IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 February 2024, accepted 5 May 2024, date of publication 16 May 2024, date of current version 28 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3402090

RESEARCH ARTICLE

Anywhere Is Possible: An Avatar Platform for Social Telepresence With Full Perception of Physical Interaction

GIANCARLO SANTAMATO[1,3], DANIELE LEONARDIS[1,3], SIMONE MARCHESCHI[1,3], SALVATORE D'AVELLA[1,3], (Member, IEEE), TOMMASO BAGNESCHI[1,3], CRISTIAN CAMARDELLA[1,3], DOMENICO CHIARADIA[1,3], MASSIMILIANO GABARDI[1,3], ANGELA MAZZEO[2,3], MARCELLO PALAGI[1,3], FRANCESCO PORCINI[1,3], MASSIMILIANO SOLAZZI[1,3], LUCA TISENI[1,3], PAOLO TRIPICCHIO[1,3], MARCO CONTROZZI[2,3], (Member, IEEE), CLAUDIO LOCONSOLE[1,4], AND ANTONIO FRISOLI[1,3], (Senior Member, IEEE)

[1]Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna, Ghezzano, 56010 Pisa, Italy
[2]The BioRobotics Institute, Scuola Superiore Sant'Anna, Pontedera, 56025 Pisa, Italy
[3]Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, 56127 Pisa, Italy
[4]Faculty of Technological and Innovation Sciences, Universitas Mercatorum, 00186 Rome, Italy

Corresponding author: Giancarlo Santamato (giancarlo.santamato@santannapisa.it)

This work was supported in part by European Union (EU)—NextGenerationEU Project "AVATAR: Enhanced AI-Enabled Avatar Robot for Remote Telepresence" funded by Italian Ministry of University and Research (MUR) Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) Bando 2022 funded by the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.1—Call for Tender (Call 2022) (D.R: no. 104 of 02/02/2022) (Master CUP J53D23000860006) (Decree No. 960) adopted on June 2023 funded by MUR under Grant 2022S9HAHZ and Grant D53D23001490008.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Review Board of Scuola Superiore Sant'Anna under Approval No. 152021.

ABSTRACT Robotic avatar technology has the potential to impact the future of human connectivity, transporting the sense of a human's presence to a remote location anywhere and in real-time. In this regard, the recent ANA Avatar XPRIZE challenge fostered the development of the ultimate generation of non-autonomous robotic avatar designs. This paper is devoted to introducing our system proposal, allowing intuitive motion-based control and multi-modal feedback. The teleoperated robot endorses an anthropomorphic upper body with dual arms and dexterous hands for fine manipulation of even small objects, as well as an omnidirectional platform for improved and safe locomotion. Special focus is pointed at the crucial challenge of providing self-body perception to the operator, in particular regarding control of the arms and rendering of haptic sensations. To this end, we propose an upper-limb exoskeleton and a teleoperation architecture allowing retargeting of the operator's skills and receiving tactile and kinesthetic force feedback with realistic and informative perception of the own's arms. Lastly, the platform was validated during laboratory and challenge missions mimicking several social, cooperative, and cultural scenarios from which we report on the lesson learned and future improvements.

INDEX TERMS Avatar, telepresence, teleoperation, human–robot interaction.

The associate editor coordinating the review of this manuscript and approving it for publication was Tao Liu.

FIGURE 1. Sully avatar interacting with a human recipient. Top: Remote environment. Bottom left: Human operator equipped with a full-body exoskeleton. Bottom right: Operator view.

I. INTRODUCTION

The emergence of pandemics and digital ecosystems requires the development of avatar platforms allowing telepresence for social interaction and healthcare assistance. For instance, the recent COVID-19 pandemic unveiled the need for technologies enabling real-time medical work and daily life assistance in remote locations. Likewise, the upcoming digital ecosystem, known as *Metaverse*, is encouraging a new paradigm of physical barriers in which the human experience is extended to virtual environments and social interaction could happen through avatars.

Along this direction, the ANA Avatar XPRIZE [1], a 10 M$ competition, has encouraged the development of avatar systems allowing social presence and actions in a remote real or virtual environment [2], [3]. In particular, this paper is devoted to describing *Sully*, our telerobotic system, and how it is capable of deploying a human's senses and intentions while conveying the perception of presence for both the operator and the recipient across several interaction scenarios. Namely, Sully is capable of safe locomotion in domestic environments, full 3D immersion, dexterous bimanual manipulation, and interaction with the recipient (see Fig. 1).

The conceptual design of the system is pictured in Fig. 2. The avatar has approximately a humanoid appearance, with 7 degree-of-freedom (DoF) arms equipped with custom anthropomorphic hands, and a two DoFs head carrying stereo cameras, microphones, and speakers. On the other side, the operator wears a Head Mounted Display (HMD) to fully immerse into the remote environment. Throughout the development, we focused on one challenging aspect that is providing physical feedback to the operator's arms and hands. While almost all the competing avatars endorsed tactile feedback, several teams did not address force feedback, due to the complexity of design and control.

