Metta[2],
Marco Maggiali[2], Daniele Pucci[1,3]

[1] Artificial and Mechanical Intelligence, Istituto Italiano di Tecnologia, Genoa, Italy,

[2] iCub Tech Facility, Istituto Italiano di Tecnologia, Genoa, Italy

[3] School of Computer Science, University of Manchester, Manchester, UK

[4] DIAG, Sapienza Università di Roma, Rome, Italy

[5] Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA, USA

[6] Computer Science and Robotics Institute, University of Toronto, Toronto, Canada

[7] DiMEILA, Istituto Nazionale Assicurazione Infortuni sul Lavoro (INAIL), Rome, Italy

(Corresponding author contact: `stefano.dafarra@iit.it`)

1

Robot control layered architecture

The layered control architecture mentioned in the "Robot control" section of the "Methods" is composed of three layers. From top to bottom, the layers are here called: trajectory optimization, simplified model control, and whole-body quadratic programming (QP) control – Figure S4.

Trajectory optimization The trajectory optimization layer aims to compute a sequence of contacts' location and timings. This layer often takes advantage of optimization techniques to consider the feasibility of the contact location. The simpler the model, the simpler the problem. For instance, flat terrain allows one to model the robot as a simple unicycle (*71*) which enables fast solutions to the optimization problem for the walking pattern generation (*72*).

The unicycle model is described by the following equations (*73*):

$$\dot{x}_u = v_u R_2(\theta_u) e_1, \tag{S1a}$$
$$\dot{\theta}_u = \omega_u, \tag{S1b}$$

with $v_u \in \mathbb{R}$ and $\omega_u \in \mathbb{R}$ the unicycle's rolling and rotational velocity, respectively. $x_u \in \mathbb{R}^2$ is the unicycle position in the inertial frame $\mathcal{I}$, while $\theta_u \in \mathbb{R}$ represents the angle around the $z-$axis of $\mathcal{I}$ which aligns the inertial reference frame with a unicycle fixed frame. $R_2(\theta) \in SO(2)$ is the rotation matrix of an angle $\theta \in \mathbb{R}$ in a 2D plane, while $e_1 = [1, 0]^\top$.

A possible control objective for this kind of model is to asymptotically stabilize a point F rigidly attached on the unicycle about a desired point F^* whose position is x_F^*. Define the error $\tilde{x}$ as

$$\tilde{x}_u := x_F - x_F^* = x_u + R_2(\theta_u) d_u - x_F^*, \tag{S2}$$

where $d_u \in \mathbb{R}^2$ is the position of F in the unicycle frame. The following control law makes the

origin of the error dynamics an asymptotically stable equilibrium (*72*):

$$\begin{bmatrix} v_u \\ \omega_u \end{bmatrix} = \begin{bmatrix} R_2(\theta_u)e_1 & R_2(\theta_u + \pi/2)d_u \end{bmatrix}^{-1} (\dot{x}_F^* - K_u \tilde{x}_u), \tag{S3}$$

with K_u a positive definite matrix.

In our context, the humanoid robot feet are represented by the unicycle wheels, and the desired footsteps can be obtained by sampling the unicycle trajectories given a desired trajectory for the point F. The generation of footsteps via the unicycle model allows for planning walking motions using a simple two-dimensional quantity. On the other hand, the unicycle model of Eq.(S1) does not allow motion along the wheel axis. In other words, the robot would not be able to perform lateral steps. To circumvent this limitation, we modify Eq. (S1a) as follows:

$$\dot{x}_u = R_2(\theta_u) \begin{bmatrix} v_u \\ l_u \end{bmatrix}, \tag{S4}$$

where $l_u \in \mathbb{R}$ is an additional control input enabling the unicycle side motions. Consequently, rather than employing the control law presented in Eq. (S3), we directly define the three control inputs v_u, ω_u, and l_u according to the desired robot motion.

