A cybernetic avatar system to embody human telepresence for connectivity, exploration, and skill transfer

Rafael Cisneros-Limón, Antonin Dallard, Mehdi Benallegue, Kenji Kaneko, Hiroshi Kaminaga, Pierre Gergondet, Arnaud Tanguy, Rohan Pratap Singh, Leyuan Sun, Yang Chen, et al.

A cybernetic avatar system to embody human telepresence for connectivity, exploration, and skill transfer

Rafael Cisneros-Limón[1*], Antonin Dallard[2,1], Mehdi Benallegue[1], Kenji Kaneko[1], Hiroshi Kaminaga[1], Pierre Gergondet[1], Arnaud Tanguy[2], Rohan Pratap Singh[1], Leyuan Sun[1], Yang Chen[1,3], Carole Fournier[2], Guillaume Lorthioir[1], Masato Tsuru[1], Sélim Chefchaouni-Moussaoui[1], Yukiko Osawa[4], Guillaume Caron[1], Kevin Chappellet[1], Mitsuharu Morisawa[1], Adrien Escande[1], Ko Ayusawa[1], Younes Houhou[1], Iori Kumagai[1], Michio Ono[5], Koji Shirasaka[5], Shiryu Wada[5], Hiroshi Wada[5], Fumio Kanehiro[1] and Abderrahmane Kheddar[1,2]

[1]CNRS-AIST JRL (Joint Robotics Laboratory), UMI3218/IRL, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan.
[2]CNRS-University of Montpellier, Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), Montpellier, France.
[3]School of Integrative & Global Majors (SIGMA), Univ. of Tsukuba, Tsukuba, Japan.
[4]Industrial Cyber-Physical Systems Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan.
[5]Double R&D, Zama, Japan.

*Corresponding author(s). E-mail(s): rafael.cisneros@aist.go.jp;
Contributing authors: antonin.dallard@gadz.org; mehdi.benallegue@aist.go.jp; k.kaneko@aist.go.jp; hiroshi.kaminaga@aist.go.jp; pierre.gergondet@aist.go.jp; arnaud.tanguy@lirmm.fr; rohan-singh@aist.go.jp; son.leyuansun@aist.go.jp; chenyang@ai.iit.tsukuba.ac.jp; carole.fournier@lirmm.fr; guillaume.lorthioir@aist.go.jp; m.tsuru@aist.go.jp; selim.cm@mailo.com; yukiko.osawa-akiyama@aist.go.jp; guillaume.caron@aist.go.jp; chappellet.kevin@gmail.com; m.morisawa@aist.go.jp; adrien.escande@inria.fr; k.ayusawa@aist.go.jp; younes.houhou@etu.sorbonne-universite.fr; iori-kumagai@aist.go.jp; m-ono@j-d.co.jp; k-shirasaka@j-d.co.jp; s-wada@j-d.co.jp; pwada@j-d.co.jp; f-kanehiro@aist.go.jp; kheddar@lirmm.fr;

Abstract

This paper describes the cybernetic avatar system developed by Team JANUS for connectivity, exploration, and skill transfer: the core domains targeted by the ANA Avatar XPRIZE competition, for which Team JANUS was a finalist. We used as an avatar a humanoid robot with a human-like appearance and shape that is capable of reproducing facial expressions and walking, and built an avatar control system that allowed the operator to control the avatar through equivalent mechanisms of motion; that is, by replicating the upper-body movement with naturalness and by stepping to command locomotion. In this way, we aimed to achieve high-fidelity telepresence and managed to be well evaluated from the point of view of the operator during the competition. We introduce our solutions to the integration challenges and present experimental results to asses our avatar system, together with current limitations and how we are planning to mitigate them in future work.

1 Introduction

We recently witnessed an exceptional health situation worldwide. Our lifestyle has changed since the start of the COVID-19 pandemic. The latter impacted the lifestyles and habits of most of our daily tasks that were usually done in physical presence. Before, experts could easily travel to a very remote location where their specific know-how was needed. Ordinary people were able to visit their loved ones at will. Now, the situation is no longer the same.

Moreover, recent awareness and alarm on climate change and pollution put high constraints on reducing Carbon emissions and subsequent print over the planet. Medical doctors, engineers, researchers, and other professionals now need to find an alternative to deal more efficiently with such situations that are highly likely to be more critical in the future. High-fidelity telepresence and beaming of robotic avatars is envisioned as one of the technologies to mitigate the impact on our daily life, should such a catastrophic situation occur again [1].

Telepresence and its variants, such as telexistence, are concepts dating to the early 1980s and have since then been relevant to many applications. For example, early telexistence is a concept that enables humans to virtually exist in another location where they can act freely in view of a more ecological and time-efficient society with an overall improved work-life balance [2]. The modern mutation is what is known as the Metaverse! Achieving truly immersive telepresence would enable an operator to transport their senses, presence, and skills to a remote location where it is not feasible to travel due to restrictions, efficiency, or just because the remote location is a dangerous place to go (e.g., planetary exploration). Skill transfer can also allow companies to operate 24 hours without having workers performing night shifts on tasks that require physical skills with remote workers in different time zones. It can also allow people to work from home and support their family if they cannot go out (e.g., due to some disability, sickness, etc.) [2]. Other relevant applications of telepresence can be the use of avatars for telecare or telenursery applications [3], for telesurgery [4], or in education [5]. Telecare can provide the required care to patients with a highly transmissible disease that could impose a risk on a nurse or a family member who would like to pay a visit. Telesurgery is also highly relevant in the case of an emergency that requires an expert surgeon that is in a remote location.

Coincidentally with the recent worldwide situation, the XPRIZE foundation launched a challenge in March 2018: ANA Avatar XPRIZE[1], a multi-year international competition that ended in November 2022. The purpose of the competition was to integrate multiple emerging technologies to develop a physical, non-autonomous avatar system to deploy senses, actions, and presence to a remote location in real-time in a manner that feels as if you are genuinely there.

From all the teams initially registered for the competition, 77 teams from 19 countries were selected as qualified teams in January 2020. After submitting materials that demonstrated enough capabilities of their avatar system, 38 teams from 16 countries were selected as semifinalists in April 2021. The Semifinals testing took place in two parts (in September 2021 and March 2022) due to the travel restrictions that some teams suffered, and from that testing, 20 teams from 11 countries were selected in May 2022 to participate in the Finals testing. During the Finals testing in November 2022, there was a qualification run before the actual testing, for which 17 teams from 10 countries qualified as finalists. The prize was ultimately given to the 1st, 2nd and 3rd places.

Our team, JANUS[2], was among the 17 teams that qualified for the finals in the ANA Avatar XPRIZE competition. JANUS is a bi-located team that gathers expertise from the Japanese National Institute of Advanced Industrial Science and Technology (AIST) and the French National Center for Scientific Research (CNRS), namely from CNRS-AIST Joint Robotics Laboratory (JRL) in Tsukuba[3], and the Interactive Digital Human Laboratory at LIRMM in Montpellier[4]. The team comprises researchers and Ph.D. students from both laboratories, gathering members from several citizenships: Japan, France, Mexico, Algeria, China, and India. It also involves as a partner the company Double R&D[5] [6].

[1] https://www.xprize.org/prizes/avatar
[2] https://unit.aist.go.jp/jrl-22022/en/projects/janus/team-janus.html
[3] https://unit.aist.go.jp/jrl-22022/en
[4] https://www.lirmm.fr/teams-en/IDH-en
[5] https://j-d.co.jp

This paper describes the cybernetic avatar system developed by Team JANUS to use for connectivity, exploration, and skill transfer: the core domains targeted by the ANA Avatar XPRIZE competition. Our main contributions are:

- The effort placed in updating and improving a decade-old humanoid robot to meet the specifications required by the competition without sacrificing the bipedal challenge nor the anthropomorphic shape of the robot.
- The design and development of light yet dexterous underactuated hands capable of performing power and precision grasping.
- The development of enhanced visual feedback consisting of decoupling the visual feedback between the operator's and the robot's head.
- The design of a button-less operator interface that uses voice, gaze, head motion, and stepping in place to control the manipulation and locomotion of the robot.
- The development of a hierarchic inequality admittance control that limits the maximum force the robot can apply on the environment without disturbing the user control, providing safety guarantees.

These contributions are explained throughout the paper, which is organized as follows:

- Section 2 briefly describes the approach followed by the other finalist teams.
- Section 3 provides our vision, which guided our development of the avatar system.
- Section 4 summarizes the Semifinals and the Finals testings of ANA Avatar XPRIZE.
- Section 5 concerns our avatar robot, particularly the mechanical and electrical improvements performed based on a previous unit, its expressive face, the vision and sound system, the dexterous hands, the haptic system, and the wireless e-stop.
- Section 6 describes our operator system, particularly our enhanced visual feedback, the operator interface, our strategy to transmit the expressions to the robot, and the haptic feedback given to the operator.
- Section 7 describes the avatar software framework, particularly the robot model, the QP-based control approach, the upper-body retargeting, the balance control during interaction and locomotion, the hierarchical inequality admittance, and how we used the emergency stop signal.
- Section 8 describes the evaluation of our system. First, we explain our outcome at the Semifinals testing, as well as our situation at the Finals. Then, we assess the capabilities of our system through a finals-like course carried out at our laboratory.
- Section 9 describes the lessons we learned from participating in the competition.
- Section 10 concludes our paper and gives our research direction.

2 Related Works

Team NimbRo, the winner of the ANA Avatar XPRIZE competition, built their whole avatar system using only off-the-shelf components. Their avatar features an anthropomorphic bimanual arm configuration with dexterous hands and a 6D movable head mounted on a holonomic base. The head carries a telepresence screen displaying a synthesized image of the operator with facial animation. An operator drives their avatar through an exoskeleton-based operator station that provides force feedback to the wrist and fingers. The arms of the avatar and the exoskeleton-based operator station are both implemented using Franka Emika Panda 7-DoF robotic arms [7, 8, 9].

Team Pollen (the second place) used as an avatar a substantially modified version of Reachy, an open-source platform entirely designed and developed by them, featuring a humanoid upper body mounted on an omnidirectional mobile base. Their operator system consisted of an HMD, hand-held controllers, and a 1-DoF elbow exoskeleton that provided force feedback [6].

Team Northeastern (the third place) also used an avatar with a dual-arm configuration implemented using Franka Emika Panda robotic arms and a three-finger hydrostatic gripper mounted on an omnidirectional base. The operator system featured an exoskeleton composed of two 6-DoF arms and gloves incorporating grippers identical to the ones on the avatar. Then, they adopted bilateral force feedback under varying time delays. Instead of using a VR system, they relied on 2D displays to avoid motion sickness [10, 11].

Team AvaTRINA [12] (the fourth place) used low-cost VR input devices for simplicity and kept the operator interface minimal. Their avatar robot consisted of two 6-DoF arms UR5-e mounted on an omnidirectional base.

Team iBotics [13] (the fifth place) used as avatar the robot EVE, a human-like robot mounted on a segway-like mobile base [6]. Their operation system also featured an exoskeleton as well as haptic gloves.

In general, among the finalists of the ANA Avatar XPRIZE competition, there were only four teams that used biped humanoid robots: Team SNU [14], Avatar-Hubo [15], iCub [16], and us. However, only iCub and our team decided to perform with bipedal locomotion. Team SNU mounted their humanoid robot TOCABI [17] on an omnidirectional base. Avatar-Hubo, which features a hybrid biped-wheeled locomotion system, opted to locomote using wheels and used their bipedal mode only during manipulation. We believe that teleoperating robots is the most practical intervention solution in unstructured or hazardous environments, and contrary to other conventional wheeled robotic platforms, legged robots are better suited when traversing through uneven terrains or climbing stairs [18]. Remarkably, the operational versatility of bipedal humanoid robots makes them suitable for work activities that require a variety of complex mobility and manipulation skills [19].

3 Our vision

The trend in the future of information technology is in extending our social presence into the digital world (e.g., the Metaverse). It is an immersive Internet where one can communicate with virtually anyone, anywhere, anytime, through text, audiovisual, and even haptic presence. It is not only a societal space but also an economic one. This concept is not new; it is also possible that the Metaverse is a digital representation of actual environments (e.g., an entire city) updated according to real changes monitored from that environment. Such a scheme would allow the user to switch virtual actions into actual ones employing a physical presence avatar in the real world. An early version of this vision was proposed in [20], where actions in a virtual environment were seamlessly made in the real world using a robot or several robots at once [21]. Our vision of real telepresence is to carry this presence to the physical world through anthropomorphic robotic avatars. One should be able to interact with remote environments not only with high-fidelity presence but also to embody the avatar [22]. We are also currently working on a setup where two persons are remotely interacting and *embracing* the presence of each other, both in the figurative and the literal sense.

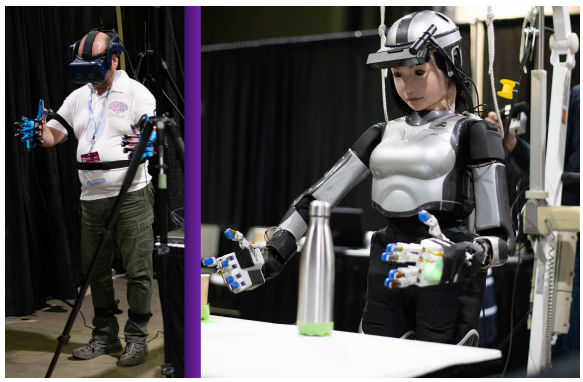

Fig. 1 HRP-4CR synchronized with the operator.

We are using an adult-sized humanoid avatar: HRP-4CR, the only one with a close-to-human look that can realize facial expressions, manipulate objects, and walk[6]. We want to demonstrate that our avatar is (i) easily controllable, (ii) rich in terms of sensory feedback, and (iii) a suitable solution in many application fields. See Fig. 1. With the recent booming of humanoid robots to serve the purpose of various applications, our technology can also be used for skill transfer through teleoperation.

With that idea in mind, we saw in ANA Avatar XPRIZE an excellent opportunity to create and integrate the technology that meets our objective of advancing fundamental knowledge and innovation. It was indeed very challenging to hold the competition by not sacrificing the bipedal challenge nor the anthropomorphic shape of the robot, including its proportions, despite the competition clearly becoming less favorable to bipedal locomotion. Every mechanical improvement had to fit within very narrow spaces, and every sensor had to be compatible with the human sensory system. Therefore, we could not consider solutions like using a robotic arm to mount the head of the robot as team NimbRo did [8].

[6]Sophia, the avatar of team Aham [23], is another impressive humanoid avatar with a close-to-human look and excellent skills for manipulation, but it still relies on wheels to locomote.

Moreover, one's remote presence needs to be fully acknowledged by the interacting people, which is why the humanoid shape is essential. There are currently challenging technicalities to overcome to reach this goal. High-fidelity replication of oneself proved possible, e.g., with the seminal work of Prof. Ichiguro [24]. Clearly, we are still far from reaching the sophistication of a self-virtual avatar generated by computer graphics rendering and animation techniques, considering that the latter has achieved unprecedented advancements. Replicating a similar-looking humanoid (android) with graceful motions is not yet possible but not out of reach. However, the face and shape of the physical avatar can replicate only one person's avatar at a time. For applications where the physical avatar evolves strictly under the physics of the physical world (e.g., no teletransportation or omni-locations presence), this is feasible. If aimed as a beaming device, i.e., a physical avatar that can take the appearance of anyone, this is very challenging. This obstacle is likely why many teams opted for a simple flat screen to render the video of any remote person's avatar. Our vision is that, if needed, the actual skin covering the head of our avatar will, in the future, be replaced by flexible (i.e., bendable and stretchable) display technology, keeping the articulation as an extension of existing rigid forms serving as displays (some are flat or oval, yet rigid). We believe that with state-of-the-art rendering techniques, one can perform a very realistic avatar face/head display. Yet the problem of anthropomorphic matching is open: changing online the size of people (e.g., from a child to an adult of different body shapes and heights) is not possible without adding more actuators and extremely complex mechatronics (that does not exist yet).

The logo of the team was also a requirement of the XPRIZE organizers. We have been questioned on the meaning of our logo. In Roman mythology, JANUS represents the transition from the past to the future, like the one we see today with new technologies. It also represents bridges and connections, like the ones we are building between humans and robots: two entities in the same *body* [22]. This idea is symbolized in our logo, representing a human brain interconnected with a "machine brain" as if they were the same entity. See Fig. 2.

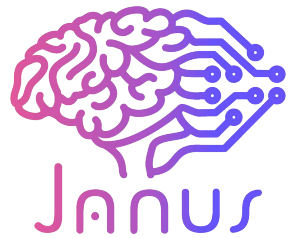

Fig. 2 Logo of JANUS: a human brain interconnected with a "machine brain" as if they were the same entity.