A few stations were equipped with haptic devices and wearable exosuits providing localized force feedback in the form of a 3D force to the operator's wrist. Instead, we propose a full-arm exoskeleton enabling 3D force feedback at the wrist and the arm, allowing a realistic perception of interaction with the environment, while preserving transparency and stability. Moreover, custom exos modules were conceived for force and tactile feedback at the thumbs and indices.

Lastly, control of avatar locomotion is achieved through navigation pedals driving a four-wheel omnidirectional platform. RGB cameras mounted on the base enable recognition of the environment to assist the user in avoiding obstacles in the operator's limited vision spots.

The basic locomotion, audio/video, and manipulation capabilities were evaluated in a preliminary session in which the operator and a remote recipient interact in magic tricks. Haptic capabilities were assessed along more complex scenarios executed during the semi-final challenge and repeated in our laboratory.

In summary, our contribution is: i) discussing the design and implementation of an anthropomorphic telemanipulation robot avatar with advanced manipulation and feedback capabilities; ii) describing the operator station and the force feedback architecture; iii) lessons learned from the challenge experience involving the next generation of robotic avatars. In view of this, the fundamental novelty of the Sully system is concerned with fine haptic/kinesthetic perception that we achieved through the integration of an upper-limb exoskeleton in the teleoperation architecture.

The paper is structured as follows. A detailed report on the technical solutions contrived by the other teams is presented in Sec. II. In Sec. III, we describe the modules composing the avatar and their integration. Sec. IV is devoted to the operator station and the set of devices allowing teleoperation and feedback. Sec. V is devoted to validation experiments performed with both trained and untrained operators and focused on complex missions consisting of fine manipulation tasks - modeled according to the ANA Avatar XPRIZE competition. Sec. VI synthesizes the work and highlights the lesson learned for immediate future integration of avatar technologies in social life.

II. RELATED WORK

The avatar systems developed for the XPRIZE challenge represent the ultimate State-of-the-Art in tele-robotics. In the following, we synthesize the variegated solutions conceived to enable the main avatar capabilities in terms of aesthetic, visualization, navigation, manipulation, force, and haptic

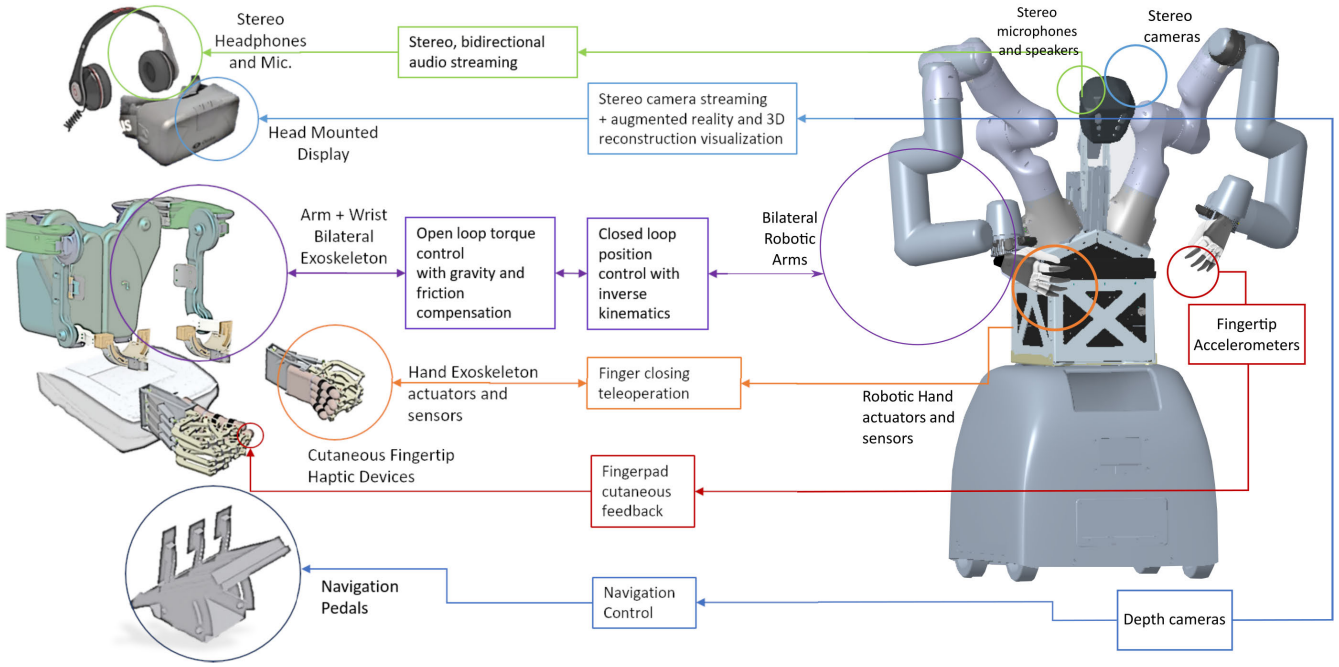

FIGURE 2. The avatar architecture, comprising the operator, the network, and the Sully avatar. The operator skills are retargeted to the robot through the control architecture, and receive feedback through the robot measurements.

feedback. A comparative analysis of the above-mentioned capabilities among top-ranked teams, including our system, is proposed in Tab. 1.