Once the footsteps are planned, the desired feet trajectory is obtained by cubic spline interpolation. Assuming a constant height of the center of mass while walking and a constant angular momentum, we plan the center-of-mass (CoM) and zero moment point (ZMP) (*74*) trajectory through the algebric Divergent Component of Motion generator (*57, 75*).

Simplified model control   The output of the trajectory optimization layer feeds the simplified model control layer which is responsible for finding feasible robot center-of-mass (CoM) trajectories. Given a desired CoM $x_{\text{CoM}}^{\text{ref}}$ and ZMP $x_{\text{ZMP}}^{\text{ref}}$ position to stabilize, we develop a control law that approximates the robot dynamics via the Linear Inverted Pendulum Model (LIPM) (*76*) following (*77*):

$$\dot{x}_{\text{CoM}}^* = \dot{x}_{\text{CoM}}^{\text{ref}} - K_{\text{ZMP}}(x_{\text{ZMP}}^{\text{ref}} - x_{\text{ZMP}}) + K_{\text{CoM}}(x_{\text{CoM}}^{\text{ref}} - x_{\text{CoM}}), \tag{S5}$$

where $K_{\text{CoM}} - \zeta I_2$ and K_{ZMP} are positive definite, while $K_{\text{ZMP}} - \zeta I_2$ is negative definite, with $\zeta \in \mathbb{R}_{>0}$.

Whole-Body control   The whole-body control layer generates joint position references for the robot. The proposed controller evaluates the generalized robot velocity $\nu \in \mathbb{R}^{n_s+n_b}$ where n_s is the number of the robot joints and $n_b = 6$ represents the degrees of freedom associated with the floating base. Recalling that the velocity of a link L depends linearly on ν employing the Jacobian J_L, we define a set of tasks $\Psi_{L_{\text{SE(3)}}}$ of the form

$$\Psi_{L_{\text{SE(3)}}} = \mathrm{v}_L^* - J_L \nu, \tag{S6}$$

where v_L^* is the desired velocity chosen to guarantee the tracking of the reference link pose (*57*). The SO(3) task, $\Psi_{L_{\text{SO(3)}}}$, ensures the convergence of a frame orientation to a desired orientation by selecting the appropriate rows from Eq.(S6).

While teleoperating, we always require the center of mass x_{CoM} to remain in a given position:

$$\Psi_{\text{CoM}} = \dot{x}_{\text{CoM}}^* - J_{\text{CoM}} \nu, \tag{S7}$$

where $\dot{x}_{\text{CoM}}^*$ is the desired CoM velocity chosen to guarantee the convergence of the CoM to a given trajectory, while J_{CoM} is the linear component of the Centroidal Momentum matrix scaled by the total mass of the robot (*78*).

To consider the desired robot joint positions provided by the retargeting system, we introduce a regularization task for the joint variables. The task is achieved by asking for desired joint velocities that depend on the error between the desired and measured joint values, such as:

$$\Psi_s = \dot{s}^* - \begin{bmatrix} 0_{n_s \times 6} & I_n \end{bmatrix} \nu, \tag{S8}$$

where n_s is the robot actuated degrees of freedom and $\dot{s}^*$ guarantees the tracking of the joint reference obtained by the algorithms presented in the "Manipulation interfaces" section.

4

The tracking of the left and right foot poses is considered as high-priority $\mathrm{SE}(3)$ tasks, Eq. (S6), that are denoted as $\Psi_{L_{\mathrm{SE}(3)}}$ and $\Psi_{R_{\mathrm{SE}(3)}}$, respectively. We take into consideration the CoM tracking as a high-priority task, Eq. (S7). The torso orientation is considered as a low-priority task $\mathrm{SO}(3)$ task and we denote it with $\Psi_{T_{\mathrm{SO}(3)}}$. The retargeting joint positions tracking is considered as a low-priority regularization task, Eq. (S8), and denoted as $\Psi_{s_{\mathrm{ret}}}$. Furthermore, the joint postural condition, Eq. (S8), is also enforced while walking as a low-priority task, $\Psi_{s_{\mathrm{reg}}}$. The above hierarchical control objectives is framed into a whole-body optimization problem:

$$
\underset{\nu}{\mathrm{minimize}}\ \ \Psi_{T_{\mathrm{SO}(3)}}^{\top}\Lambda_T\Psi_{T_{\mathrm{SO}(3)}} + \Psi_{s_{\mathrm{reg}}}^{\top}\Lambda_{s_{\mathrm{reg}}}\Psi_{s_{\mathrm{reg}}} + \Psi_{s_{\mathrm{ret}}}^{\top}\Lambda_{s_{\mathrm{ret}}}\Psi_{s_{\mathrm{ret}}} \tag{S9a}
$$
$$
\text{subject to}\ \Psi_{L_{\mathrm{SE}(3)}} = 0 \tag{S9b}
$$
$$
\Psi_{R_{\mathrm{SE}(3)}} = 0 \tag{S9c}
$$
$$
\Psi_{\mathrm{CoM}} = 0 \tag{S9d}
$$

The performance of Eq. (S9) depends on the choice of the weights $\Lambda_{s_{\mathrm{reg}}}$ and $\Lambda_{s_{\mathrm{ret}}}$. In particular, we observe that the weights achieving good embodiment during standing and walking are not the same. For this reason, we implement a gain-scheduling technique depending on whether the robot is walking or standing. The transition between the two sets of weights is smooth with a minimum acceleration trajectory.

Since the decision variable is the robot velocity ν and the tasks depend linearly on ν, we transcribe the optimization problem into a quadratic programming (QP) problem, and we solve it via off-the-shelf solvers. The transcription is achieved through the Inverse Kinematics implemented in the `bipedal-locomotion-framework` library (*79*). The QP problem is solved by means of `osqp-eigen` (*80*) a `C++` wrapper for OSQP (*81*).

Manipulation interface inverse kinematics algorithm

The manipulation interfaces use the wearable sensor's measurement altogether as inputs for a constrained inverse kinematics algorithm (*82*), mapping the motion into the robot model fol-

lowing the geometric retargeting approach presented in (*13*). Depending on the measurement type, the desired model link velocities are defined as follows. For a position $\hat{p} \in \mathbb{R}^3$ and/or a linear velocity measurement $\hat{v} \in \mathbb{R}^3$, the corrected link linear velocity is computed as

$$v^* = \hat{v} + K_p \left(\hat{p} - p \right). \tag{S10}$$

Instead, for an orientation $\hat{R} \in \mathrm{SO}(3)$ and/or an angular velocity measurement $\hat{\omega} \in \mathbb{R}^3$, the corrected link angular velocity is computed as

$$\omega^* = \hat{\omega} + K_R \, \mathrm{sk}(R^T \hat{R})^\vee, \tag{S11}$$

where $\mathrm{sk}(.)^\vee : \mathrm{SO}(3) \to \mathbb{R}^3$ is the operator that extracts the skew-symmetric part of the matrix and applies the inverse of the skew operator. Finally, for a gravity versor measurement $\hat{u}_g \in \mathbb{R}^3, ||\hat{u}_g|| = 1$, the corrected link angular velocity is computed as

$$\omega^* = R(\hat{u}_g \times u_g). \tag{S12}$$

The corrected link velocities compensate for the measurement error and are achieved by means of the same task formulation presented in Eq.(S6). In particular, all the corrected link velocities and the respective Jacobians are vertically concatenated into v^* and J respectively,

$$\Psi = v^* - J\nu. \tag{S13}$$

Then, we formulate a constrained inverse differential kinematics optimization:

$$\nu^* = \underset{\nu}{\mathrm{minimize}} \; \Psi^\top \Lambda \Psi \tag{S14}$$

$$\mathrm{subject\ to}\ A\nu \leq b, \tag{S15}$$

with Λ being a weight matrix. A and b are defined to ensure both velocity and configuration constraints using the joint limit avoidance approach described in (*82*). The output system velocity ν^* is integrated to obtain the model configuration q. The joint limit avoidance mechanism and the integration passage naturally filter disturbances coming from the input sensors. In fact,

6

the former limits the maximum model joint velocities, while the latter naturally filters high-frequency noise. At the same time, this behavior introduces some lag in the motion retargeting, hence it is necessary to tune λ to trade-off between reactivity and data filtering.