4 The challenges of ANA Avatar XPRIZE

One of the claimed objectives of ANA Avatar XPRIZE was to advance the state of the art in avatar systems. To get that done, teams were required to build robust, intuitive, and immersive avatar systems that could be operated by briefly trained non-expert operators (the judges) at each stage of the competition, namely the Semifinals and the Finals testings [8]. The judges could be trained only for a short time (about 1 hour) [8].

During the Semifinals testing[7], the avatar (operated by a judge) and a recipient (also a judge) were expected to have a conversation while understanding each other's intentions and, in this way, test the connectivity. Also, throughout three scenarios, the avatar was expected to manipulate the objects shown in Fig. 3: a flower vase of 1.3 kg,

[7] https://www.xprize.org/prizes/avatar/articles/on-the-ground-at-the-ana-avatar-xprize-semifinals

Fig. 3 The objects to be manipulated at the ANA Avatar XPRIZE in the Semifinals.

Fig. 4 The course (shown below) and tasks (depicted above) at the ANA Avatar XPRIZE in the Finals: (a) mission communication, (b) switch, (c) door crossing, (d) long-distance navigation, (e) canister plug-in, (f) narrow-space navigation, (g) sliding door removal with drill, and (h) rock selection (behind the sliding door).

a glass of wine, a mug, and puzzle pieces placed over a desk. Scenario #1 was a collaborative puzzle activity, requiring the avatar to manipulate the puzzle pieces to insert them into its place. Scenario #2 simulated a final stage of a business deal, requiring the avatar to make a toast using either the glass of wine or the mug. Scenario #3 simulated a visit to a distant museum of antiquities, for which the avatar was requested to take the artifact (the flower vase), explore its texture, and describe it, along with its weight. These scenarios required bulk and precision grasping, as well as the capability to get some haptic feedback. The expected locomotion was simple (just moving 1.2 m away from the desk).

For the Finals testing[8], the focus was not only the connectivity but also the ability of the operator to explore a remote location and transfer his or her skills to the avatar wherever any specific know-how is needed. To test these capabilities, XPRIZE designed a course (supposedly on a "remote planet") that required mobility over 30 m (see Fig. 4), as well as dexterous manipulation of objects and tools. The considered tasks were: (a) to walk to a mission commander to give a report and receive instructions, (b) to activate a switch[9], (c) to cross through a wide door, (d) to move for 20 m without obstacles, (e) to identify the heaviest canister (about 1 kg) among several and introduce it into the corresponding slot, (f) to move around obstacles with the narrowest space being 1.4 m, (g) to operate a drill to remove a hexagonal screw that opens a sliding door giving access to a collection of rocks, and (h) to identify the roughest rock among several behind a curtain by using only haptic feedback. This course had to be completed in 25 min or less. Performing these tasks required an untethered avatar with the ability to navigate, perform precise and bulk grasping, and utilize an advanced haptic system.

A detailed description of the competition stages, particularly of these two events (Semifinals and Finals testings) can be found at [25].

[8] https://spectrum.ieee.org/xprize-robot-avatar

[9] The safety switch chosen by XPRIZE has a spring that requires a force of about 5 kg·f to be applied (measured by us); however, XPRIZE removed that spring to ease the task.

5 Avatar Robot

As mentioned in Sec. 3, we are using as an avatar the humanoid robot HRP-4CR (see Fig. 7). This robot is 1.635 m in height when the legs are fully extended and weighs 49.7 kg. It has 42 available DoF (see Fig. 5): two legs of 6 DoF each, two arms of 7 DoF each, two hands of 1 DoF each, a waist with 3 DoF, a head with a neck of 3 DoF, and a face of 8 DoF. However, we are only using 38 DoF because only 4 DoF of the face can be used. See details in Sec. 5.3.

5.1 Mechanical and Electrical Improvements

HRP-4CR (Fig. 7) is an enhanced version of the cybernetic humanoid robot HRP-4C [26] (Fig. 6), or Miim, which was initially released in 2009 and was designed to resemble the appearance of an average Japanese female. This robot was over ten years old, so some of its components were difficult to purchase, modern operating systems no longer supported the drivers, and its wiring had become unreliable.

HRP-4C was originally developed for entertainment (e.g., for expressing emotions, dancing, and human-like walking) [27]. Therefore, the arms and hands were not designed to perform manipulation nor to bear the loads required by ANA Avatar XPRIZE: the upper-body joints lacked rated torque, each arm had 6 DoF, and there were no F/T sensors at the wrists. Also, the hands were 'decorative' and not capable of any grasping. This issue is addressed in Sec. 5.4.

Therefore, we decided to enhance the robot by (a) changing the low-level field-bus technology (CAN) into EtherCAT, (b) improving the cooling system, (c) adding a 7-th DoF to each arm, (d) increasing the strength of almost all the joints of the chest and arms, (e) implementing F/T sensors at the wrists, (f) adopting a new battery box system, and (g) implementing a proper power management system; all of these while keeping the physical appearance of the robot as close as possible to the original one. Doing this was very challenging due to the confined and narrow space that was available, resulting in a light robot[10].

The field-bus technology was changed from CAN (max. 1 Mbps) to EtherCAT (max. 100 Mbps), and the latter was implemented as a network divided into six lines connecting the EtherCAT devices in a daisy chain. These devices are two types of motor drivers (Elmo: G-TWI R50/100 EE and Technosoft: iPOS2401 MX-CAT) driving all the motors of the robot except for the hands, the IMU (Epson: M-G370) mounted on the waist, and F/T sensors on each ankle (ATI: Mini58) and on each wrist (ATI: Mini45).

[10] For comparison, other teams struggled to keep their robots below the maximum limit (160 kg) established by ANA Avatar XPRIZE [11].

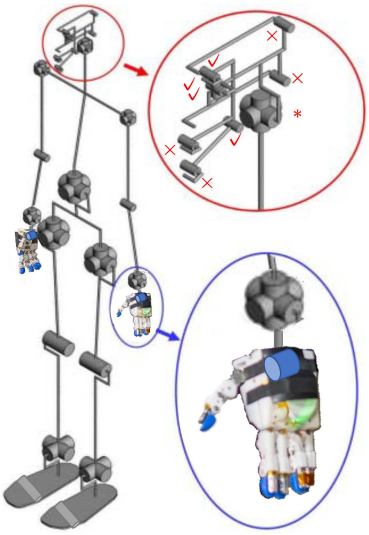

Fig. 5 DoFs of HRP-4CR. Only 4 of 8 DoF of the face are used (✓).

Fig. 6 HRP-4C

Fig. 7 HRP-4CR

By adopting these distributed motor drivers, the amount of heat they generate was higher than those based on CAN. As a result, the cooling system had to be improved, and this was done by using more powerful DC fans, carefully planning the airflow path within the links, and using air holes on the cloth covers of the robot (over which there are plastic covers, with a gap between them).

The dexterous workspace of the arms was improved by adding a 7-th. DoF to each arm implemented as a Wrist Pitch joint. The strength of the arms and chest joints was improved by increasing the rated torque of several joints: Chest Pitch/Roll, Shoulder Pitch/Roll/Yaw, and Wrist Yaw/Roll. This was done by changing the gear ratio (the pulleys), changing the motor, or using both strategies. By doing that, the robot could manipulate objects and operate tools up to 3 kg.

For the Semifinals, the robot was still externally powered. However, for the Finals challenge, the robot had to be completely untethered. This required us to design a new battery box for the robot, which we did by considering the energy required for the course. We took into consideration the measured power under some conditions: just standing (302 W), standing while performing manipulation (426 W), and walking (821 W), as well as an estimated time of each condition during the competition. Based on the analysis, we developed a battery system with a minimum of 180 Wh. We implemented it using LiFE (Lithium FErrite) battery cells arranged in two boxes placed at the hips of the robot.

5.2 Vision, Sound, and WiFi

To allow for remote viewing, we installed a stereo camera (Stereolabs: ZED 2) on the head of the robot. Concretely, it was mounted on a helmet designed for the robot when it was used for the opening speech at an event back in 2009 [27]. The streaming of the stereo camera is sent to the operator by using the ZED proprietary SDK.

For sound capture, the robot is equipped with two types of microphones. The first one is a stereo microphone system (System In Frontier Inc.: RASP) located at the level of the robot's ears, allowing for sound sources' location. The second one is a supercardioid (directional) shotgun microphone (RØde: VideoMic GO II) mounted on the helmet, allowing it to receive high-quality sound from the front. When it was announced that sound localization was not a required task in the competition, we chose to use only the directional microphone to optimize the quality of the interaction. Audio communication is achieved using an open source VoIP software[11].

Echo-cancellation was an essential consideration while developing the 2-way audio communication. Without active echo-cancellation, the audio output from the speakers installed within the robot's chest is sufficiently loud to be picked up by the microphones on the helmet and consequently fed to the operator's headset. Hence, the operator can hear an *echo* of their voice, which is undesirable. Thankfully, the VoIP software we use can also be configured to perform echo cancellation.

These perception devices are managed by a small PC with GPU (NVIDIA: Jetson Nano), the vision computer, mounted inside of the head of the robot. The wireless LAN of this vision computer is realized through a WiFi card with Intel Wireless-AC 8265 (IEEE802.11 ac/n/a/g/b) that is installed on the Jetson Nano. The WiFi antennas are also installed inside of the head.

5.3 Expressive Face

The head of HRP-4CR is mounted on a 3 DoF neck, allowing a more natural head motion. Emotions in our avatar system are expressed through the DoF on the face of the robot.

The facial expression was originally driven in HRP-4C by using 8 DoF inside of the head: EYEBROW_Pitch, EYELID_Pitch, EYE_Pitch, EYE_Yaw, MOUTH_Pitch, LOWERLIP_Pitch, UPPERLIP_Pitch and CHEEK_Pitch, which in the past allowed the robot to imitate facial expressions of a person singing [28]; however, as we mounted the vision PC mentioned in Sec. 5.2, it took the space of 4 motor drivers. Thus, we had to select 4 DoF to control and sacrifice the others.

We based our selection of DoFs on a compromise between simplicity for getting the corresponding facial characteristics from the operator (e.g., the eye motion from eye tracking) and the usefulness of the corresponding DoF in contributing to non-verbal communication (e.g., the motion of the mouth while talking). In that sense, both the DoF of the eyes and the aperture of the

[11] https://www.mumble.info/

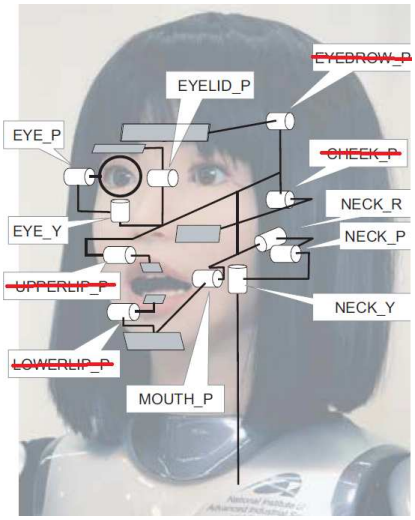

Fig. 8 Face and neck joints of HRP-4C [32]. Disabled dof are canceled with a red line.

mouth were obvious candidates. As for the last DoF to be controlled, we selected the eyelid(s). This decision is grounded on research that suggests that the blinks of a listener are perceived as communicative signals, directly influencing the speaker's communicative behavior in face-to-face communication [29]. Consequently, we decided to control: EYELID_Pitch, EYE_Pitch, EYE_Yaw, and MOUTH_Pitch; see Fig. 8.

Among all the finalist teams, our team was the only one relying on a fully articulated face to show expressions on the robot. Team iCub articulated only the eyelids, leaving the eyes steady and using LED-drawn and animated eyebrows and mouth [16]. Other solutions were to use facial animation displayed on monitors or tablets mounted on the robot and based on pictorial elements (team SNU [30], team iBotics [13]) or actual video feed of the operator's face. The latter was a straightforward solution for team Northeastern, which decided not to use a VR headset that would partially occlude the operator's face [11]. On the other hand, team NimbRo solved the partial occlusion problem by using facial animation of the operator's photo displayed on the tablet [31].

5.4 Dexterous Hands

Hands became one of the critical components for the requested manipulation tasks. The human hand involves many degrees of freedom, almost comparable to a whole-body humanoid robot.

Anatomical finger joints are defined as:
(a.i) Metacarpophalangeal (MCP) joint,
(a.ii) Proximal Interphalangeal (PIP) joint,
(a.iii) Distal Interphalangeal (DIP) joint,
and thumb joints are:
(b.i) Trapeziometacapal (TM) joint,
(b.ii) Metacarpophalangeal (MP) joint,
(b.iii) Interphalangeal (IP) joint.

Here, MCP and TM joints are multi-DoF joints. If these DoFs are counted as 2, then the skeletal DoF of the fingers in total becomes 20, which does not count metacarpal movements.

When the robot is being teleoperated, it is desirable to realize a suitable embodiment [22]. Having anthropomorphic hands is one of the strategies to ease the projection of the body-image of the operator to the robot [33].

The ultimate embodiment should provide identical functionality of the human hand to the avatar robot's hand [34], which would require control of very dextrous hands. The embodiment is a level of body-image projection, and the dexterity is a level of achievable task variation and precision, hence two different concepts. However, especially for hands, since human hands are very dextrous compared to robot hands, the level of the required dexterity for achieving suitable embodiment is already a challenge. Even if the same number of DoFs were realized, actively controlling all DoFs is a very challenging task, e.g. many motors increase weight and mechanical complexity, computational load, and communication load. Hence, realizing sufficient dexterity that provides suitable embodiment while simplifying the mechanism and control is the objective of hand development in this work.

Complex mechanisms are fragile and do not suit the need for load-bearing tasks. For example, in the ANA Avatar XRPIZE finals competition, NimbRo used a right 20-DoF Schunk SVH Hand for precision tasks and a left 5-DoF Schunk SIH Hand for force-requiring tasks [7] to realize dexterity and mechanical robustness.

From the control point of view, having many DoFs does not always enhance embodiment when kinematic differences between robot and human hands cannot be efficiently compensated during dexterous tasks under visual feedback [34]. Of course, having active DoFs allows performing more sophisticated control. It must be thought of as a trade-off of the drawbacks mentioned above.

One of the attempts to realize dexterous movements with less DoF is to use synergy. Brown and Asada [35] used Principal Component Analysis (PCA) to analyze hand synergy and concluded that 5 top principal components can represent 90% of hand movements. Catalano *et al.* [36] used synergy with Series Elastic Actuators to enhance both the dexterity and the stability of the grasps.

One extreme of this idea is to use only one actuator and have underactuated joints so the hand will adapt its shape to the grasping object. This method is used in electromyography (EMG) controlled prosthetic devices for two reasons: the importance of light weight and ease of control with EMG. Fukaya *et al.* [37] developed an underactuated hand with one motor, later constructed a humanoid hand, D-Hand, with five fingers driven by one motor [38]. Underactuated DoFs allow the hand to adapt to various kinds of objects without explicit control of the joints.

One of the objectives of this work is to realize dexterous manipulation using low active DoFs and underactuation with synergies, which can be termed as *mechanically embedded intelligence*.

HRP-4C originally had human-looking hands [27]. The hand had five fingers and 13 joints. Finger joints (DIP, PIP, and MCP joints of index, middle, annular, and little fingers) were driven with one motor, and the Thumb TM joint was driven with another motor. Since all the joints connected to a motor are rigidly coupled, the hand had only 2 DoF. In order to perform dexterous manipulation, more DoFs are necessary.

In this project, we developed a hand based on a concept of the D-Hand [38] (developed by our partner, Double R&D), an underactuated humanoid hand with 16 joints and 13 DoFs. The kinematics of the hand and the picture of D-Hand V3.1 are shown in Figs. 9 and 10, respectively. The TM2 joint of the thumb and other joints move in a sequential manner such that when the fingers are fully extended, as in Fig. 10, the TM2 joint will close first, then all the other joints close in an underactuated manner, which allows the hand to perform non-prehensile manipulations.