A. AVATAR AESTHETIC AND DoFs

The avatar aesthetics introduce a compromise between system complexity (number of DoFs) and the human-likeliness factor which is crucial for the acceptability with the recipient [4].

Among the finalist teams, the number of DoFs ranges from 20 of team Northeastern [5] up to 54 DoFs of iCub3 [6] which included DoFs even for the eyes movements and eyelids, achieving realistic aesthetics of a child. In this regard, the JANUS avatar [7] was even equipped with 1 DoF for mouth aperture. Besides team iCub3 [6] introduced 3 DoFs for the torso, while team SNU [8] and UNIST [9] designed a waist with 3 and 2 DoFs, respectively.

As for overall appearance, the tallness of the proposed avatars ranges from 1 m [10] to a maximum of 1.8 m [7], [8], [11], while the weight ranges from 21 kg [10] to 160 kg [5].

B. TELEMANIPULATION

Arms are fundamental for manipulation, gesture, and interaction. Most avatars implemented the classical 7 DoFs scheme (3 DoFs for the shoulder, 1 DoF for the elbow, and 3 DoFs for the wrist) through custom or commercial arm assemblies (mainly Panda by Franka Emika).

Still, novelties have been proposed. Team SNU [8] and UNIST [9] conceived a 4 DoFs shoulder by adding an additional translation to better emulate the complex human motion and allow wider workspace. Besides, the AlterEGO [10] took care on mimicking the human muscles and enhancing safety in physical interaction and control stability, introducing variable-stiffness actuators for all the arms and upper-body DoFs.

The avatar hands have been rendered in the anthropomorphic or gripper form with variegated designs and complexity. For the semifinal, team AVATRINA [12] proposed a parallel jaw for the left hand and a 4 DoFs claw for the right hand, substituted for the final by the 6 DoFs Ability Hand and a parallel jaw, respectively. Team Northeastern [5] designed a 3 DoFs anthropomorphic hydrostatic gripper. On the other side, the 20 DoFs Shunk SVH hand was adopted by team NimbRo [13]. To combine multiple capabilities, different devices can be adopted [12], [13] for the left and right hand. As an example, team AVATRINA [12] utilized the 6 DoFs Psyonic Ability Hand for the left hand, and the Robotiq 2F-140 gripper for the right hand. Besides, the left hand of NimbRo mounted the 5 DoFs Shunk SIH hand.

Arm teleoperation has been realized through commercial tracking systems (VIVE [7], [8]); robotic manipulators (Panda [13]); commercial haptic interfaces (Virtuose 6D [14]); custom suit (iFeel [6]). In particular, team Northeastern [5] developed a custom arm exoskeleton to implement translational force feedback in 3 DoFs, while team UNIST [9] designed a wearable arm haptic interface with 3 active DoFs and 3 passive DoFs.

Hand and fingers tracking from the operator is mainly achieved through haptic gloves, such as the SenseGlove DK1/DK2 [14], also combined with visual tracking systems

TABLE 1. Comparative analysis of the main avatar characteristics proposed in the course of the ANA Avatar XPRIZE challenge. A selected panel of top-ranked teams has been considered.