Supplementary figures

Figure S1: **The puzzle used during the XPrize semifinals**. Each puzzle piece has a knob that helps their grasping.

Figure S2: **A representation of the ANA Avatar XPrize finals course**. It is themed on the exploration of another planet. The labels indicate how the tasks listed in Table 2 were executed on the course.
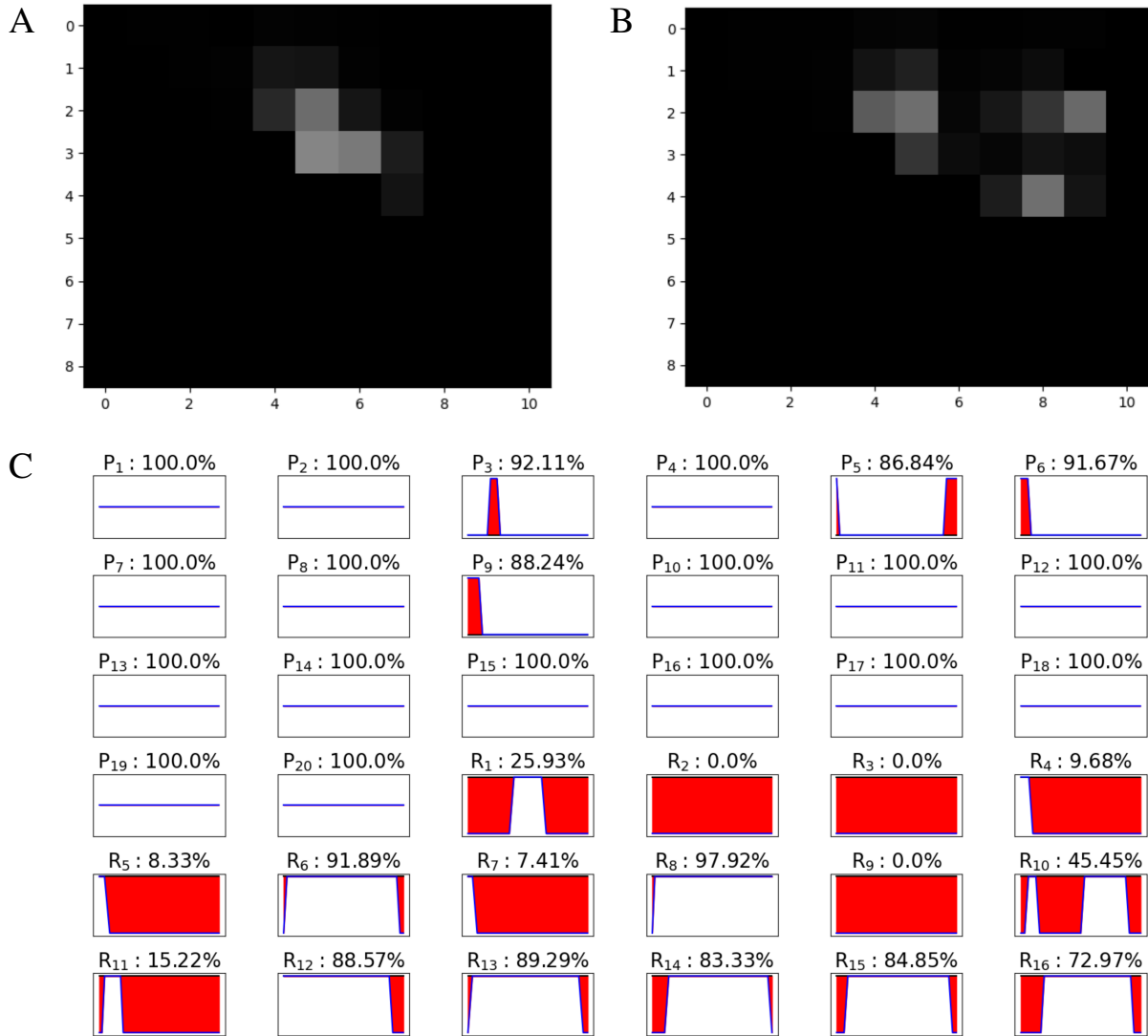
Figure S3: **Performance of the classifier for the XPrize finals texture task**. The images represent the input and the performance of the neural network used to identify the texture of rocks. Grayscale images extracted from the sensorized robot palm in contact with a plain (**A**) and a rough (**B**) rock. The higher the pressure measured by each tactile sensor, the whiter the correspondent pixel. The active tactile sensors are more sparse during contact with a rough rock. The rock classifier's performances on the 36 contacts of the test set (**C**). Each plot corresponds to a separate contact, either plain P_i or rough R_i, and shows a comparison of the ground truth and the predicted class for the entire duration of the contact. When the prediction and the ground truth do not coincide, the gap is filled in red to highlight the error. Although the misclassification rate increases for rough contacts, the overall accuracy on the test set, more precisely the average of the per-contact accuracies labeling each plot, reaches 78%.
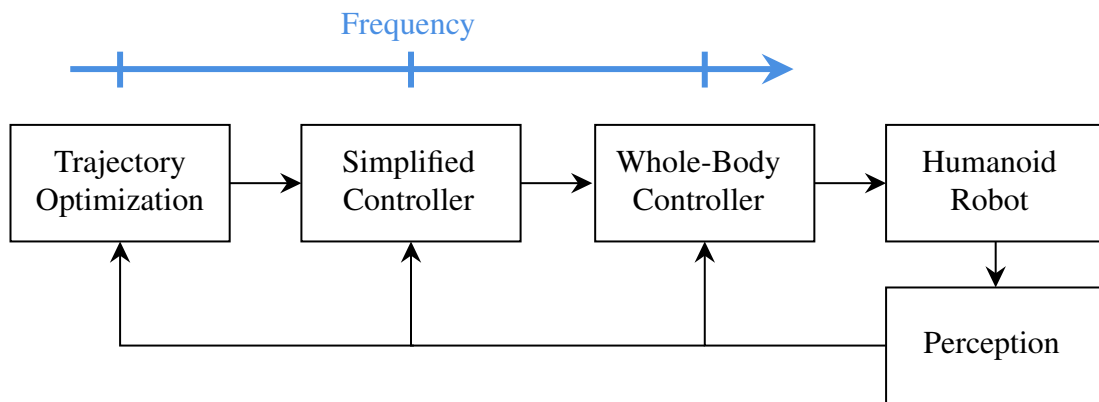
Figure S4: **The three layer controller architecture**. The inner the layer, the higher the frequency. Each layer gathers the outcome of the outer layer, the information from the robot through the perception block, and generates the references for the inner layer.
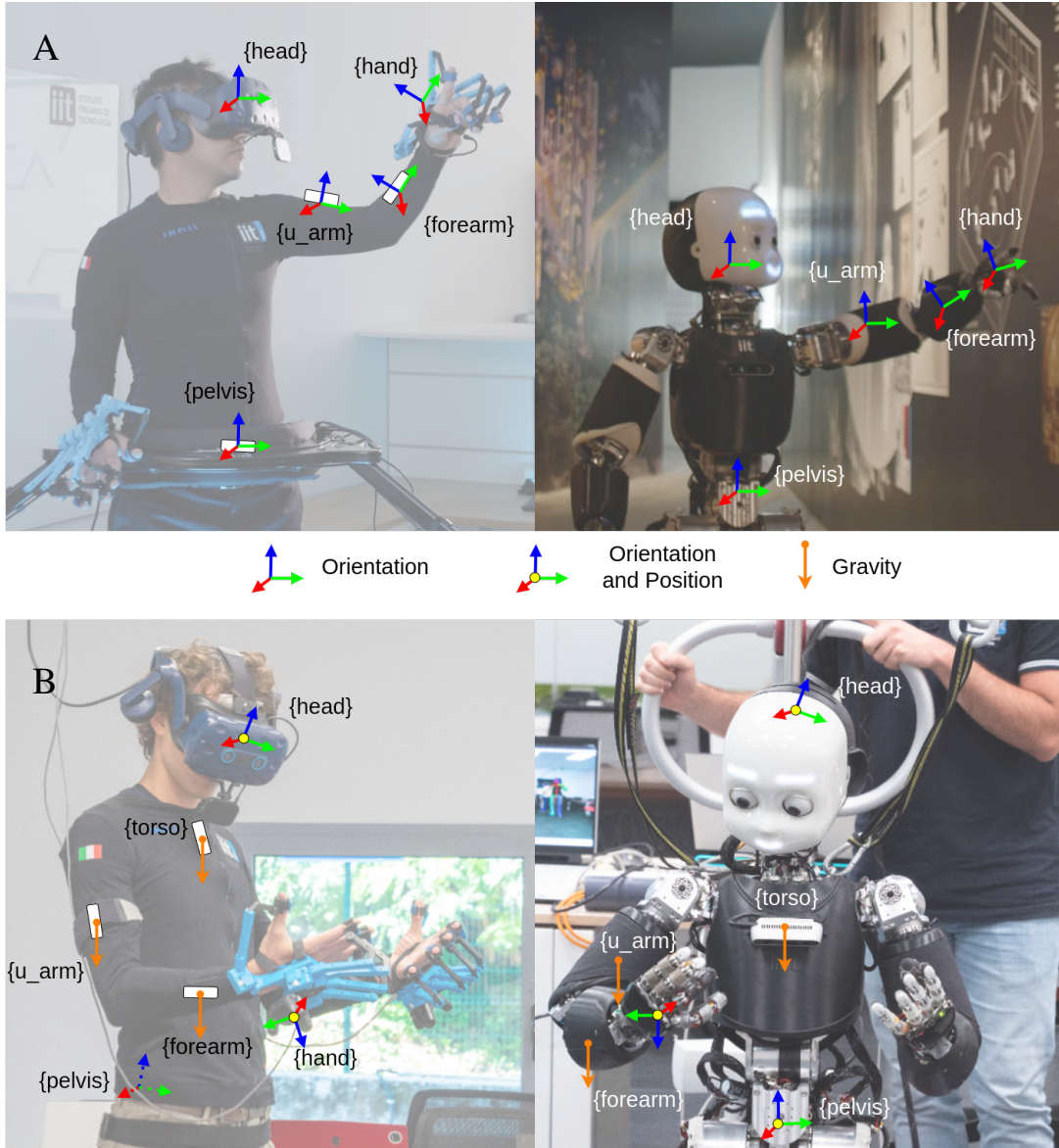
Figure S5: **Manipulation interfaces**. The operator movements are tracked by the robot through the estimation of a series of "targets", as explained in the "Manipulation interfaces" section. Upper body motion