The structure of the hands was optimized for power grasps of various objects and non-prehensile grasps that were expected to become necessary in the ANA Avatar XPRIZE Finals, while maintaining some precision grasping capability, namely tripod grasping objects such as puzzle pieces that

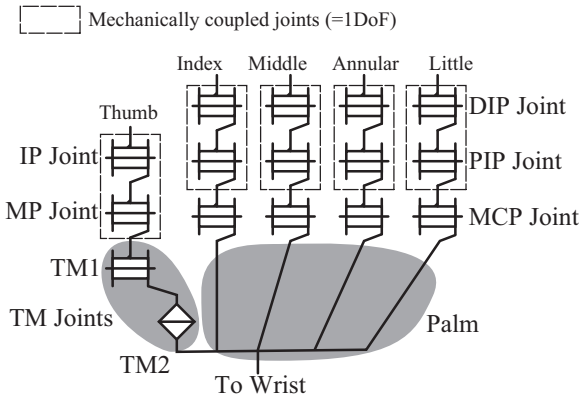

Fig. 9 Kinematics of D-Hand V3.1 (posterior view)

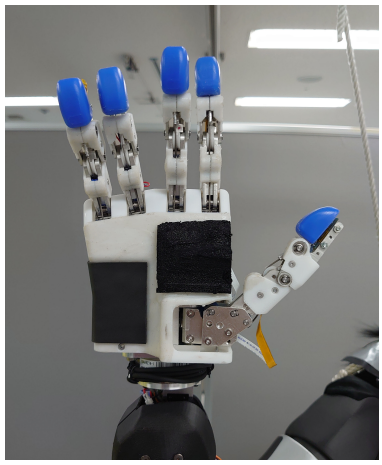

Fig. 10 D-Hand V3.1 (anterior view)

were necessary for the Semifinals. The design trade-offs were made between practical issues such as robustness and weight and a high level of dexterity in general cases.

The joints are driven with wire tendons, connected through a differential mechanism to realize underactuated behavior. The MCP and PIP joints are loaded with springs, which determine the closing speed and timing of the joints. With tuning, D-Hand V3.1 can perform precision grasping, such as pinching or tripod grasping (see Figs. 11, 12). D-Hand V3.1 can also perform a power grasping task like holding a power drill and then triggering it (see Fig. 13). Each finger can produce approximately 7 N in a fully stretched configuration.

Fingertips are equipped with force sensors to provide touch feedback, as described in Sec. 5.5.

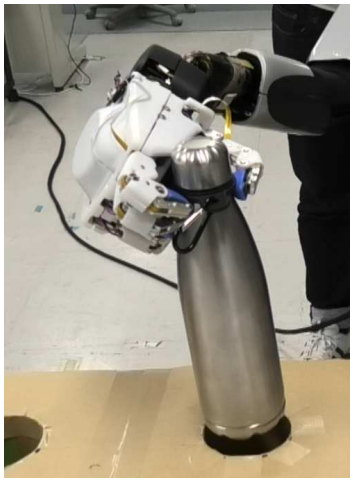

Fig. 11 Precision grasping (thumb and two fingers grasping a heavy canister)

Fig. 12 Precision grasping (tripod grasping of a puzzle piece by its knob)

Fig. 13 Power grasping (heavy wrap while operating drill)

5.5 Haptic Sensing System

On the avatar side, we developed a haptic sensing system to give tactile feedback to the user. We attached miniature 6-Axis Force/Torque (F/T) sensors (Touchence: P18) to three of the fingertips (thumb, index, middle) and Single-Axis Force Sensitive Resistors (FSRs) (FlexiForce: A101) to the remaining fingertips (ring, little) on each hand.

Fig. 14 Mounting of sensors on the fingertip.

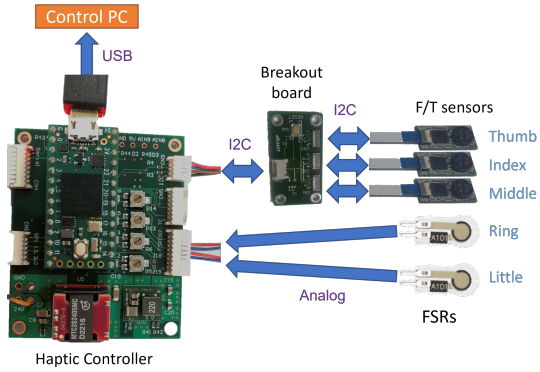

Fig. 15 Haptic system of each hand.

These miniature sensors are mounted under each fingertip to fix their position and protect them from impact (see Fig. 14).

We also designed and developed a Haptic Controller: a small Printed Circuit Board (PCB) to connect these sensors to a development board (Teensy 4.0) featuring an ARM Cortex-M7 microcontroller running at 600 MHz. The miniature F/T sensors can provide a digital output transmitted by I^2C. So, we had to design a breakout board to connect those to an SMBus/I^2C accelerator (LTC1694) mounted on the Haptic Controller. The FSRs provide an analog output that is connected to trans-impedance amplifiers (and inverters), mounted on the board, and digitalized.

The Haptic Controller was designed to reduce sensor noise and prevent interference between multiple sensors. It was also prepared to receive and process the output signal coming from thermocouples. This is because, initially, ANA Avatar XPRIZE had announced that the haptic system would also need to transmit thermal sensation, but this requirement was later removed. Thermal sensation is a desirable feature, so we plan to include it in the future.

The developed system makes it possible to send the sensor's outputs through USB-serial communication to the control PC inside of the robot (see Sec. 6.4) and publish them through ROS.

5.6 Wireless E-Stop

We built a wireless emergency stop button (E-stop) consisting of a Zigbee 3.0 module, a battery charger, and a Li-ion battery. The emergency signal is transmitted between the E-stop and another Zigbee 3.0 module connected to the control PC of the robot through USB-serial communication. LEDs on the E-stop button indicate the emergency status in the real-time controller, communication status (initializing, activated, or lost communication), and battery status (full or charging).

6 Operator System

Our operator system is shown in Fig. 16. It consists of a Head Mounted Display (HMD), a motion tracking system, and haptic gloves. A PC running Unity manages these devices.

The HMD (VIVE Pro Eye[12]) offers a resolution of 1440 × 1600 pixels per eye with 110° diagonal Field of View (FoV) and a 90 Hz refresh rate, as well as headphones for audio communication. It is used to provide the operator with the image captured by the robot through the stereo camera (Section 5.2), enhanced as explained in Section 6.1. It also offers eye-tracking functionality, which we actively use to interact with the Operator Interface (Section 6.2) and to transmit expressions to the robot (Section 6.3). Additionally, we installed a lips-tracking device (VIVE Facial Tracker[13]) on the HMD and used it to complement the information needed to transmit expressions to the robot (Section 6.3).

To track the operator's limbs, we used a motion tracking system consisting of 7 individual trackers (VIVE Tracker[14]), each providing a 6D pose estimation. They have been installed on the operator's: (a) lower back (to provide a reference frame with respect to which the motion of other limbs are defined), (b) hands (to track their motion on the robot's limbs), (c) elbows (to track as best as possible the motion of the arms), and (d) ankles (to command the walking motion, see Section 6.2). This technology requires the installation of base stations around the operator.

[12] https://vive.com/us/product/vive-pro-eye/overview/
[13] https://www.vive.com/us/accessory/facial-tracker/
[14] https://vive.com/us/accessory/tracker3/

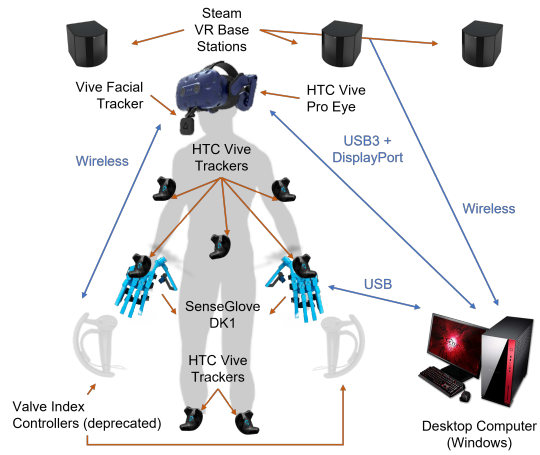

Fig. 16 Overview of the full operator system.

We also decided to use haptic gloves implemented as an exoskeleton for the hands (SenseGlove DK1). These gloves provide the state of the fingers' joints of the operator, which can be used to control the robot's grasping. As the hands of the robot have only one degree of freedom (Section 5.4), we set their closing value to follow the operator's most flexed finger on each hand. They can also provide haptic feedback, for which more details are given in Section 6.4.

Our framework also supports using hand-held controllers (Steam: Valve Index), which we used actively for the Semifinals. With these, the grasping is triggered by buttons (open or close). They also allow commanding the walking motion and activating the hands' control. However, since we needed better grasping control and haptic feedback, they were deprecated. This decision leads also to the design of a button-less operator interface (Section 6.2).

6.1 Enhanced Visual Feedback

Three shortcomings may alter the transparency of the visual feedback, causing discomfort for the operator or jeopardizing the embodiment [33]: (a) the lag between the motion of the operator's and robot's head due to network communication delays or discrepancies in joint velocities, causing cybersickness [39], (b) a mismatch between the robot camera's and the HMD's FoV (the former usually smaller), and (c) a mismatch between human's and robot's range of motion of the neck due to joint limits on the latter.

To cope with these shortcomings, we use one of the ideas of our previous work [40], consisting of decoupling the visual feedback between the operator's and the robot's head movement in a virtual or augmented environment. Therefore, as shown in Fig. 17, the environment rendered in the HMD follows the movement of the HMD at the operator's side. In contrast, a screen rendered with RGB video data from the robot's point of view follows the movement of the camera mounted on the robot's head. As a result, the image is spatially consistent with the robot's motion, improving the understanding of the environment. The inconvenience is that the area outside the camera's FoV is empty (black) on the HMD.

Furthermore, to improve the perception of space and environment, we integrate the robot model into Unity so that the user can visually perceive the robot's body in the virtual environment created in Unity and projected into the VR headset. As soon as the orientation of the operator's head exceeds the limits of the robot's head, the user perceives the robot's body (its shoulders, the position of the arm in space, etc.) in the displayed environment and only over the empty (black) area. See also Fig. 17-(C).

Two other teams also used decoupled visual feedback for ANA Avatar XPRIZE but in a different way. Team NimbRo deployed a spherical rendering with two 180° stereo cameras mounted on a 6-DOF robotic arm [41]. They mention that if the distance between the object and the camera is shorter than the radius of the sphere into which the image is projected, a significant distortion appears. Team SNU presented a method using two RGB cameras to render a stereoscopic view on a curved plane to reduce distortion [42]. In their case, the latency is mainly due to the network connection and the delays in rendering the scene, which can cause discomfort.

For future improvement, we are considering integrating SLAM as suggested in [40] to fill in the missing FoV instead of having a black space.

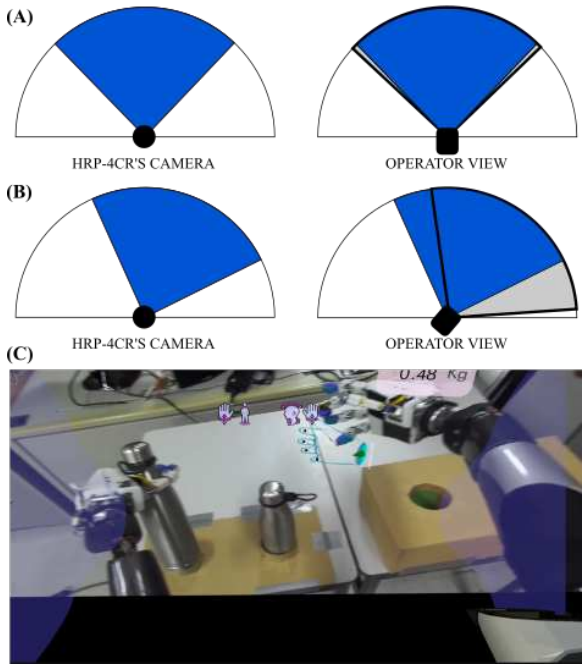

Fig. 17 Decoupled visual feedback and integration of the robot model. The blue area represents the camera's FoV mounted on the robot, whose orientation varies according to the motion of its head. The area surrounded by a bold line is associated with the orientation of the operator's head. The gray area represents an empty environment, rendered black on the operator's head. We can see two different situations in (A) and (B): In (A), HRP4-CR and the operator have the same head orientation, and then the video data is seamlessly displayed on the HMD. In (B), the head orientation of HRP-4CR and the operator is not the same, so the operator sees only part of the image captured by the robot, and the remaining part is empty (black). In (C), we can see the 3D model of the robot rendered only over the empty (black) area outside the FoV. In this case, it is part of the robot's arm (seen in the lower right corner).

6.2 Operator Interface

Beyond the competition, our main target is to allow for the most profound possible sense of embodiment [22]. Since humans do not have buttons and use directly their hands for interacting with the environment, we chose a button-less graphical user interface (GUI). However, it was unfortunately not possible yet to perform a fully embodied interaction, mainly because the operator could not freely walk in the operating room, and we were still lacking perfect force feedback. Therefore, we needed to implement new interfaces to activate the control of the arms and steer the robot. These interfaces are of three kinds: eye tracking, head orientation, and voice interface.

Vocal commands allow the operator to switch between hands control mode and locomotion (walking) mode (see Fig. 18). These modes are separated in the GUI mainly to prevent accidental activation of locomotion during manipulation. Each of these interfaces is described next.

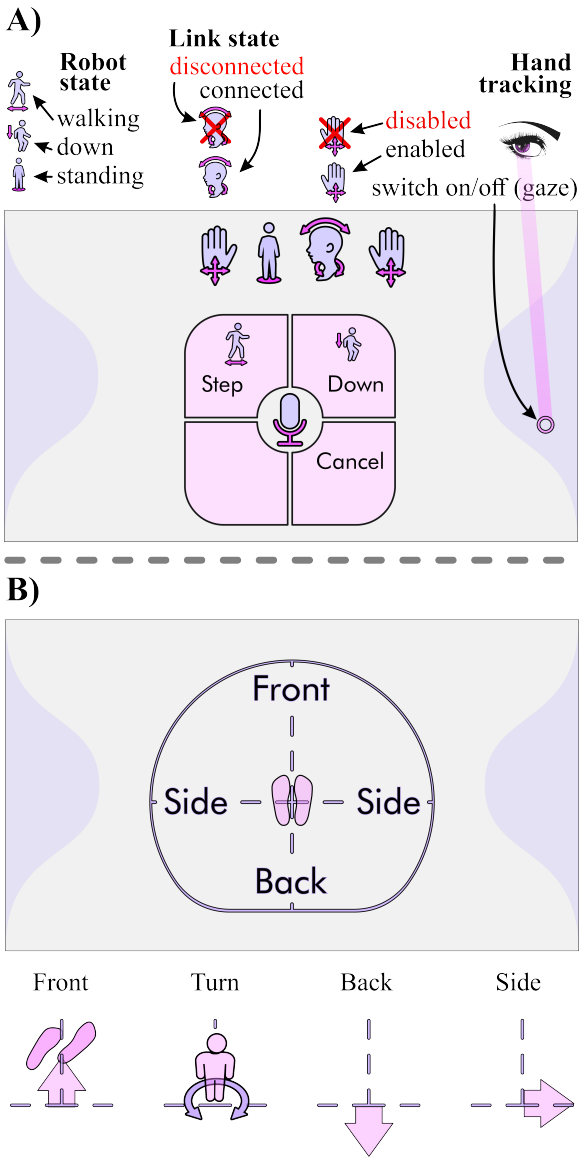

Fig. 18 Operator's GUI representation. The light gray areas represent the FoV of the operator. Panel **A)** shows the voice commands and the state interface. The vocal commands are shown in the middle of the FoV around the microphone symbol. The state of the robot is displayed using icons. The hand retargeting is triggered when the gaze is detected for several seconds over the violet areas on the sides. While the gaze is detected over those areas, its location is shown with a small circle, and the violet area becomes gradually opaque. Each side triggers the activation of the corresponding arm. Panel **B)** shows the walking interface, which changes according to the different walking modalities (shown below): walking forward, turning in place, backward, and sideways. These modalities can be triggered through the voice interface.

6.2.1 Hands Control Interface

Since we lack force feedback on the robot arms, it can be dangerous to keep the control of the operator's arms constantly active. Deactivating them also allows the operator to rest or perform motions that are not intended to be replicated remotely.

This choice in the control design can limit the possibility of increasing the embodiment as we allow the operator's hands to move without seeing any motion on the robot side. However, we can improve the ease of use and the reliability of the whole system for the operator.

For activating/deactivating the control of the arms, we use the gaze tracking feature. If the operator stares for more than three seconds over one of the lightly shaded areas on the sides of the FoV, the hand control of the corresponding side (left or right) activates or deactivates (see Fig. 18-A)). A visual cue shows the center of the gaze, and the area opacity acknowledges the activation/deactivation. Two pictograms at the top of the screen with the shape of a hand (left and right) inform the operator if the corresponding hand control is activated or deactivated (shown crossed).