		AlterEGO [10]	Avatrina [12]	iBotics [14]	iCub [6]	Janus [7]	NimbRo [13]	Northeast. [5]	SNU [8]	UNIST [9]	Sully
Aesthetics and DoFs		1 m 21 kg 50 DoFs	1.75 m 157 kg 24 DoFs	1.85 m n/a 23 DoFs	1.55 m 52 kg 54 DoFs	1.6 m 50 kg 42 DoFs	1.25 m 140 kg 45 DoFs	n/a 160 kg 20 DoFs	1.8 m 100 kg 33 DoFs	1.6 m n/a 33 DoFs	1.75 m 100 kg 27 DoFs
Navigation	Avatar	Wheeled differential drive	Omni-directional drive	Wheeled differential drive	Legged (6 DoFs)	Legged (6 DoFs)	Omni-directional drive	Omni-directional drive	Omni-directional drive	Omni-directional drive	Omni-directional drive
	Operator	Handheld VR controller + Wii balance platform	Pedals	Pedals + balance control platform	Cyberith Virtualizer Sensorized shoes	Sensorized shoes + GUI + VIVE trackers	3D rudder	3 DoFs footplate + pressure sensors	T. Flight rudder pedals	3D rudder + vibrating band	Pedals
Arms	Avatar	5 DoFs custom	7 DoFs Panda	7 DoFs custom	7 DoFs custom	7 DoFs custom	7 DoFs Panda	7 DoFs Panda	8 DoFs custom	8 DoFs custom	7 DoFs Panda
	Operator	Myo arm bands	✗	✗	iFeel sensorized suit	VIVE trackers	✗	✗	VIVE tracker	✗	ALeX upper limbs exoskeleton
Hands	Avatar	19 DoFs IIT SoftHand	Psyonic, Robotiq gripper	6 DoFs Psyonic	18 DoFs hands	13 DoFs D-Hand V3.1	20 DoFs Shunk SVH, 5 DoFs Shunk SIH	3 DoFs gripper	3 DoFs gripper	6 DoFs hands	Left: 6 DoFs Azzurra, 4 DoFs Mia
	Operator	IMU	VIVE trackers	Virtuose 6D	iFeel sensorized suit	VIVE trackers	Panda	3 DoFs arm	VIVE trackers	6 DoFs interf.	7 DoFs ALeX exos
Force feedback		✗	✗	Virtuose 6D	✗	✗	Panda	3 DoFs arm	Robotic arms	6 DoFs interf.	7 DoFs ALeX exos
Haptic feedback	Vibration	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Force	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Device	VR controller	SenseGlove, Valve Index	HaptX DK2 gloves	HaptX DK1 Gloves	HaptX DK1 gloves + VR controller	SenseGlove	3 DoFs exos	4 DoFs exos	HaptX DK1 Gloves	5 DoFs HandExos
Vision	Avatar	RGB-D Stereolabs	Custom stereo 120° FoV	Stereo 180° FoV	Stereo	Stereo RGB	Stereo RGB-D	Stereo 180° FoV	Sony FDR-X3000	Stereo 90° FoV	Stereo Firefly S USB 3.0
	Operator	Meta HMD + standard screen	Valve Index HMD	VIVE HMD	VIVE HMD	VIVE HMD	Valve HMD	Standard screen	VIVE HMD	VIVE HMD	Meta HMD

[9], [12]. Instead, valve index controllers were used [7], [12] in the case of avatar gripper hands. More rarely, the operator wore hand-exoskeletons, as the 3 DoFs anthropomorphic design proposed by team Northeastern [5].

C. VISUALIZATION

Vision is essential for avatar operation. All the systems adopted stereo cameras to provide the user with stereoscopic 3D video feedback. Some systems [12] mounted cameras on linear actuators to adjust the stereo baseline to match the operator's interpupillary distance and to achieve a specified vergence, version, and tilt angle [6].

From a software point of view, a scalable resolution was adopted for transferring images depending on the available bandwidth [14], while NimbRo proposed the spherical rendering technique to achieve both low latency response to head movement and real-time streaming [13]. From the operator side, two approaches exist: a VR head-mounted display - most common - that enhances immersivity, and a screen display [5]. Team Northeastern [5] also provided the operator with images from the avatar waist camera.

D. LOCOMOTION AND NAVIGATION

Locomotion and navigation capabilities are necessary, but still not being the crucial aim of the contest. The avatars were implemented with multi-legged [15] or wheeled robots [16]. Multi-legged locomotion was implemented through 6 DoFs kinematics by several teams (iCub3 [6], JANUS [7], SNU [8], and Hubo [11]) and it represents a valuable solution in the perspective of humanoid aesthetic. Conversely, mobile-base is the most adopted solution in view of the relatively simple navigation tasks requested for the contest. Four omni-directional wheels is the most common solution. One exception is represented by the AlterEGO [10] which endorses a two-wheels base, and a control system to solve the related stability issues.

Some teams added a suspension system for distributing and supporting the robot's weight [9]. Sensors could be adopted for warning the operator about obstacles in the user's limited vision spots like distance sensors [9] or ultra-sonic sensors [12]. Usually, the operator drives the locomotion of the avatar through commercial foot pedals. Some exceptions to foot pedals are represented by locomotion platforms like the Wii balance board [10] and the Cyberneth Virtualizer [13].

E. HAPTIC AND FORCE FEEDBACK

Signals from haptic feedback can be generated by exploiting different kinds of sensors. In this regard, team AVATRINA [12] mounted pressure sensors on the avatar grippers, while team iBotics [14] utilized Psyonic touch sensors and custom pressure sensors on the hand palm. The iCub3 avatar [6] combined tactile sensors [6] and artificial skin. Team Northeastern [5] included two microphones on the wrist of the arms to combine auditory and haptic feedback.

The most common approach to render haptic information is through vibration motors, while iBotics team [14] used compressed air. Most of the gloves provided touch feedback using a system of brakes able to produce a passive force on the finger [6]. To achieve human-like surface/texture sensing ability, team AVATRINA [12] included a texture sensor based on a LIDAR camera for texture recognition. Consequently, the operator could perceive the texture through visual, vibrotactile, and also audio feedback. Team UNIST [9] developed a TENG sensor using liquid metal and silicone elastomers as a texture sensor, combined with a real-time AI algorithm capable of 94% accuracy in texture recognition.