retargeting using in the "iFeel only" configuration (**A**). Upper body motion retargeting using iFeel nodes, headset, and trackers, corresponding to the "iFeel plus trackers" configuration (**B**). In both cases, the same sensor configuration applies to both arms, hence we show the configuration of one arm only.
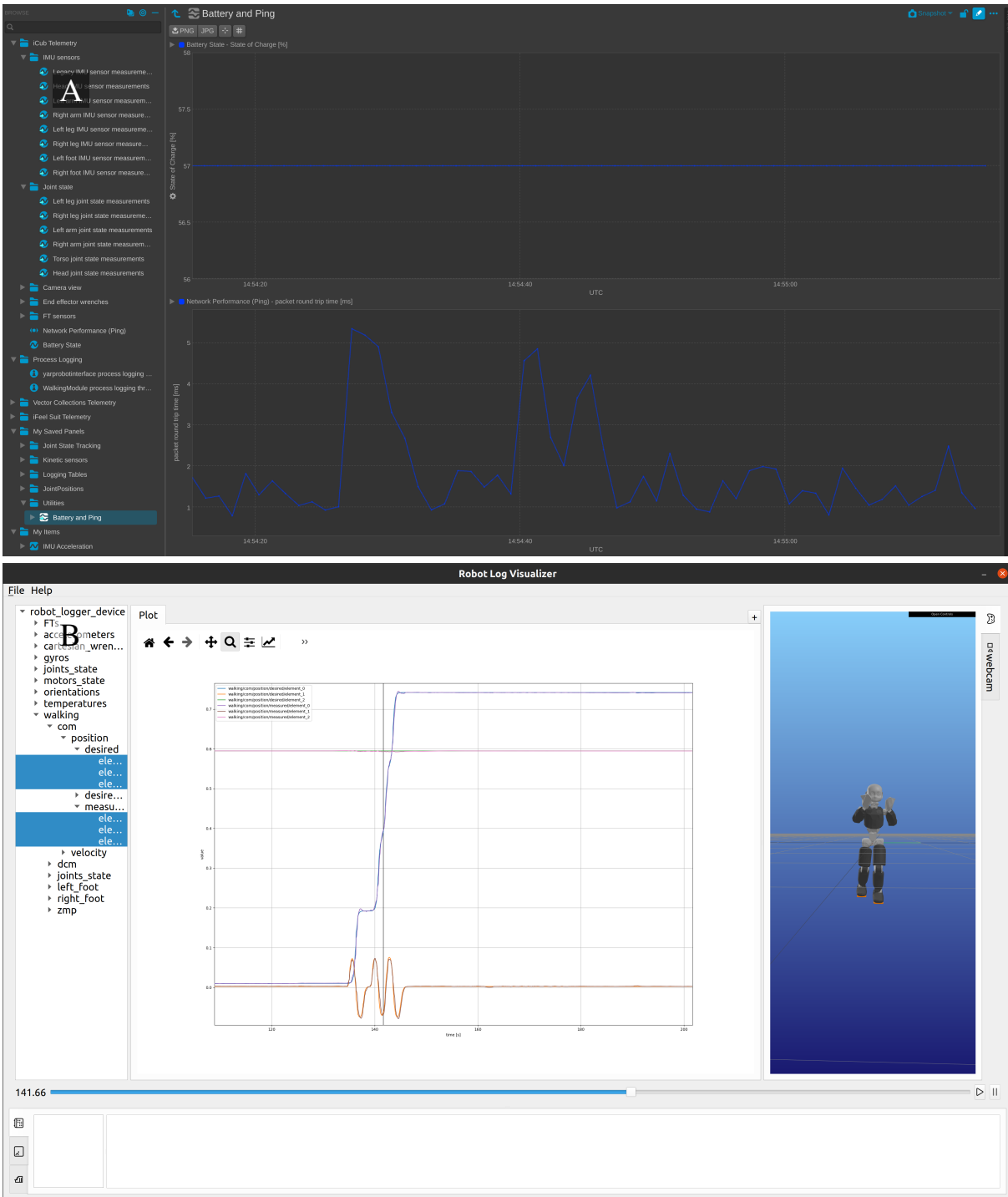
Figure S6: **Logging systems**. Examples of usage of the logging systems presented in the "Methods" section. An example of the online logging system showing the battery level and the robot's head communication delay (**A**). The offline logging system displaying a representation of the robot and a plot (**B**).