Finally, a vocal command makes the robot decrease the waist height to reach lower targets.

6.2.2 Walking Interface

To promote embodiment, we aimed to achieve a high level of telepresence in our humanoid avatar [33] thanks to a walking interface triggered by stepping. However, since the operators' space is limited, they need to step in place to trigger walking and use voice and head orientation interfaces to steer the locomotion. A voice command triggers the display of a menu showing the stepping modalities. The operator can choose a modality with a second voice command (forward, back, or side) and start stepping in place. See Fig. 18-B).

For turning, the operator turns the head in the desired direction, and the robot turns in the corresponding direction. Similarly, while the sidewalk mode is enabled, the head direction of the operator (left or right) will determine if the robot sidewalk is toward the left or the right. The visual feedback is then augmented with arrows and footstep cues showing the steering direction. Finally, the amplitude of the step is related to the inclination of the operator's head. Looking straight gives maximum amplitude, and looking down decreases it.

6.3 Transmission of Expressions

Besides the avatar's ability to transmit skills, another essential quality considered for the competition (during the Semifinals and the Finals) was the ability to convey verbal and non-verbal communication. The latter delivers emotional information and is particularly important to enhance the human-robot interaction. Notably, in human-human communication, the face plays a vital role. We can figure out most of the non-verbal cues and emotions through facial expressions [43], and we can take advantage of our hardware for such a purpose.

As mentioned in Section 5.3, we can control the yaw and pitch angles of the eyes, as well as the opening angle of the eyelids and mouth of our robot. However, given that one joint drives the yaw angle of both eyes, we cannot control the eye motion for each eye independently. Therefore, we cannot modify their vergence (mechanically set as 0 deg by default). In the same way, we cannot control the pitch angle of each eye independently, nor the opening angle of each eyelid.

The HMD that we are using is capable of obtaining gaze and eye-state information [44], including gaze point, gaze direction, pupil position, pupil size, and eye openness. To overcome the fixed vergence limitation, we define a virtual eye in the middle of both eyes and obtain the gaze angles for it. Based on the position of gaze origin point (x, y, z) in Fig. 19, we can calculate the yaw angle α and the pitch angle β of the virtual eye as:

$$\alpha = \arctan\left(z/x\right), \tag{1}$$

$$\beta = \arctan\left(y/\sqrt{x^2 + z^2}\right). \tag{2}$$

To obtain the opening angle of the mouth, we use the lip tracking device. A total of 26 lip *blend shapes* have been predefined within the VIVE Eye and Facial Tracking SDK [45], but we only focus on one of them: *Jaw_Open*[15].

In Unity, we can retrieve the *weighting* (the percentage of resemblance) of the detected lip shape of the operator to the blend shape. We directly associate this resemblance with the openness of the lip (a value from 0 to 1).

[15]See the blend shape 03.JAW_OPEN at https://hub.vive.com/storage/docs/en-us/UnityXR/UnityXRLipExpression.html

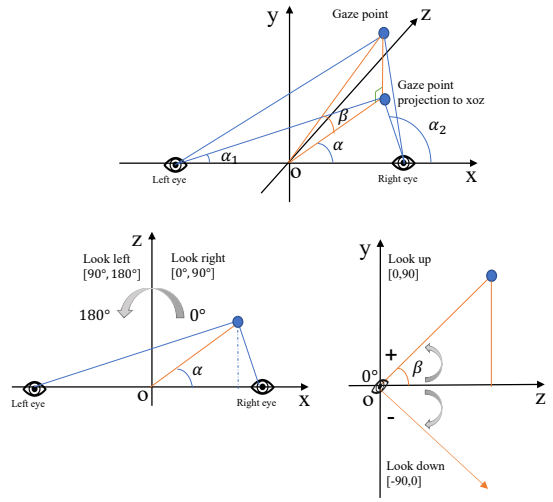

Fig. 19 Diagram used to calculate the gaze angles for the virtual eye from the gaze information.

The extracted facial information is collected as a 4D vector: [α, β, *eye openness*, *lip openness*], and on the robot side, it is mapped onto the four face actuators by setting appropriate targets. For controlling the gaze direction in the horizontal and vertical planes, α and β are scaled and translated to match the range of EYE_Yaw and EYE_Pitch joints, respectively. Similarly, *lip openness* is mapped to the MOUTH_Pitch joint. Since the physical velocity limit of the EYELID_Pitch is relatively lower than the average human blink speed, the *eye openness* value was ultimately ignored, and the *eyelid* joint was manually programmed to mimic blinking behavior at periodic intervals (in a similar way as in [32]). Nevertheless, we plan to use *eye openness* value to allow transmission of eye shutting and opening behavior in the future. Some screenshots of expressions being transmitted to the robot are shown in Fig. 20 and Fig. 21.

Since the motion of *eyelid* joint does not follow the operator's, only the three remaining facial joints (EYE_Yaw, EYE_Pitch, MOUTH_Pitch) are being used to transmit the operator's facial expression. Furthermore, there is another limitation that comes from the original range of motion of the facial joints (from HRP-4C). While EYE_Yaw has enough range of motion, EYE_Pitch, EYELID_Pitch, and MOUTH_Pitch do not. Consequently, their motion is subtle, the eyelids cannot cover the eyes (only move above

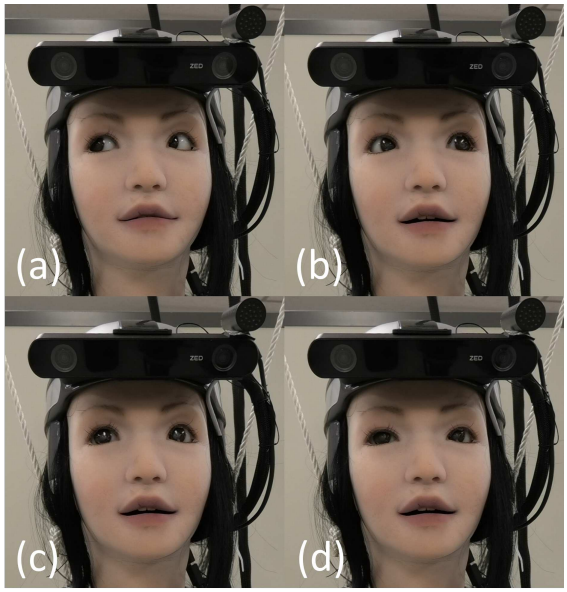

Fig. 20 Movement of the face joints: (a) and (b) Eye Yaw, (c) Eyelid up, (d) Eyelid down, (a) Mouth closed, (b)(c)(d) Mouth open. Notice the subtle motion of the eyelids (through the glow in the eyes) and the mouth (through the emergence of the teeth).

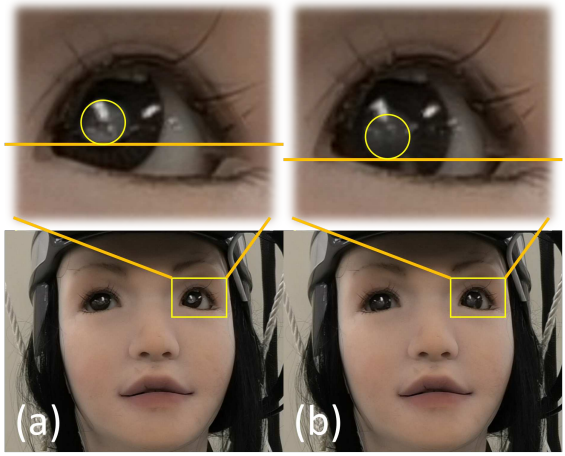

Fig. 21 Eyes are looking (a) up and (b) down using the EYE_Pitch joint. Close-ups are included for each image, where the pupil is annotated to emphasize the subtle motion of the eyes.

them), and the mouth opening is small. While these limitations restrict the spectrum of expressions that can be transmitted faithfully, we believe the face joints still contribute to anthropomorphizing the robot. Emotions such as being "unamused" can be expressed with a combination of

Fig. 22 Neck motion: looking left-right using the yaw joint in (a) and (a), and head tilting using the roll joint in (c) and (d).

looking to the left or right (using the EYE_Yaw joint) and closing the mouth (MOUTH_Pitch joint), visible in Fig. 20 (a). "Thoughtfulness" can be demonstrated by looking up and opening the mouth (EYE_Pitch and MOUTH_Pitch joints, respectively), as seen in Fig. 21 (a). The emotional information can be enhanced by combining the motion of the face joints with the 3-DoF motion of the neck (as described in Section 5.3), which tracks the motion of the operator's head. As shown in Fig. 22 (c) and (d), the robot visibly appears to be in a state of "thoughtfulness" due to the combined effect of the neck and face joints.

6.4 Haptic Feedback

A study in [46] reported that users can experience haptic sensations with visual feedback alone even from a non-anthropomorphic embodied limb/agent. The haptic sensation was reported even though the setup did not include any haptic or pseudo-haptic feedback device of any form. However, even though these results highlight that vision could suffice to render a subset of touch information, more is needed to achieve immersive telepresence.

The human sensory system is extremely complex and multidimensional. To provide appropriate haptic feedback to the operator for achieving efficient manipulation through teleoperation,

there is a need to stimulate *kinesthetic* and *cutaneous* receptors. Kinesthesia or proprioception corresponds to the person's perception of body movement. It is achieved through mechanoreceptors within joints and muscles that also allow one to perceive force being exerted on any object. Kinesthetic devices can either be wearable or grounded. Wearable devices are attached to the user's body and help render the shape of objects. Grounded devices are mounted on a stationary platform and help render their weight. Cutaneous or tactile feedback is based on slow and fast-adapting mechanoreceptors under the skin. Each of these corresponds to a different sensation, either coming from pressure or vibration stimulus (caused by feeling texture). Pressure is normally rendered by applying a normal force on the skin through a mechanical mean. Vibration stimulus is normally given through vibrotactile stimulation. An ideal haptic device should incorporate kinesthetic and cutaneous feedback [47].

The rules of ANA Avatar XPRIZE for the Semifinals and the Finals specified the necessity of transmitting the sensation of feeling texture through tactile sensation. Thus, we decided to use commercial haptic gloves (shown in Fig. 16) that provide kinesthetic and vibrotactile feedback but no mechanical-based feedback to render pressure. The kinesthetic feedback in those gloves is implemented by using magnetic breaks that stop the motion of the fingers, transmitting force between wires and fingertips to render the shape of objects. This rendering is helpful for teleoperated manipulation. The vibrotactile feedback is provided on each fingertip, and it can be used to give the operator a different tactile sensation according to the texture of the objects that the robot's hand is rubbing.

On the other hand, rendering the sensation of weight was not mandatory for the competition. It was only necessary to inform the operator of such weight, as the objective was only to help identify a heavier object among two equally-looking samples. Thus, we decided not to use a grounded exoskeleton for the arms, a solution adopted by some teams like NimbRo [48] and team Northeastern [11]. This design choice was taken because we wanted the operator to be standing and walking in place (as explained in Section 6.2) to teleoperate the avatar as naturally as possible, and that would have required a more complex design than

Fig. 23 The weight is dynamically displayed in pink text boxes above the operator's hands (identified by a blue contour); the top picture shows the detected weight of the canister, while the bottom picture shows the force required to trigger the safety switch.

if the operator had been sitting. Instead, what we opted to do was to display the weight perceived by the wrist F/T sensors (after removing the weight of the hand) within text boxes positioned near the hands of the robot in the GUI (see Fig. 23).

We use the feedback provided by the haptic sensing system (Sec. 5.5) in three different ways:

- *Visually.* The force vectors are displayed as arrows on the palm and on each fingertip of hand markers displayed in the operator's field of view (see Fig. 24)[16]. The arrow on the palm is proportional to the object's weight, while the arrows on the fingertips are proportional to the measured contact forces. Also, the color of each arrow changes according to the measured force. The arrows are initially green and gradually turn red as the force increases.
- *As a grasping force.* The normal component of the measured force (f_z) on each fingertip of the robot's hands is used proportionally to apply a resisting force on the fingers of the operator as follows:

[16]The postures of the hand markers correspond to the ones of the operator, not of the robot, and they can be different as it will be discussed in Sec. 7.2

$$f = \begin{cases} \gamma \frac{f_z - f_{\min}}{f_{\max} - f_{\min}} & \text{if } f_z > f_{\min} \\ 0 & \text{if } f_z \leq f_{\min} \end{cases}, \quad (3)$$

where $f_{\max}$ and $f_{\min}$ are the maximum and minimum values that can be read from the sensor. Note that the forces can not pull back the operator's fingers but only make the grasping harder.

- *As vibration.* We calculate the vibrotactile haptic feedback (the amplitude of the vibration: ψ) to be triggered at each fingertip of the operator as a sum of two different components:
  - A high-frequency component (ψ_{highf}), meant to represent the nature of the material. As such, it is associated with a Coulomb's friction coefficient and, therefore, obtained by using the tangential component of the measured force (determined by f_x and f_y), as well as the normal one.
  - A low-frequency component (ψ_{lowf}), meant to capture the object's geometry. As such, it is proportional to the rate of change of the measured normal component of the force and inversely proportional to the speed of the hand (the magnitude of its velocity v_{hand}).

  That is, calculated as:

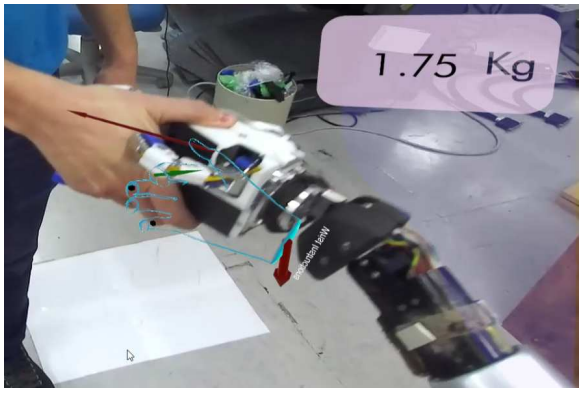

Fig. 24 Visual force feedback on the palm and at the fingertips of the hand marker (identified by a blue contour) shows the direction and intensity of each force. Note that as the orientation of the operator's hand (aligned to the hand marker) differs from the robot's due to the handshake, the visual direction of the forces is not aligned with the latter but with the former.

$$\psi_{\text{highf}} = \left| \frac{\sqrt{(f_x^2 + f_y^2)}}{f_z} \right|, \quad (4)$$

$$\psi_{\text{lowf}} = |\dot{f}_z| / \|v_{\text{hand}}\|, \quad (5)$$

$$\psi = \begin{cases} \rho\,(\psi_{\text{lowf}} + \psi_{\text{highf}}) + \psi_{\min} & \text{if } f_z > f_{\min} \\ 0 & \text{if } f_z \leq f_{\min} \end{cases} \quad (6)$$

Providing vibrotactile feedback to transmit the sensation of feeling texture (roughness) was also adopted by team NimbRo. However, they decided to incorporate microphones into the fingers of their hands (instead of F/T sensors) and use an additional vibration actuator (not the ones provided by the haptic gloves). Also, contrary to an analytical approach (like us), they used a Convolutional Neural Network (CNN) to classify the surface as rough or smooth and send the appropriate haptic signal [49].

7 Avatar Software Framework

To control the robot, we use `mc_rtc`[17], a powerful yet user-friendly open-source software framework co-developed by JRL and LIRMM for implementing controllers and designing complex robot applications [50, 51].

The `mc_rtc` framework controls the avatar robot by using the commands given by the operator and providing feedback from the robot over a network, using an embedded server-client system instead of ROS to communicate the data.

The embedded controller runs a quadratic program (QP) to compute the desired joint accelerations of the robot in an optimal way regarding a set of concurrent tasks and under linear safety and feasibility constraints. In the following subsections, we will describe this architecture and specifically give a summary of the robot model, the formulated QP together with tasks and constraints, and the finite state machine feature in our framework. Then, we introduce the implementation of high-level features of the avatar framework that rely on this architecture: the arms and hands control, the balance control, and safety features.

[17] https://jrl-umi3218.github.io/mc_rtc/

7.1 Robot Model

A humanoid robot is a floating-base mechanism; this is because there is no constraint linking the position nor the orientation of any part of the robot to the environment. The floating base is a specific joint selected as the root of the kinematic tree, and it is commonly chosen to be attached to the waist of the robot.