Force feedback can be generated mostly by 6D force/torque sensors placed at the wrist of the avatar or load cells per each finger [9]. The iCub3 [6] exploited signals collected by an artificial skin distributed on the forearm and the hand palm of the avatar.

For the operator, the force was displayed through custom haptic interfaces, [5], [9] (see Sec. II-B) or commercial manipulators [13], in the form of 3D force at the operator's wrist. Force on the hand was created by a of brakes able to produce a passive force on the finger [14].

F. OTHER CAPABILITIES

All the avatars included microphones and speakers to enable aural and visual interaction and feedback that were placed generally around the avatar head, or in the torso [8]. Team iBotics also modeled thermal sensing and feedback: two 90Â°C FoV IR thermal sensors were integrated in the lower back of the robot allowing for 180Â° temperature sensing, while infrared heaters were placed in the walls of the control pod on the back and side of the operator. Some teams included emotional rendering. The basic approach consists of streaming the operator's face through a dedicated camera on a display screen attached to the avatar. More sophisticated strategies captured the facial expressions of the operator through eye and face tracking systems. In this regard, team iCub3 [6] utilized the HTC Vive Pro Eye camera and the ViVE facial tracking system. Facial expressions were displayed on the avatar side in various forms. iCub3 utilized a system based on an LED display, reproducing typical expressions like a smile, and eyelids movements jointly controlled by a single DC motor. Team SNU [8] and UNIST [9] both adopted emoticons, while team NimbRo [13] developed a system to display the operator's face with rich expressiveness even if the operator is wearing an HMD. To this goal, they extracted facial key points from the face camera and the eye tracking system and exploited a detector network based on an hourglass architecture and unsupervised learning.

III. THE AVATAR: SULLY

The Sully avatar is conceived to interact with humans during numerous everyday life activities, including social, cooperative, assistance, and teaching tasks (see Sec. V-B and ref. [10] for detailed examples of usage scenarios).

Consequently, the design process of such a robot was guided based on the need to safely and effectively interact with the human body, manipulate domestic objects, and navigate within indoor environments.

The main capabilities that we attempted to achieve are:

- haptics: perception of object shapes, texture/roughness, contact thresholds, vibrations;
- manipulation: grasping of tiny objects (necklace, thermometers, capsules), precise pick-and-place (puzzle, glasses, food) also in bilateral mode;
- visual and auditory perception: localization of people, objects, and perception of sounds;
- speaking, gestures, and body language: waving greetings, hand-shaking, giving a pat;
- mobility: mainly in-plane locomotion in typical indoor environments, obstacle avoidance.

Based on the analysis of [10], vision is obviously a required/basic capability needed in all the tasks, and implemented by all the competing teams with similar characteristics. Also hearing and speaking capabilities are fundamental in social interaction, even though their occurrence is less dominant (15% of the tasks).

On the other side, half of the tasks involve manipulation and haptics, implying the need for human-like dexterous kinematics for the upper body, in particular for arms and hands. Besides, the completion of finesse pick-and-place tasks, exploration, and learning introduces the issue of retargeting touch and force to the operator, implying the need for force sensing on the arms and the grasp. With respect to other proposed solutions/teams, we focused on fully immersive control and perception of the upper limbs: this involves the use of a full upper-limbs exoskeleton with kinesthetic and tactile haptic feedback, as it will be disclosed in Sec. IV. Mobility is necessary, even though basic locomotion ability is required. Consequently, mimicking the human lower body is not strictly required.

In addition, the avatar robot must meet the following requirements: i) the total weight, including the power source, must not exceed 160 kg; ii) the width and length must be no more than 100 cm x 120 cm maximum; iii) the total height must be less than 210 cm; iv) the avatar must be able to safely operate indoor and must be safe for humans. In the following, we discuss the main modules of the Sully avatar.

A. UPPER-BODY: MANIPULATION AND HAPTICS

The Sully avatar features an anthropomorphic upper body, Fig. 3, with 175 cm of tallness, and a weight of 100 kg. The weight is distributed as follows: about 50% of the weight is on the navigation platform, 35% on the anthropomorphic arms, and 15% on the torso and head. The robot counts 27 DoFs (7 for each arm, 5 for the left hand, 3 for the right hand, 2 for the head, and 3 for locomotion).

Two 7 DoFs Franka Emika Panda arms are mounted in a V-shaped angle configuration to mimic the human arm pose. The shoulder height of 130 cm above the floor allows convenient manipulation of objects on a table, as well as interaction with both sitting and standing persons. Besides the shoulder width of 85 cm facilitates the navigation through standard passages. The Panda arms allow a payload of 3 kg and ensure a maximal reach of 855 mm. The extra degree of freedom of the kinematic allows further flexibility in the elbow position. The purposes of the arms are dual: precisely retargeting the pose of the operator's wrist, and allowing direct inducement of forces at the wrist, thus facilitating force feedback.