The robot has $n+6$ DoF, where n is the number of joints. Its configuration can be described by $\boldsymbol{q} = (\boldsymbol{p_B}, \boldsymbol{R_B}, \boldsymbol{q_\theta})$, where $\boldsymbol{p_B} \in \mathbb{R}^3$ is the position of the floating base, $\boldsymbol{R_B} \in SO(3)$ represents its orientation, and $\boldsymbol{q_\theta} \in \mathbb{R}^n$ comprises the joint angles. The *configuration velocity* is given by $\boldsymbol{\alpha} = \left[\, \boldsymbol{v_B^T}\ \boldsymbol{\omega_B^T}\ \boldsymbol{\dot{q}_\theta^T} \,\right]^T \in \mathbb{R}^{n+6}$. Here, $\boldsymbol{v_B}$ and $\boldsymbol{\omega_B}$ are the linear and angular velocities of the floating base. The time derivative of the configuration velocity, $\dot{\boldsymbol{\alpha}} \in \mathbb{R}^{n+6}$, is the *configuration acceleration* [52]. This configuration acceleration is affinely linked to the acceleration of posture and Cartesian tasks and is, therefore, the decision variable of the QP, as explained hereinafter.

7.2 QP-Based Whole-Body Control

The QP solver computes an optimal reference configuration acceleration, $\dot{\boldsymbol{\alpha}}_{\boldsymbol{r}}$, subject to linear constraints. This acceleration is then integrated twice to obtain joint commands that are realized through low-level PD joint-tracking control.

The QP is formulated as follows:

$$\begin{aligned} \dot{\boldsymbol{\alpha}}_{\boldsymbol{r}} = \arg\min_{\boldsymbol{x}}\ & \tfrac{1}{2}\left\|\boldsymbol{W}\left(\boldsymbol{A}_{\mathrm{ob}}\boldsymbol{x} - \boldsymbol{b}_{\mathrm{ob}}\right)\right\|^2 + \tfrac{\gamma_{\mathrm{QP}}}{2}\left\|\boldsymbol{x}\right\|^2 \\ \text{s.t. } & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \end{aligned} \tag{7}$$

where $\boldsymbol{W} = \mathrm{blkdiag}\left(\boldsymbol{W_1}, \ldots, \boldsymbol{W_k}\right)$ is a block diagonal matrix comprising weight matrices for k tasks [50, 53] and γ_{QP} is a small weight that minimizes $\boldsymbol{x}$ [54]. Objectives are formulated through the linear system $(\boldsymbol{A}_{\mathrm{ob}}, \boldsymbol{b}_{\mathrm{ob}})$, which vertically concatenates matrices and vectors for k tasks. Constraints are formulated similarly to get $(\boldsymbol{A}, \boldsymbol{b})$ [52].

7.2.1 Tasks

For the jth task, $\boldsymbol{A}_{\mathrm{ob},\boldsymbol{j}}$ and $\boldsymbol{b}_{\mathrm{ob},\boldsymbol{j}}$ are given as:

$$\boldsymbol{A}_{\mathrm{ob},\boldsymbol{j}} = \boldsymbol{J_{g,j}}(\boldsymbol{q}), \tag{8}$$

$$\boldsymbol{b}_{\mathrm{ob},\boldsymbol{j}} = \ddot{\boldsymbol{g}}_{\mathrm{ob},\boldsymbol{j}} - \dot{\boldsymbol{J}}_{\boldsymbol{g},\boldsymbol{j}}(\boldsymbol{q}, \boldsymbol{\alpha})\boldsymbol{\alpha}, \tag{9}$$

where $\ddot{\boldsymbol{g}}_{\mathrm{ob},\boldsymbol{j}}$ is an acceleration objective and $\boldsymbol{J_{g,j}}(\boldsymbol{q})$, $\dot{\boldsymbol{J}}_{\boldsymbol{g},\boldsymbol{j}}(\boldsymbol{q}, \boldsymbol{\alpha})$ are the jth task Jacobian and its time derivative.

Posture-related tasks (in joint or Cartesian space) are specified with acceleration objectives, $\ddot{\boldsymbol{g}}_{\mathrm{ob},\boldsymbol{t}}$, and these are implemented with PD tracking. For example, a posture task in joint space is defined as $\ddot{\boldsymbol{g}}_{\mathrm{ob},\boldsymbol{t}} = \ddot{\boldsymbol{q}}_{\boldsymbol{\theta},\mathrm{ob}}$. In contrast, position and orientation tasks of a link l in Cartesian space are defined as $\ddot{\boldsymbol{g}}_{\mathrm{ob},\boldsymbol{t}} = \dot{\boldsymbol{v}}_{\boldsymbol{l},\mathrm{ob}}$ and $\ddot{\boldsymbol{g}}_{\mathrm{ob},\boldsymbol{t}} = \dot{\boldsymbol{\omega}}_{\boldsymbol{l},\mathrm{ob}}$, respectively (for different task t), such that

$$\ddot{\boldsymbol{q}}_{\boldsymbol{\theta},\mathrm{ob}} = \boldsymbol{K_p}\left(\boldsymbol{q_\theta^d} - \boldsymbol{q_\theta}\right) + \boldsymbol{K_v}\left(\dot{\boldsymbol{q}}_{\boldsymbol{\theta}}^{\boldsymbol{d}} - \dot{\boldsymbol{q}}_{\boldsymbol{\theta}}\right), \tag{10}$$

$$\dot{\boldsymbol{v}}_{\boldsymbol{l},\mathrm{ob}} = \boldsymbol{K_p}\left(\boldsymbol{p_l^d} - \boldsymbol{p_l}\right) + \boldsymbol{K_v}\left(\boldsymbol{v_l^d} - \boldsymbol{v_l}\right), \tag{11}$$

$$\dot{\boldsymbol{\omega}}_{\boldsymbol{l},\mathrm{ob}} = \boldsymbol{K_p}\tilde{\boldsymbol{\Omega}} + \boldsymbol{K_v}\left(\boldsymbol{\omega_l^d} - \boldsymbol{\omega}_l\right), \tag{12}$$

where $\boldsymbol{K_p}$ and $\boldsymbol{K_v}$ are diagonal matrices of PD gains and $\tilde{\boldsymbol{\Omega}} = \boldsymbol{S^{-1}}(\log\left\{\boldsymbol{R_l^d R_l^T}\right\})$ calculates the error vector in orientation. K_v is by default set as $2\sqrt{K_p}$. The super-script d stands for desired values, terms without subscripts indicate current values, and $\boldsymbol{S^{-1}}\left(\cdot\right) : \mathbf{R}_{3\times3} \rightarrow \mathbf{R}_3$ is the inverse of the skew-symmetric operator [52].

Note that thanks to the regularization term γ_{QP} in Eq. (7), the approach does not produce unbounded accelerations in the vicinity of task singularities. This approach contrasts with one of the other teams, which designed their hardware specifically to avoid singular configurations [11].

7.2.2 Constraints

The geometry and the DoFs of the operator and robot are different, as are their capabilities and limitations. Therefore, not all the motions produced by the operator can be translated safely into robot motions. Furthermore, since the operator is not supposed to have prior experience or knowledge about the robot's capabilities, they are not expected to take these discrepancies into account. This means that the robot control has to consider the safety and feasibility constraints by itself.

The optimization problem described earlier needs to be made aware of these constraints, which thus need to be taken care of explicitly. The QP framework allows to constrain the problem with equality or inequality conditions. In our case, two kinds of inequality constraints were considered:

- *Joint limits constraints*. Joint range and speed limits are implemented as inequality constraints using a velocity damper approach, as done in [54].
- *Self-collision constraint*. It implements collision avoidance between relevant pairs of links [55]. It is based on the method proposed in [56] and implemented as in [54].

The QP can deal with additional constraints, such as ensuring that the expected contact forces respect friction conditions or considering floating base/torque constraints. However, in our implementation, we did not resort to these features mainly because they increase computational costs. Furthermore, as teleoperation does not rely on planning, we cannot currently foresee imminent contacts whose registration is needed to implement such constraints.

7.2.3 Finite State Machine

The QP tasks and constraints are managed by a finite state machine (FSM) that receives inputs from the operator side (HMD's 6D pose, hands' poses, emergency button signal, etc.) and accordingly executes the appropriate states to achieve the desired behavior. Each state triggers a different behavior with different tasks and constraints, thus implementing a control scheme that realizes our teleoperation framework. A simplified diagram of this control scheme is depicted in Fig. 25.

7.3 Upper-Body Retargeting

To effectively perform teleoperation, it is necessary to map the collected sensory information coming from the operator interface to the reference behavior that is set as tasks for the robot, a procedure known as *retargeting* [19]. Direct mapping of whole-body motion is not feasible due to kinematic and dynamic differences between operator and robot, which would not only lead to erroneous postures but also to the imminent loss of balance. So, there has to be some trade-off between imitation and feasibility/safety [57].

While there are methods that deal with the retargeting of the whole body [57, 58], a simple alternative (adopted by us) is to deal with the retargeting of the upper-body joints, essential for manipulation, independently from the retargeting of the legs, which is crucial for balancing and locomotion [57, 59].

Two main methods deal with the (upper-body) retargeting of motions at the kinematic level: configuration space retargeting and task space retargeting [58]. Configuration-space retargeting maps human joint angles to equivalent robot joint angles, preserving the shape of gestures. Task-space retargeting maps the relative pose of the operator's hands to the robot and is essential for adequate manipulation. Hybrid methods also exist, and they are usually realized by smoothly switching between the two main methods based on some automatic strategy [60].

Given that the kinematic structure of our robot considers redundant (7 DoF) arms, task-space retargeting through position and orientation tasks for the end-effectors is prone to find arm configurations that are different and not natural; that is, configurations in which the position of the robot elbows do not visually correspond to the ones of the operator. Achieving a natural configuration of the arms can be done through a trade-off between task-space retargeting and a strategy that achieves a similar effect as configuration-space retargeting. This strategy consists of retargeting the orientation of the operator's upper arms to the robot with a lower weight than the hands.

To implement this approach for retargeting, we collocate trackers on the hands of the operator (attached to the haptics gloves) and on the back of the upper arms (close to the elbows), as seen in Fig. 16. We transform the 6D pose retrieved from these trackers which is expressed in the operator's room (world) frame to the operator's frame. This frame is known through the 6D pose retrieved by collocating an additional tracker on the lower back of the operator. To achieve a proper mapping due to the shape mismatch between the operator and robot, the relative translations of the hands with respect to the operator's frame need to be scaled. This scale is set offline such that the robot's elbow reaches a complete extension (a singularity configuration) when the operator's elbow reaches that configuration and not before. The relative transformations of all these trackers are then set as position and orientation tasks relative to the robot's lower back frame, whose transformation from the floating base is known. These relative position and orientation tasks are necessary to allow proper control of the robot's upper limbs regardless of the operator's position and orientation.

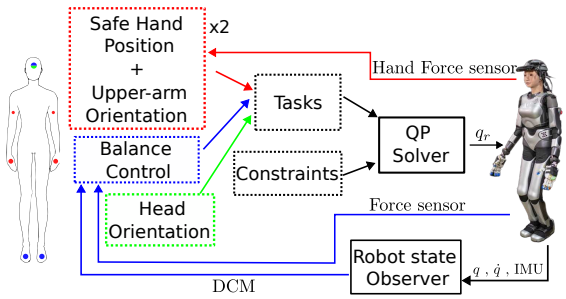

Fig. 25 Simplified control scheme: Each color represents a set of tasks that the operator can control using the corresponding sensor.

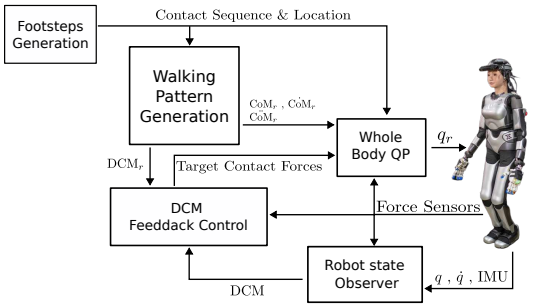

Fig. 26 Simplified scheme that focuses on walking and balance control. This scheme provides more details regarding the blue-colored components of Fig. 25.

7.4 Balance Control

The robot must guarantee its balance at any moment, either during locomotion, when manipulating an unknown object, or during both situations simultaneously. This is achieved by controlling the robot while assuming its dynamics to be described by a simplified dynamical model known as the Linear Inverted Pendulum (LIP). The LIP model assumes the robot maintains a constant center of mass (CoM) height and angular momentum. We use the LIP model under known external disturbances by using the sensorial information coming from the F/T sensors installed on the robot to account for the interaction with the environment [61].

From the LIP model, we can extract an open-loop linear Model Predictive Control (MPC) scheme to generate the centroidal trajectories for locomotion. This MPC receives as inputs the location of the reference footsteps (computed from the operator's desired walking direction) and the step duration parameters (double and single support duration), then generates the desired acceleration of the CoM that leads to balanced locomotion when following these footsteps. The pendulum dynamics are then integrated to generate a reference position and velocity of the CoM. In order to account for the discrepancy between the simplified model and the real robot, a Divergent Component of Motion (DCM) feedback control is used to guarantee a good tracking of the generated trajectories [62]. This feedback control corrects the reference acceleration of the CoM. It is then converted into desired contact forces, which are applied through admittance control [18]. Fig. 26 shows a simplified scheme of the walking control and balance strategy.

7.5 Hierarchic Inequality Admittance

Our framework cannot transmit proper force feedback to the hands of the operator (only visually). Moreover, applying a significant force on the environment could lead to a loss of balance. We, therefore, needed a control scheme that would limit the maximum force the robot can apply to the environment without disturbing the user control.

For the Semifinals, such a scheme was achieved using an admittance control [63] triggered only in specific conditions, overriding the user control. We enhanced this method by formulating a new admittance control scheme: if force constraints are violated, it will limit any motion that increases the constrained force. This way, we created a hierarchy between the maximum force one could apply and the desired motion. The critical aspect of this scheme is that only motions going against the constraint are limited, which means that the other directions are tracked optimally strictly under the force limit constraint.

Concretely, we modify the tasks acceleration objective in the following way:

$$\begin{aligned} \ddot{\boldsymbol{g}}_{\text{ob}} &= \arg\min_{\ddot{\boldsymbol{g}}} \left(\left\| \ddot{\boldsymbol{g}} - \ddot{\boldsymbol{g}}^r_{\text{ob}} \right\|^2 \right) \\ &\text{s.t. } \boldsymbol{d_i}^T \ddot{\boldsymbol{g}} < l_i \end{aligned} \tag{13}$$

where $\boldsymbol{g}$ is the position of the end effector, $\ddot{\boldsymbol{g}}^r_{\text{ob}}$ is the reference acceleration provided by the retargeting (see Eqs. (11) and (12)), d_i is the unit vector along the direction of the ith force limit, and l_i is the corresponding end-effector acceleration limit. The latter is defined as:

$$l_i = \begin{cases} +\infty & \text{if } \boldsymbol{d}_i^T \boldsymbol{f}_i < \bar{f}_i - \bar{f}_{i,m} \\ -\lambda_{i,1} \boldsymbol{d}_i^T \dot{\boldsymbol{g}} & \text{if } \boldsymbol{d}_i^T \boldsymbol{f}_i > \bar{f}_i - \bar{f}_{i,m} \\ -\lambda_{i,p} \boldsymbol{d_i}^T \left(\boldsymbol{d}_i^T \boldsymbol{f}_i - \bar{f}_i - \bar{f}_{i,m} \right) & \text{if } \boldsymbol{d}_i^T \boldsymbol{f}_i > \bar{f}_i \\ \quad -\lambda_{i,1} \boldsymbol{d_i}^T \dot{\boldsymbol{g}} & \end{cases} \tag{14}$$

where $\bar{f}_i$, $\bar{f}_{i,m}$, $\lambda_{i,1}$, and $\lambda_{i,2}$ are, respectively, the force limit, the safety margin for the ith constraint, and two positive gains. Further details of this control scheme will be provided in a future publication.

7.6 Soft Emergency Stop

As required by ANA Avatar XPRIZE, it was necessary to introduce an emergency stop that could be triggered remotely to take the robot into a safe and stable mode, and it was up to us to determine what that meant for our system.

For a humanoid robot with a high CoM, interrupting power to the actuators in an emergency will cause the robot to fall over, break hardware, and injure nearby people. In addition, simply bringing the robot to a sudden stop during a dynamic motion such as walking could also cause a similar situation. Therefore, we opted for the real-time controller to manage the emergency signal so that the robot can stop stably at the appropriate time; that is, if the robot is walking, it will stop at the end of the next step. Additionally, we decided that the emergency signal should also disable the retargeting of the hands to avoid compromising the robot's balance in the case of an unexpected situation.