To obtain versatile grasping capabilities, two different end-effectors with complementary strengths are incorporated. The Mia hand [17], [18], developed by the Prensilia research group, is mounted on the right arm, as shown in Fig. 3. It endorses actuation on 3 of the 4 fingers DoFs, performing 7 of the 10 main gestures used in 80% of our daily movements: cylindrical grip, precision grip, lateral grip, pointing up, and pointing down as examples. The maximum gripping force reaches up to 70 N in any type of grip with a weight of 480 g. Force sensors on the fingertips (3 normal + 3 tangential) allow measurement and control of the grasping force.

Instead, the IH2 Azzurra by Prensilia [19], [20], mounted on the left side, as in Fig. 3, is endorsed with underactuated self-adaptive fingers with manually adjustable stiffness. The actuation with a weight of 640 g achieves a maximum force of 35 N in the cylindrical grasp and 7 N in the lateral grasp. Its design allows the flexion and extension of the thumb, while the index and middle are independent, as well as the thumb abduction and adduction. All the fingertips are compliant and consequently, fingers automatically wrap around objects.

For both hands, dynamic interaction with the environment to be rendered through tactile rendering (i.e. contact threshold, vibrations) is estimated through accelerometers located close to the fingertips.

B. VISION MODULE

The vision module of the avatar system consists of two Firefly S USB 3.0 mounted within a face-like case, Fig. 3, and combined to provide a stereo vision to the human operator. The head inclination can be manually regulated, e.g. passive DoF, while its rotation can be targeted to the orientation of the operator's head via a motorized joint.

The baseline among the two lenses of the cameras has been designed to mimic the mean distance of human eyes, which is 6 cm [9], [21], to give the best sensation to the human operator. The two cameras are color cameras that can run up to 60 fps. Table 2 lists the main technical specifications. From the software architecture point of view, the module follows a component-based design, where the ROS concept of nodelet is used.

Nodelets are used to provide a way to run multiple code flows on a single machine, in a single process, without incurring copy costs when passing messages intra-process. In practice, nodelets are threads that run with the