8 Evaluation of the system and experimental results

8.1 Our participation at ANA Avatar XPRIZE

Our avatar system was first tested during the ANA Avatar XPRIZE Semifinals Testing (plan 2[18]). At that moment, we were using a previous version of the D-Hands (2.0), which were less robust but still could grasp and hold the heaviest object of the testing: the flower vase (see Fig. 3). These hands did not have an embedded haptic system, thus making it impossible for the operator to feel the texture of the flower vase, which was one of the tasks. On top of that, we were still using the hand-held controllers, now deprecated, as shown in Fig. 16, which could only command to open/-close the hands. It required much expertise for the operator to master the closure timing to grasp the puzzle pieces, also shown in Fig. 3.

During the testing, we got a total score of 80 points (whose breakdown is shown in Table 1 according to the scenarios described in Section 4), allowing us to advance to the Finals Qualification. A comparison of our performance with the other teams is shown in Figs. 27, 28 and 29. As can be seen, despite our deficiencies at that moment, our avatar system was well evaluated from the point of view of the operator (even getting the 2nd best score during scenario #2 and the 5th best score during scenario #3). However, it fell short from the point of view of the recipient, leading to a total score that positioned us in the 20th place.

The difference in the evaluation of the operator and the recipient is counter-intuitive and surprising, especially regarding the significant difference with respect to other teams. Due to the previously described deficiencies, we would have expected the operator to give a lower score than the recipient, who witnessed congruent body language and facial expressions. One hypothesis for this asymmetric evaluation is that the cooling system was noisy (due to the DC fans), interfering with the operator's voice coming from the speakers. However, as we had implemented a noise canceling system to remove the noise from the microphones, the sound of the DC fans was attenuated for the operator. Also, there were some other technical problems. For example, the mechanism that was driving the eyelids got damaged during the transportation of the robot, so the robot did not blink. As explained in Section 5.3, that could have influenced the communicative behavior of the recipient.

Furthermore, given that our range of motion of the neck (around the pitch axis) is smaller than the human counterpart, for the robot to be able to look at its feet when walking, the stereo camera

[18] As the Semifinals occurred during the pandemic, some teams (including ours) could not be tested during the main event in Miami, so the judges traveled and tested our system in our laboratory at LIRMM in Montpellier, France.

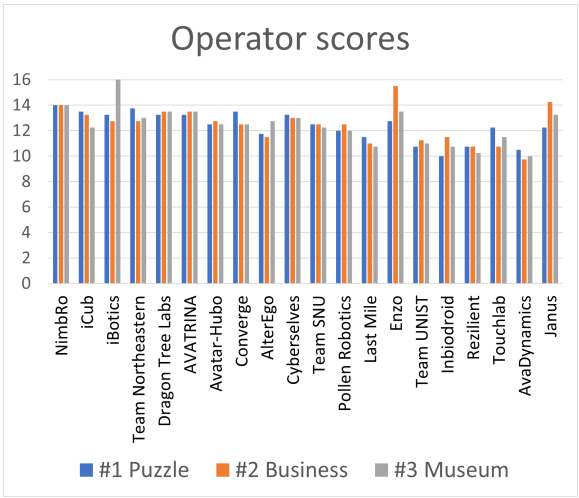

Fig. 27 Best score given by the operators to the teams for each scenario.

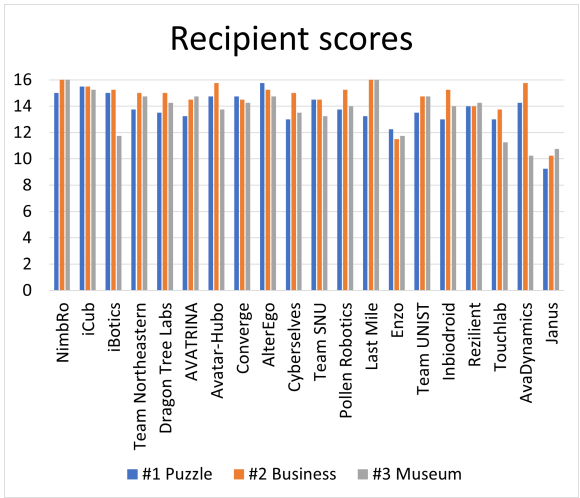

Fig. 28 Best score given by the recipients to the teams for each scenario.

Table 1 Scoring of team JANUS at the Semifinals.

Scenario	Operator (max. 15)	Recipient (max. 15)	Total
#1 Puzzle	12.25	9.25	21.5
#2 Business	14.25	10.25	24.5
#3 Museum	13.25	10.75	24.0
Submitted Video	-	-	10.0
Total			80.0

had to be installed on the helmet with an inclination of 25°(looking down). Hence, the face of the robot was "looking" upward when the operator aimed to look in a horizontal direction, resulting in an aspect of the robot that could have been uncomfortable for the recipient.

Finally, another hypothesis is that the close-to-human shape of our avatar might have raised the recipient's expectations, creating a frame of reference about which a comparative judgment was made. However, due to the technical problems, the outcome was not as good as expected and rated below that reference point. This effect is explained by the expectation (dis)confirmation theory [64], or the adaptation gap hypothesis [65].

A video showing the performance of our system at that stage of the competition is available at https://youtu.be/GnGmWgzANWU. It includes some footage of the Semifinals Video that we submitted to XPRIZE, as well as additional testing that we performed in-house.

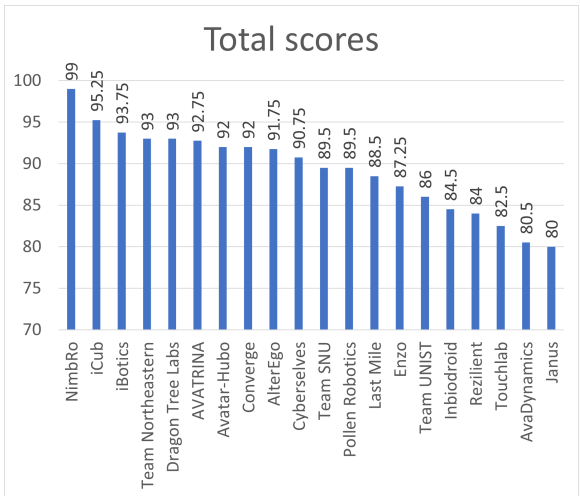

Fig. 29 Best total score obtained by each team.

For the ANA Avatar XPRIZE Finals Testing, it was required to first pass through a qualification stage to be selected as a finalist and compete during the actual two runs of the finals. During the qualification day, we trained the judge, who would be the operator of our avatar. Then, we proceeded to be tested. However, our avatar fell suddenly and unexpectedly after successfully communicating with the mission commander. Unfortunately, we do not know for sure the reason for the fall. Our logging system was not activated to avoid any interference with the real-time control of the robot. We got enough points to qualify as finalists.

However, although we repaired the robot the following day, it was not reliably walking as before, and we had to quit the competition.

8.2 In-house evaluation

8.2.1 The setup

After having missed the chance to evaluate our avatar system at the competition, we decided to assess its performance afterward in-house by realizing each of the skill-transfer tasks of the Finals (and some of the Semifinals). To do that, we designed a test course that, although not similar to the one shown in Fig. 4, would keep relevant characteristics: performing the skill-transfer tasks in a different order but in similar circumstances, navigating through narrow spaces with an equivalent narrowest gap, and locomotion over long distances. We also tried to use as much as possible the same items (canister, switch, drill) and similar relevant dimensions of the mock-up that was used at the Finals.

The eight tasks (in the tested order) were:

1. To identify the full canister that is just next to the empty one and transfer it to a designated slot.
2. To activate the safety switch (loaded with the original spring).
3. To use a drill to open a sliding door by removing a screw to reveal a display.
4. To identify the smoothest or the roughest rock (among three types) on the display using only haptics.
5. To grasp a piece of a puzzle, remove it and put it back again.
6. To make a toast by using a wine cup and to perform a handshake.
7. To navigate through a narrow space without bumping into objects while traveling a distance of about 7.5 m.

The travel distance of this entire course was approximately 17.7 m, slightly shorter than the one suggested as the objective for the exploration domain of the competition (20 m).

8.2.2 The haptic feedback

Regarding Task 4, we first investigated the feasibility of our haptic feedback through quantitative evaluation. The rocks used in the setup are shown in Fig. 30. Outside of the experiment, the fingertip corresponding to the index finger of the right hand rubbed each of the rocks. We recorded the vibration amplitude (ψ) that would be sent as vibrotactile feedback to the operator, as well as the normal and tangential forces ($f_n = f_z$, $f_t = \sqrt{f_x^2 + f_y^2}$) that resulted in those vibration amplitudes. The corresponding plots are shown in Fig. 31. The values of the parameters used for the haptic feedback (see eq. (6) and eq. (3)) are shown in Table 2.

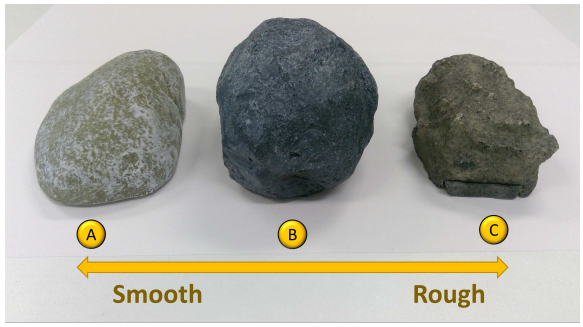

Fig. 30 Prop rocks for Task 4, labeled as A, B, and C.

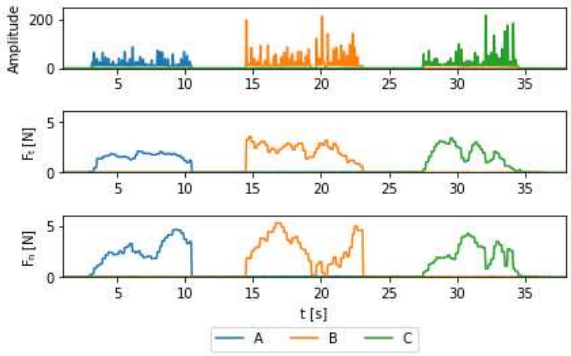

Fig. 31 Plots for the haptic feedback coming from the right index fingertip. The top plot corresponds to the amplitude of the calculated vibration (ψ), the middle plot corresponds to the tangential force (f_t), and the bottom plot corresponds to the normal force (f_n).

Table 2 Parameters used for the haptic feedback.

Parameter	Value
γ	100
ρ	2
$\psi_{\min}$	10
$f_{\min}[N]$	0.3
$f_{\max}[N]$	7

From the results of the plots, we can see that the vibration values are more homogeneous for the smooth rock (A) and more intermittent for the rough rock (C). The plot for the rock in the middle (B) shows values consistently in between the previous ones.

8.2.3 The experiment

The performance of the avatar system was assessed by testing it using the setup described in Section 8.2.1. The video showing this test is available at https://youtu.be/CaOOoSqWjCo, and hereinafter, we will provide some additional details about its execution.

Fig. 32 shows screenshots of the robot performing the canister task. It also shows how the operator triggered the sideways locomotion of the avatar by stepping in place, as well as the GUI showing the canister's weight.

Fig. 33 shows screenshots of the robot performing the switch task. Given that we used the safety switch still loaded with the original spring, the robot had to apply a force of about 5 kg·f while keeping the balance to achieve the task. To do that, we used the external forces compensation mentioned in Section 7.4. This control automatically moved the CoM of the robot accordingly to keep balance. However, it was transparent for the operator, who completed the task by slowly taking the hand down (to prevent it from slipping).

Fig. 34 shows screenshots of the robot performing the drill task. The main challenge for this task was to correctly grasp the drill such that the index finger was able to pull the trigger. Particularly, if the orientation of the drill within the hand were not very close to the ideal one, the fingertip would either slip over the trigger or try to pull it at the location of the DIP joint. In any of those cases, the strength of the hand was not enough to pull it. Because of that, we modified the drill: We used a zip tie to pre-pull the trigger and make it easier for the hand to turn on the drill.

At this point, there was a malfunction of the haptic system (it got disconnected), and the operator decided to skip the rocks task and go for the puzzle task.

Fig. 36 shows screenshots of the robot performing the puzzle task. Performing this task using the haptic gloves was more straightforward than with the hand-held controller (as we did during the Semifinals), for which we only commanded two states of the hand: opened or closed. With the haptic gloves, the operator could seamlessly control an entire range of hand closures, thus requiring less training.

Fig. 37 shows screenshots of the robot performing the toast and handshake task. The latter's objective was to test the hierarchical inequality admittance explained in Section 7.5. On purpose, the operator did not move the robot's hand, but the recipient applied enough force on it until it showed compliance.

Fig. 38 shows screenshots of the robot performing the navigation task. The operator succeeded in commanding the robot to move through the narrow space, thanks to the understanding of the environment, which was partially achieved with the help of enhanced visual feedback. See Section 6.1. This understanding is due to the image being spatially consistent with the robot's motion. Furthermore, the robot could transit from a hard floor to a carpet without losing balance.

Finally, we tested only the rocks task. Fig. 35 shows screenshots of this task. Here, the operator was able to select the correct rock. However, the operator admitted feeling lucky, meaning that the approach for the haptic feedback fell short and that it needs improvement. One reason is that the quality of the vibrotactile feedback provided by the haptic gloves was not suitable to provide different sensations. This poor quality may be because only the amplitude of the vibration can be modulated, not its frequency. Furthermore, there is a dead band where the commanded vibration is low, and the operator does not feel anything.

It is worth mentioning that the operator of our avatar system is experienced, contrary to the case of the competition where judges were the ones operating. In this regard, our current purpose was only to assess the performance of our system, which is already complicated due to the bipedal nature of our avatar. Thus, we still have some work to do in order to improve the ease of use of the teleoperation system.

9 Lessons learned

Throughout the entire competition (preparation, Semifinals, and Finals testings) we learned some lessons that we share next:

Mature vs. innovative technologies

The ANA Avatar XPRIZE competition represented one opportunity to showcase how to merge several research topics and technologies within one system. However, those kinds of events also showcase the contrast between what can be done according to the state of the art (innovative technologies) and what can be done with enough reliability (mature technologies). This is a compromise our team faced regularly, as we had to ensure that a newly developed technology could apply to our case and that the technical implementation was also reliable, especially when working together with other components.

For example, our team came up with research related to the Enhanced Visual Feedback that did not only use a decoupled viewpoint control, but managed to fill the missing areas of the FoV with a scene reconstructed from what the operator had seen using SLAM [40]. However, the scene reconstruction required a large communication bandwidth and high computation load, which could compromise the reliability of the communication with the robot.

Another example is our newly developed walking control scheme based on closed-loop model predictive control [66]. This scheme boosts walking robustness by allowing the humanoid robot to recover from multiple disturbances, including sudden pushes during walking, and by achieving locomotion over uneven and compliant grounds. The problem was that it was still in early development during the competition, and at that time, it was yet to be reliable.

These examples illustrate why most of the non-commercial technological bricks we used come from something other than the state of the art but were well embedded enough in our framework.

Constrained hardware integration is hard

If integrating hardware components is not easy, doing it reliably within very narrow hardware constraints is exceptionally challenging. These constraints came from the fact that HRP-4C was designed to be slim and low-weight, with an appearance that we did not dare to sacrifice. Every single replacement and addition (motor drivers, computing systems, sensors, DoFs, batteries) had repercussions that affected the whole system's performance, requiring re-engineering and countermeasures that had their repercussions.

For example, changing the motor drivers required not only redesigning the internal frames but also installing powerful DC fans due to the additional heat. The DC fans were powerful and drew more current, jeopardizing the encoders, a situation that was difficult to debug. Also, because they were powerful, they were noisier, affecting the recipient's experience.

Another example is the inclusion of the Jetson NANO PC plus the WiFi Card inside of the head that made us sacrifice half of the DoF of the face. As there was almost no remaining space left, we tried to use Bluetooth communication with the microphone and speakers (which had to be tiny, yet powerful). However, for some reason, the communication with those devices was very unreliable, and we ended up using wired solutions, complicating the wiring.

Finally, another example is related to the design of the hands. We implemented a clever way to deal with the abduction-adduction of the thumb and the flexion-extension of all the fingers by using only 1 DoF. However, given that the motions are sequential, the way to retarget the motion of the hands from the operator to the robot became unnatural. This behavior turned out to be an unnecessary complexity, given that the tasks at the Finals did not require the thumb to be fully adducted (there was no waving gesture required). It would have been better not to have DoF that we do not use.

Side quests can be time costly

Side quests refer to solving issues for which the team does not have much experience with but that are required for the competition. Indeed, issues related to Bluetooth communication problems, audio and video streaming, delays not related to network limitation but due to incorrect configuration, etc. were more time costly than expected and distracted us from more complex issues. What we had to do was to hire dedicated engineers that could take care of those issues.

Continuous testing is a must

The competition scenario comprised tasks that were handled by many different technological bricks. We spent much time testing each of these

bricks individually. However, we needed to conduct more general experiments to verify that all the parts were working well together or if there were no miss-functions in the long run. A dedicated group was needed to perform tests regularly instead of having tests done by the developers, who only tested their parts. On top of that, the more such a dedicated group is unfamiliar with the developments, the better. In that way, limitations that are unconsciously overlooked can become apparent, and the situation would have been closer to the actual testing at the competition.

10 Conclusions and future work

We presented in this work a telepresence framework that allows an operator to control a humanoid robot remotely to perform several tasks required for the ANA Avatar XPRIZE competition: locomotion, manipulation, communication, and haptic feedback sensing. Even if we could not showcase all of the functionalities during the competition, we could validate them afterward by using a similar experimental setup. Yet, we are aware that other components can still be added to improve the quality of the telepresence.

Prior to the finals, the competition committee requested the participants to embed the avatars with thermal feedback capability as part of haptic sensing and feedback modalities. In our team, some members already had excellent knowledge of thermal sensing and display technology. For instance, we have studied, in virtual reality and teleoperators, various thermal coupling schemes by analogy to the position (velocity)/force coupling scheme [67, 68]. For this purpose, we have used two Peltier devices equipped with temperature and thermal flow sensors. One Peltier device is mounted on the robot's finger and replicates the human fingertip thermal exchange dynamics; the other Peltier device serves as a display of the touch-sensing thermal experience at the remote location. The bilateral coupling scheme drives both Peltier devices' change in temperature dynamics (heating or cooling each of them) to have the best rendering fidelity of the thermal exchange; that is to say, high telepresence thermal sensation. The Peltier devices, with embedded sensors, are available in different sizes and are very light. Therefore, they can be mounted on force display devices (i.e., force and thermal feedback can be collocated and rendered concurrently) at the handle, as in [69], or on encounter-type force display, as in [70]. Following this background, we have started implementing thermal touch and feedback on our avatar system.

On a different topic, a critical remark raised by the judges at the end of the competition is the lack of assistance towards the operator, who was in complete charge of all the tasks. This resulted in fatigue, as well as a lack of efficiency to achieve complex tasks. The aforementioned assistance is also called "Shared Autonomy," and it is within our research interest to include it in our framework. This will improve the operator efficiency, especially for precision tasks such as the drill operation during the XPRIZE competition. The operator often had difficulty stabilizing the hand in front of the screw while holding the drill. A shared autonomy system could identify the purpose of the operator and stabilize the hand of the robot during a drilling task. Our expectation of the shared autonomy is to identify the goal that the operator is trying to achieve (e.g., grabbing a bottle) to move the robot hand toward the best grabbing spot while taking into account the operator's command (the operator can still move the hand in another direction if wished). The goal recognition part will be performed using multimodal models to generate potential goals for the operator, thanks to the camera's visual feedback. Then, the probability of each goal will be estimated through the observation of the operator's actions and by performing Bayesian filtering over a hidden Markov model. We are also planning to include augmented reality feedback to show the intention that has been detected, as well as identified affordances in real time. There are a few existing approaches for shared autonomy in the case of robotics arms and even fewer for complex robots such as humanoids. These approaches often do not deal with complex tasks to perform (requiring several steps to achieve the task), and most of the time, they do not deal with environments for which they have yet to be prepared beforehand. Team AvaTRINA was one of the few teams to feature assisted teleoperation to operate the drill (according to [11]). Why no more teams were using shared autonomy might be related to several facts: failing to perform one of the tasks in the

final would result in disqualification. Hence, most teams preferred to use technologies for which they attested reliability. It was also an investment in time to develop such a technology for their avatar, especially if there are unknown environmental factors like objects or light exposure. Lastly, one might wonder about the impact of this technology on the embodiment of the operator, and if poorly designed, it will affect the embodiment negatively.

From the limitations viewpoint, we need to improve the management of failures, especially the ones provoked by the operator when colliding with the environment or an unsuitable surface. Another improvement that we consider essential for having a good telepresence and increasing the embodiment of the operator is the FoV. In the current state of our framework, the FoV is limited by the HMD and the one coming from the camera, both being significantly smaller than the one of the human. This makes the knowledge of the position of the robot's limbs hard for the operator to understand. In this research direction, [40] developed a solution to memorize the previous image captured by the robot to reconstruct a wider FoV. However, we could not integrate it into the framework for the competition due to the required large communication bandwidth and the high computation load. Finally, the actual performances of this work need to be assessed in a human subject study to assess the task performance and the subjective impression of the operators and recipients. This is one crucial topic that we are currently working on.

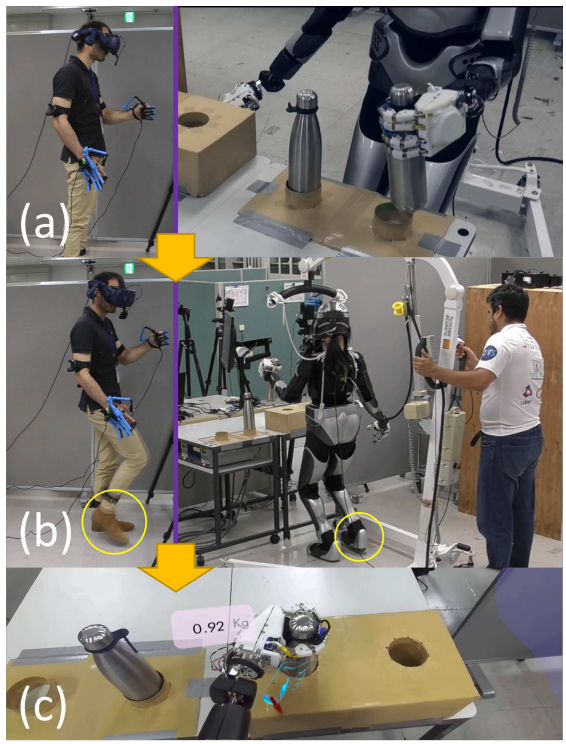

Fig. 32 Screenshots showing the execution of the canister task: (a) grasping the canister and holding it to verify its weight, (b) walking with the canister in the hand (notice the operator commanding the stepping), (c) inserting the full canister in the slot (operator's view; notice the displayed weight).

Acknowledgments

The authors would like to thank Masahiro Kato, Ryoma Koshi, Shoichi Yaguchi, and Natsumi Mashiko for their engineering work in this project, as well as Luigi Penco from Inria for the active discussions.

Declarations

Funding. This research was partially funded by the Japan Science and Technology Agency (JST) with the JST-Mirai Program, grant number JPMJMI21H4, and by JSPS KAKENHI, grant numbers JP1190410 and JP982714.

Conflict of interest. The authors declare no conflict of interest. The funders have no role in the study's design, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Ethics approval. Not applicable.

Consent to participate. Informed consent was obtained from all individual participants who tested our teleoperation system.

Consent for publication. The authors affirm that all the people appearing in the images of this manuscript are coauthors and have provided informed consent for the publication of their images.

Availability of data and materials. The video material associated with this manuscript is publicly available on YouTube.

Authors' contributions. Rafael Cisneros-Limón (RCL), Antonin Dallard (AD), Mehdi Benallegue (MB), Kenji Kaneko (KK), Hiroshi

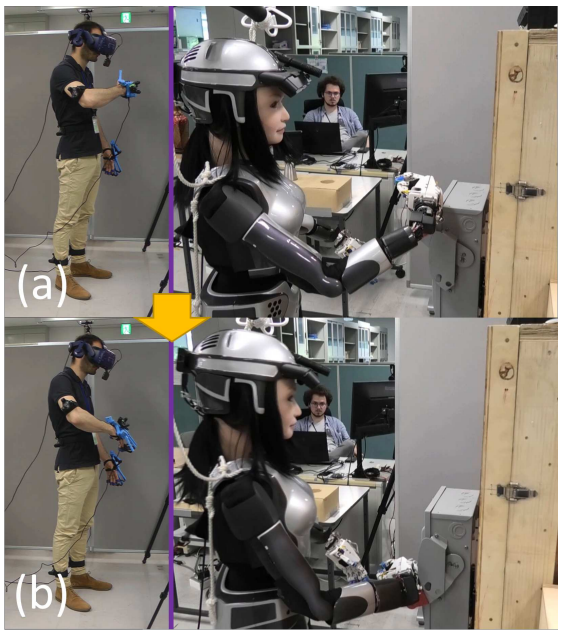

Fig. 33 Screenshots showing the execution of the switch task: (a) before changing the state of the lever, (b) after changing the state.

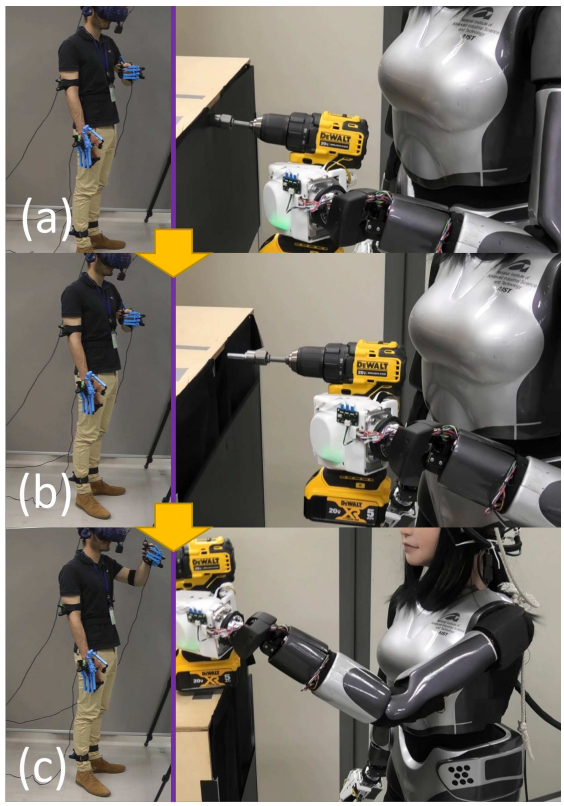

Fig. 34 Screenshots showing the execution of the drill task: (a) before removing the screw, (b) after removing the screw (notice the sliding door opening), (c) putting back the drill (notice the sliding door completely open).

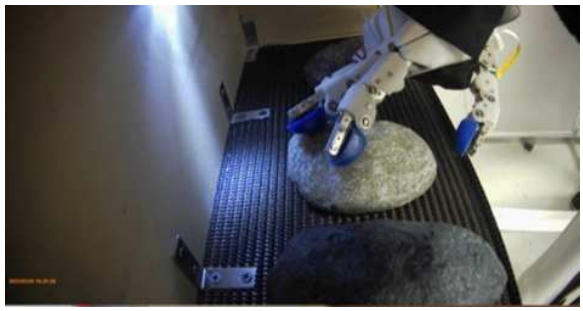

Fig. 35 Screenshot taken during the execution of the rocks task, specifically when rubbing the second rock (smoothest). Note that the operator did not have a visual of the rocks; he just knew their approximate position inside the box.

Kaminaga (HK), Pierre Gergondet (PG), Arnaud Tanguy (AT), Rohan Pratap Singh (RPS), Leyuan Sun (LS), Yang Chen (YC), Carole Fournier (CF), Guillaume Lorthioir (GL), Masato Tsuru (MT), Sélim Chefchaouni-Moussaoui (SCM), Yukiko Osawa (YO), Guillaume Caron (GC), Kevin Chappellet (KC), Mitsuharu Morisawa (MM), Adrien Escande (AE), Ko Ayusawa (KA), Younes Houhou (YH), Iori Kumagai (IK), Michio Ono (MO), Koji Shirasaka (KS), Shiryu Wada (SW), Hiroshi Wada (HW), Fumio Kanehiro (FK) and Abderrahmane Kheddar (AK) (all authors) contributed in some way to the avatar system's conception, design and development. The leading and management of team Janus was performed by AK and FK. The team name (JANUS) was proposed by AK. The technical management of the team and overview of all the development was performed by RCL, the first author. The team logo was created by MB. Mechanical and electrical improvements of the avatar robot were performed by KK and HK. Maintenance of the robot, low-level control, and system was performed by KK, HK, FK, RCL, and RPS. Simulation of the robot and the teleoperation system was performed by RCL, GL, and PG. Implementation of vision, sound, and the wireless network was done by KK, RPS, LS, and YC. Design and manufacture of the D-Hands were carried out by HK, MO, KS, SW, and HW. Integration of the D-Hands into the system was done by HK, FK, RCL, and PG.

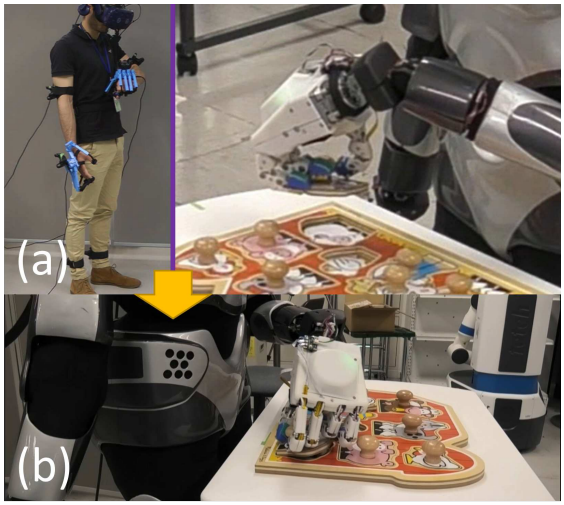

Fig. 36 Screenshots showing the execution of the puzzle task: (a) grasping and lifting the piece, (b) putting it back in its place.

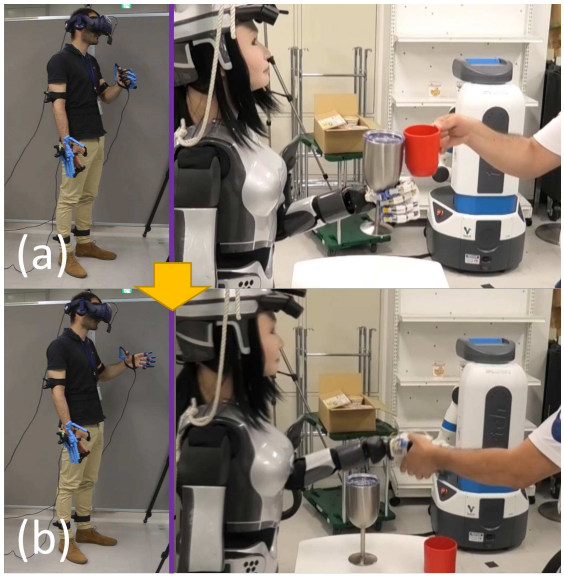

Fig. 37 Screenshots showing the execution of the toast and handshake task: (a) robot grasping the wine glass and making toast, (b) the recipient applies force to move the robot's hand to do a handshake (in this case, the operator did not move his hand, so the robot's arm showed compliance).

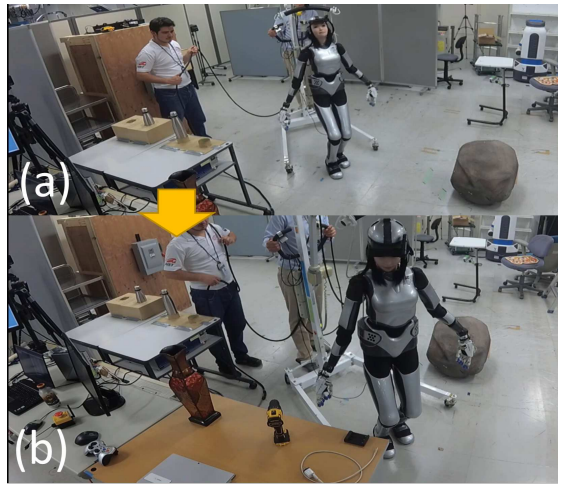

Fig. 38 Screenshots showing the execution of the navigation task: (a) heading to the narrow space, (b) moving to check the surroundings, and passing through the narrow space.

Conceptualization and development of the haptic sensory system were performed by RCL, HK, YO, SCM, and AT. Design and implementation of the E-Stop was performed by MM and AT. Preparation of the operator system was done by KA, AT, LS, YC, AD, and PG. Conceptualization and development of the enhanced visual feedback were performed by MB, YC, LS, AD, and CF. Conceptualization and development of the operator interface were done by AD, MB, CF, PG, and GL. Implementation of the transmission of expressions was done by RPS and LS. Implementation of the haptic feedback on the operator side was performed by AD, CF, and PG. Development of the software framework was done by AT, PG, AD, and KC. Conceptualization, evaluation, and implementation of upper-body retargeting were performed by MB, AE, AD, CF, and IK. Balance, locomotion, and footstep planning were improved for this project by MB, AD, and MT. Implementation of admittance control for safe interaction was developed and implemented by MB and AD. Evaluation of the avatar system was done by RCL, AD, GL, MB, PG, HK, and YH. The manuscript was written by RCL, AD, MB, HK, RPS, LS, YO, CF, MM, GL, and AK. All authors read and approved the final manuscript.