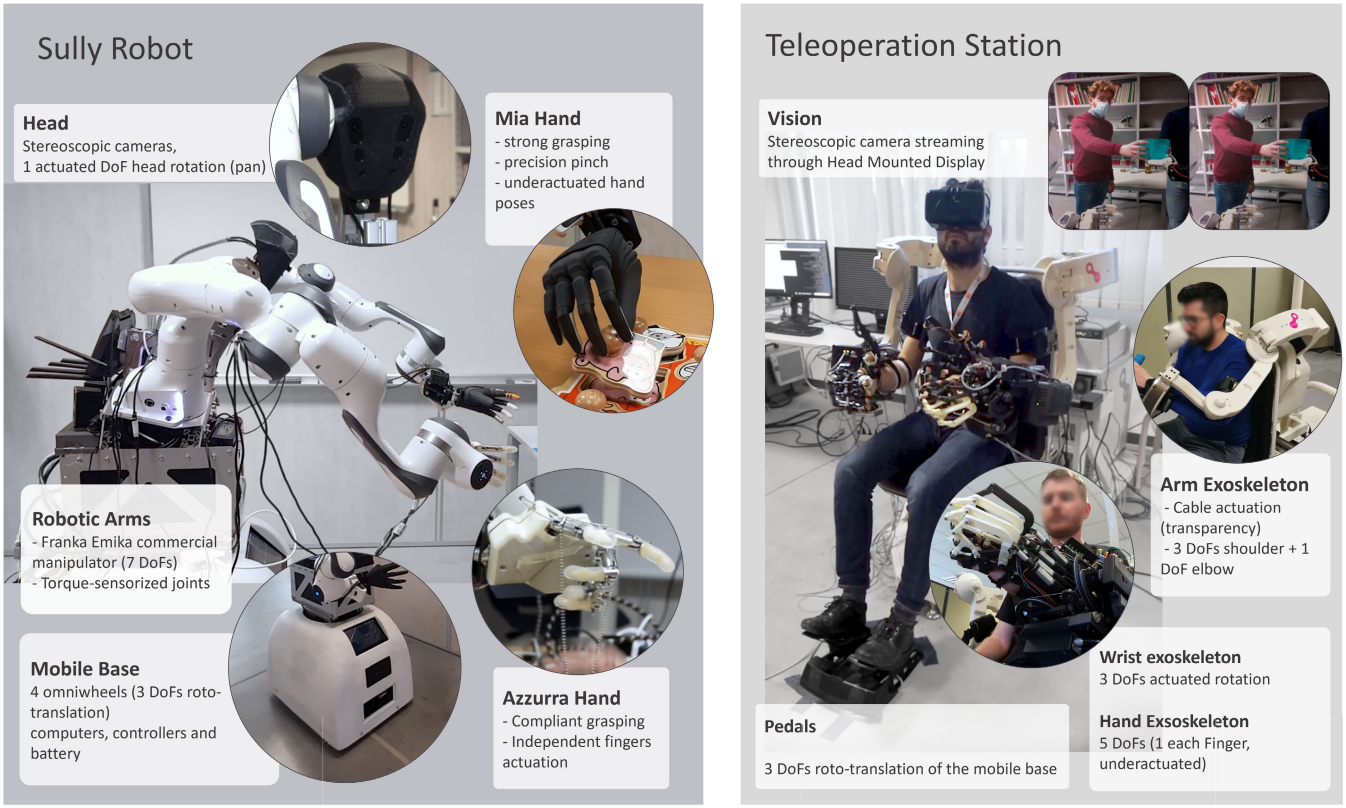

FIGURE 3. Left: comprehensive picture of the Sully avatar and its main subsystems allowing to transport the sense of presence at a remote distance. Right: the operator station: in particular, the ALeX upper limb exosuit allows the operator to intuitively control the remote avatar in bimanual manipulation tasks while providing full-force feedback.

TABLE 2. FLIR Firefly S USB3.

Frame Rate	60 FPS
Pixel Size	3.45 μm
Resolution	1440 x 1080
Sensor Type	CMOS
Sensor Format	1/2.9"
ADC	10-bit
Chroma	Color
Megapixels	1.6

same process, i.e., the nodelets manager. Such threads can communicate with each other efficiently, doing zero-copy pointer passing between publish and subscribe calls of the manager. Different from regular nodes that use TCP/IP as communication protocol, nodelets share the memory, which is much more efficient when dealing with heavy topic information such as camera streams.

In the proposed work, each camera runs within its own nodelet manager: in particular, the manager handles the node driver that acquires the images and publishes the frames on the appropriate topic and the node that applies undistortion and rectification based on the intrinsic and extrinsic information obtained after the calibration procedure sending it to the Oculus at 20 Hz throughout an efficient web socket.

C. MOBILE BASE: LOCOMOTION

The lower body of the robot consists of a holonomic platform with four omni-wheels to obtain both in-place rotation and translation of the base in any direction. This feature is particularly relevant to adjusting the workspace in manipulation tasks, with fine adjustments of the robot position with respect to the desk and manipulated objects. The avatar is then capable of up to 2 m/s movement, although we usually cut the operator command at 0.5 m/s for safety reasons.

The four wheels are actuated by brushless servomotors with gear reducers. The base hosts the power system with battery and voltage adapters, motor drivers for the base, arms, and head of the robot, and a main PC unit. The base is equipped with two Realsense D-435i RGB-D cameras for detection of the environment. A SLAM algorithm based on RTAB-Map [22] implemented in ROS is used to construct an occupancy map of the environment, as shown in Figure 4, and to assist the user in avoiding obstacles when moving with limited visibility (i.e. in lateral or backward movements).

D. NETWORK ARCHITECTURE AND OTHER FEATURES

Sully possesses stereo microphones and speakers to enable hearing and speaking with/from the remote environment. Besides, it is powered by a custom-made battery - 24 V

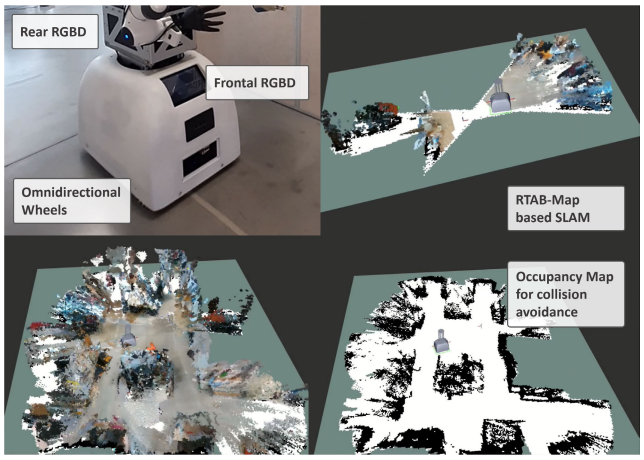

FIGURE 4. The mobile base while constructing an occupancy map of the environment, used to assist the operator in collision avoidance.

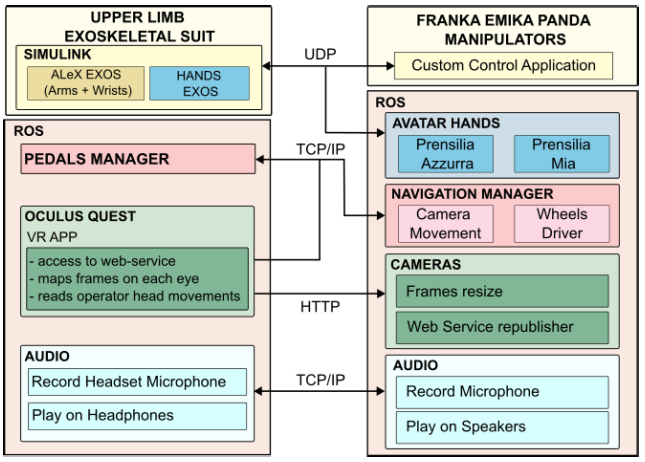

FIGURE 5. Schematic of the communication network architecture and protocols.

and 200 A·h. According to the avatar characteristics and the tasks described in Sec. V, the achieved autonomy is around 4 hours.