References

[1] R. Cisneros, M. Benallegue, K. Kaneko, H. Kaminaga, G. Caron, A. Tanguy, R. Singh, L. Sun, A. Dallard, C. Fournier, M. Tsuru, C. Yang, Y. Osawa, G. Lorthioir, F. Kanehiro, and A. Kheddar. Team JANUS

humanoid avatar: A cybernetic avatar to embody human telepresence. In *RSS 2022 Worshop on "Towards Robot Avatars: Perspectives on the ANA Avatar XPRIZE Competition"*, 2022.

[2] S. Tachi. Forty Years of Telexistence —From Concept to TELESAR VI (Invited Talk). In *International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments (ICAT-EGVE)*, 2019.

[3] T. Turja, S. Taipale, M. Niemelä, and T. Oinas. Positive turn in elder-care workers' views toward telecare robots. *International Journal of Social Robotics*, 14(4), 2022.

[4] S. Cunningham, A. Chellali, I. Jaffre, J. Classe, and C.G.L. Cao. Effects of experience and workplace culture in human-robot team interaction in robotic surgery: A case study. *International Journal of Social Robotics*, 5(1), 2013.

[5] M. Lei, I.M. Clemente, H. Liu, and J. Bell. The acceptance of telepresence robots in higher education. *International Journal of Social Robotics*, 14(4), 2022.

[6] ANA Avatar XPRIZE. Avatar Finalist Team Deck for Investors. https://assets-us-01.kc-usercontent.com/5cb25086-82d2-4c89-94f0-8450813a0fd3/551108bb-8ed1-473c-823d-55c2245584b7/Avatar_Finalist%20Team%20Deck%20for%20Investors%20-%20V8%20Mobile%20Friendly%204%5B1%5D.pdf, 2022.

[7] M. Schwarz, C. Lenz, A. Rochow, M. Schreiber, and S. Behnke. NimbRo Avatar: Interactive immersive telepresence with force-feedback telemanipulation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021.

[8] M. Schwarz, C. Lenz, R. Memmesheimer, B. Pätzold, A. Rochow, M. Schreiber, and S. Behnke. Robust immersive telepresence and mobile telemanipulation: NimbRo wins ANA Avatar XPRIZE Finals, 2023. arXiv:2303.03297.

[9] C. Lenz, M. Schwarz, A. Rochow, B. Pätzold, R. Memmesheimer, M. Schreiber, and S. Behnke. Nimbro wins ana avatar xprize immersive telepresence competition: Human-centric evaluation and lessons learned. *International Journal of Social Robotics*, 2023.

[10] R. Luo, C. Wang, E. Schwarm, C. Keil, E. Mendoza, S. Alt, J.P. Whitney, and T. Padır. Towards robot avatars: Systems and methods for teleinteraction at avatar XPrize semi-finals. In *RSS Workshop Towards Robot Avatars: Perspectives on the ANA Avatar XPRIZE Competition*, 2022.

[11] R. Luo, C. Wang, C. Keil, D. Nguyen, H. Mayne, S. Alt, E. Schwarm, E. Mendoza, T. Padır, and J.P. Whitney. Team Northeastern's Approach to ANA XPRIZE Avatar Final Testing: A Holistic Approach to Telepresence and Lessons Learned, 2023. arXiv:2303.04932.

[12] J.M.C. Marques, P. Naughton, Y. Zhu, N. Malhotra, and K. Hauser. Commodity telepresence with the AvaTRINA Nursebot in the ANA Avatar XPRIZE semifinals. In *RSS Workshop Towards Robot Avatars: Perspectives on the ANA Avatar XPRIZE Competition*, 2022.

[13] J.B.F. Van-Erp, C. Sallaberry, C. Brekelmans, D. Dresscher, F. Ter-Haar, G. Englebienne, J. Van-Bruggen, J. De-Greff, L.F. Silva-Pereira, A. Toet, N. Hoeba, R. Lieftink, Falcone S., and T. Brug. What comes after telepresence? embodiment, social presence and transporting one's functional and social self. In *IEEE International Conference on Systems, Man, and Cybernetics*, 2022.

[14] B. Park, J. Jung, J. Sim, S.Y. Kim, J.W. Ahn, D. Lim, D. Kim, M. Kim, S. Park, E.H. Sung, H. Lee, G. Park, J. Cha, J. Shin, and J. Park. Team SNU's Avatar System for Teleoperation using Humanoid Robot: ANA Avatar XPRIZE Competition. In *RSS Workshop Towards Robot Avatars: Perspectives on the ANA Avatar XPRIZE Competition*, 2022.

[15] 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 *IEEE International Conference on Advanced Robotics and its Social Impacts*, 2022.

[16] S. Dafarra, K. Darvish, R. Grieco, G. Milani, U. Pattacini, L. Rapetti, G. Romualdi, M. Salvi, A. Scalzo, I. Sorrentino, D. Tomè, S. Traversaro, E. Valli, P. Viceconte, G. Metta, M. Maggiali, and D. Pucci. iCub3 avatar system, 2022. arXiv:2203.06972.

[17] M. Schwartz, J. Sim, J. Ahn, S. Hwang, Y. Lee, and J. Park. Design of the humanoid robot TOCABI. In *IEEE-RAS International Conference on Humanoid Robots*, 2022.

[18] S. Caron, A Kheddar, and O. Tempier. Stair climbing stabilization of the HRP-4 humanoid robot using whole-body admittance control. In *IEEE International Conference on Robotics and Automation*, pages 277–283, 2019.

[19] 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*, 2023.

[20] Abderrahmane Kheddar. Teleoperation based on the hidden robot concept. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 31(1):1–13, 2001.

[21] A. Kheddar, C. Tzafestas, P. Coiffet, T. Kotoku, S. Kawabata, K. Iwamoto, K. Tanie, I. Mazon, C. Laugier, and R. Chellali. Parallel multi-robots long distance teleoperation. In *IEEE International Conference on Advanced Robotics*, pages 1007–1012, 1997.

[22] L. Aymerich-Franch, D. Petit, G. Ganesh, and A. Kheddar. The second me: Seeing the real body during humanoid robot embodiment produces an illusion of bi-location. *Consciousness and Cognition*, 46:99–109, 2016.

[23] D. Hanson, A. Imran, A. Vellanki, and S. Kanagaraj. A neuro-symbolic humanlike arm controller for sophia the robot, 2020. arXiv:2010.13983.

[24] Hiroshi Ishiguro. Android science – toward a new cross-interdisciplinary framework – . In *International Symposium of Robotics Research*, pages 118–127, 2007.

[25] S. Behnke, J.A. Adams, and D. Locke. The $10 million ANA Avatar XPRIZE competition advanced immersive telepresence systems, 2023. arXiv:2308.07878.

[26] K. Kaneko, F. Kanehiro, M. Morisawa, K. Miura, S. Nakaoka, and S. Kajita. Cybernetic human HRP-4C. In *IEEE-RAS International Conference on Humanoid Robots*, 2009.

[27] K. Kaneko, F. Kanehiro, M. Morisawa, T. Tsuji, K. Miura, S. Nakaoka, S. Kajita, and K. Yokoi. Hardware improvement of cybernetic human HRP-4C for entertainment use. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011.

[28] S. Kajita, T. Nakano, M. Goto, Y. Matsusaka, S. Nakaoka, and K. Yokoi. Vocawatcher: Natural singing motion generator for a humanoid robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011.

[29] P. Hömke, J. Holler, and S.C. Levinson. Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLoS One*, 13(12), 2018.

[30] M. Schwartz, J. Sim, and J. Park. Design and control of a humanoid avatar head with real-time face animation. In *International Conference on Control, Automation and Systems*, 2022.

[31] A. Rochow, M. Schwarz, M. Schreiber, and S. Behnke. VR facial animation for immersive telepresence avatars. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2022.

[32] M. Tachibana, S. Nakaoka, and H. Kenmochi. A singing robot realized by a collaboration of VOCALOID and cybernetic human HRP-4C. In *InterSinging*, 2010.

[33] N. Nostadt, D. A Abbink, O. Christ, and P. Beckerle. Embodiment, presence, and their intersections: teleoperation and beyond. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(4):1–19, 2020.

[34] A. Toet, I. A Kuling, B. N. Krom, and J. B. F. van Erp. Toward enhanced teleoperation through embodiment. *Frontiers in Robotics and AI*, 7:14, 2020.

[35] C.Y. Brown and H.H. Asada. Inter-finger coordination and postural synergies in robot hands via mechanical implementation of principal components analysis. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2877–2882, 2007.

[36] M.G. Catalano, G. Grioli, E. Farnioli, A. Serio, C. Piazza, and A. Bicchi. Adaptive synergies for the design and control of the Pisa/IIT softhand. *The International Journal of Robotics Research*, 33(5):768–782, 2014.

[37] N. Fukaya, S. Toyama, T. Asfour, and R. Dillmann. Design of the TUAT/Karlsruhe humanoid hand. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 1754–1759, 2000.

[38] NEDO. NEDO develops robot hand "karakuri" that can grab hold of various items using only simple controls. https://www.nedo.go.jp/english/news/AA5en_100344.html.

[39] J.-P. Stauffert, F. Niebling, and M. E. Latoschik. Latency and cybersickness: Impact, causes, and measures. a review. *Frontiers in Virtual Reality*, 1:582204, 2020.

[40] Y. Chen, L. Sun, M. Benallegue, R. Cisneros, R.P. Singh, K. Kaneko, A. Tanguy, G. Caron, K. Suzuki, A. Kheddar, and F. Kanehiro. Enhanced visual feedback with decoupled viewpoint control in immersive humanoid robot teleoperation using SLAM. In *IEEE-RAS International Conference on Humanoid Robots*, 2022.

[41] M. Schwarz and S. Behnke. Low-latency immersive 6D televisualization with spherical rendering. In *IEEE-RAS International Conference on Humanoid Robots*, 2021.

[42] J. Shin, J. Ahn, and J. Park. Stereoscopic low-latency vision system via ethernet network for humanoid teleoperation. In *International Conference on Ubiquitous Robots*, 2022.

[43] D. Gulhane. *The Effects of An Avatar's Facial Features On Social Presence*. PhD thesis, University of Twente, 2022.

[44] Y. Imaoka, A. Flury, and E.D. De-Bruin. Assessing saccadic eye movements with head-mounted display virtual reality technology. *Frontiers in Psychiatry*, 11, 2020.

[45] VIVE. Eye and Facial Tracking SDK. https://developer-express.vive.com/resources/vive-sense/eye-and-facial-tracking-sdk.

[46] L. Aymerich-Franch, D. Petit, G. Ganesh, and A. Kheddar. Object touch by a humanoid robot avatar induces haptic sensation in the real hand. *Journal of Computer-Mediated Communication*, 22(4):215–230, 2017.

[47] A.R. See, J.A.G. Choco, and K. Chandramohan. Touch, texture and haptic feedback: A review on howwe feel the world around us. *MDPI Applied Sciences*, 12(9), 2022.

[48] C. Lenz and S. Behnke. Bimanual telemanipulation with force and haptic feedback through an anthropomorphic avatar system. *Robotics and Autonomous Systems*, 161, 2023.

[49] B. Pätzold, A. Rochow, M. Schreiber, R. Memmesheimer, C. Lenz, M. Schwarz, and S. Behnke. Audio-based roughness sensing and tactile feedback for haptic perception in telepresence, 2023. arXiv:2303.07186.

[50] K. Bouyarmane and A. Kheddar. Using a multi-objective controller to synthesize simulated humanoid robot motion with changing contact configurations. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011.

[51] R.P. Singh, P. Gergondet, and F. Kanehiro. mc-mujoco: Simulating articulated robots with FSM controllers in MuJoCo. In *IEEE/SICE International Symposium on System Integration*, 2023.

[52] R. Cisneros, M. Morisawa, M. Benallegue, A. Escande, and F. Kanehiro. An inverse dynamics-based multi-contact locomotion control framework without joint torque feedback. *Advanced Robotics*, 34(21–22), 2020.

[53] M.A. Hopkins, D.H. Wong, and A. Leonessa. Compliant locomotion using whole-body control and divergent component of motion tracking. In *IEEE International Conference on Robotics and Automation*, 2015.

[54] J. Vaillant, A. Kheddar, H. Audren, F. Keith, S. Brossette, A. Escande, K. Bouyarmane, K. Kaneko, M. Morisawa, P. Gergondet, E. Yoshida, S. Kajita, and F. Kanehiro. Multi-contact vertical ladder climbing by an HRP-2 humanoid. *Autonomous Robots*, 40(3), March 2016.

[55] Adrien Escande, Sylvain Miossec, Mehdi Benallegue, and Abderrahmane Kheddar. A strictly convex hull for computing proximity distances with continuous gradient. *IEEE Transactions on Robotics*, 30(3):666–678, June 2014.

[56] F. Kanehiro, M. Morisawa, W. Suleiman, K. Kaneko, and E. Yoshida. Integrating geometric constraints into reactive leg motion generation. In *IEEE/RSJ Internacional Conference on Intelligent Robots and Systems*, 2010.

[57] L. Penco, B. Clément, V. Modugno, E.M. Hoffmann, G. Nava, D. Pucci, N. Tsagarakis, J.B. Mouret, and Ivaldi S. Robust real-time whole-body motion retargeting from human to humanoid. In *IEEE-RAS International Conference on Humanoid Robots*, 2018.

[58] K. Darvish, Y. Tiripachuri, G. Romualdi, L. Rapetti, D. Ferigo, F.J. Andrade-Chavez, and D. Pucci. Whole-body geometric retargeting for humanoid robots. In *IEEE-RAS International Conference on Humanoid Robots*, 2019.

[59] A. Dallard, M. Benallegue, F. Kanehiro, and A. Kheddar. Synchronized human-humanoid motion imitation. *IEEE Robotics and Automation Letters*, 2023.

[60] R. Meattini, D. Chiaravalli, G. Palli, and C. Melchiorri. Exploiting in-hand knowledge in hybrid joint-cartesian mapping for anthropomorphic robotic hands. *IEEE Robotics and Automation Letters*, 6(3):5517–5524, 2021.

[61] M. Murooka, K. Chappellet, A. Tanguy, M. Benallegue, I. Kumagai, M. Morisawa, F. Kanehiro, and A. Kheddar. Humanoid loco-manipulations pattern generation and stabilization control. *IEEE Robotics and Automation Letters*, 6(3):5597–5604, 2021.

[62] S. Kajita, M. Morisawa, K. Miura, S. Nakaoka, K. Harada, K. Kaneko, F. Kanehiro, and K. Yokoi. Biped walking stabilization based on linear inverted pendulum tracking. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4489–4496, 2010.

[63] K. Bouyarmane, K. Chappellet, J. Vaillant, and A. Kheddar. Quadratic programming for multirobot and task-space force control. *IEEE Transactions on Robotics*, 35(1), 2019.

[64] R.L. Oliver. A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4), 1980.

[65] T. Komatsu, R. Kurosawa, and S. Yamada. How does the difference between users' expectations and perceptions about a robotic agent affect their behavior? *International Journal of Social Robotics*, 4(2), 2011.

[66] A. Dallard, M. Benallegue, N. Scianca, F. Kanehiro, and A. Kheddar. Robust Bipedal Walking with Closed-Loop MPC: Adios Stabilizers, 2023. hal-04147602.

[67] A. Drif, J. Citerin, and A. Kheddar. Thermal bilateral coupling in teleoperators. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1301–1306, Aug 2005.

[68] M. Guiatni and A. Kheddar. Modeling Identification and Control of Peltier Thermoelectic Modules for Telepresence. *Journal of Dynamic Systems, Measurement, and Control*, 133(3), 03 2011.

[69] Mohamed Guiatni, Vincent Riboulet, Christian Duriez, Abderrahmane Kheddar, and Stéphane Cotin. A combined force and thermal feedback interface for minimally invasive procedures simulation. *IEEE/ASME Transactions on Mechatronics*, 18(3):1170–1181, 2013.

[70] J. Citérin, A. Pocheville, and A. Kheddar. A touch rendering device in a virtual environment with kinesthetic and thermal feedback. In *IEEE International Conference on Robotics and Automation*, pages 3923–3928, 2006.