The avatar PC hardware embeds a six-core processor AMD Ryzen 5 2600x and a GPU NVIDIA Corporation TU117, GeForce GTX 1650 with 16 GB of memory. The operating system is Ubuntu 18.04 Long-Term Support (LTS). Algorithm coding was implemented through different programming languages: Matlab/Simulink, C++, and Python. In particular, a component-based architecture was adopted in the programming design. Accordingly, each logic function can be conveniently assigned to a dedicated node. Lastly, intra-node communication allows the accomplishment of the main-level functions.

The connection to the robot can be established through an Ethernet cable or wireless via a standard 5 GHz Wi-Fi network. The robot central unit communicates with a series of boards distributed on the robot body and connected via an Ethernet bus.

Both the robot and the operator system require a cluster of different PCs, connected in a local area network (LAN), running multiple applications at once on different operating systems. The communication between the avatar and the operator station is delivered via the schematic pictured in Fig. 5.

A compiled Simulink scheme is used to retrieve kinematic data from both the ALeX upper limb and HandExos exosuits (see Sec. IV) and build setpoints for the Panda manipulators on the avatar side. In particular, data are sent to a custom-compiled application running on a dedicated PC. The exoskeletal module also sends setpoints to a ROS node which is responsible for opening/closure of the Prensilia hands fingers. On the other side, kinesthetic and force feedback are generated within the Panda custom application PC and sent back to the Simulink module. Also, the Prensilia hands detect contact with physical objects and send tactile feedback data back to the hand exoskeleton to perform cutaneous feedback on the operator. To achieve the real-time requirements, the haptic and teleoperation modules are connected with the User Datagram Protocol (UDP) with an experimentally optimized bandwidth of up to 150 Mb/s.

The Pedals Manager takes care of reading serial data from the pedals and publish on a ROS node which is accessed by a subscriber on the avatar side. In this way, the platform wheels driver can target forward, backward, and lateral slide motion.

The vision module - Oculus Quest - on the operator's side, includes three nodes accounting for i) querying frames from each camera; ii) resizing each frame matching the Oculus sight features; iii) republishing the compressed version on a web service. Besides, a C# ROS wrapper is used in a Unity VR application to get access to the Web Service, retrieve the frames, and parse them as textures on two planes to render the stereoscopic vision. The video streaming is transmitted via the HTTP protocol. In this regard, it should be observed that most of the bandwidth is occupied by the camera streaming. Hence, when needed, the bandwidth can be tuned by changing the compression ratio of the image resolution. Also, head orientation data are sent to a ROS node through TCP/IP to retarget the operator's movement through camera rotation.

Lastly, the audio module on the operator's side opens a ROS node featuring TCP/IP data streaming toward a corresponding audio module on the avatar side. Two nodes are included: one that records data from the headset microphone, and a second responsible for playing on the headphones. Correspondingly, record and play nodes on the avatar side receive data from the operator module and transmit data to the avatar speakers.

IV. THE OPERATOR STATION

In the Sully avatar system, presented in Fig. 3, the operator intuitively controls the avatar through a full-body suit, including the following devices:

- the ALeX upper body exoskeleton;
- two hand exos modules;
- the Oculus Rift Headset;
- the pedals platform.

The manipulation interfaces exploit the headset, the ALeX exoskeleton, and the hand-exos to control the robot head, arms, and fingers respectively. The details concerned with these subsystems are discussed in Sec. IV-A and IV-C.

The headset is fundamental for visual and auditory feedback. The images acquired by the robot cameras are displayed to the headset, allowing the operator to have a first-person perspective. At the same time, the audio captured by the avatar microphone is played on the operator's headphones. Besides, the pose of the headset is tracked in real-time defining a reference for the motion of the robot head. The headset endorses a Meta Quest 2 Oculus system, providing 1832×1920 pixels per eye with a maximum refresh rate of 120 Hz, and a 113.46° diagonal field of view.

In terms of locomotion targeting, we adopted navigation pedals to achieve omnidirectional driving, in which the operator targets translation velocity by pitching the pedals. This choice is motivated since the avatar's lower body is considerably different than the bipedal human lower body. Accordingly, it was not required the use of complex systems - like the ALeX exoskeleton for the upper arm - to control the four wheels.

A. THE ALEX EXOSUIT: A PLATFORM DESIGNED FOR HIGH-TRANSPARENCY TRACKING AND FORCE FEEDBACK

The ALeX platform [23], [24] is a full upper-arm, bilateral exoskeleton, responsible for providing distributed force feedback on both arms and consequently, it is an essential part of our telepresence system. Despite lightweight manipulators exist to target the arm motion [25], the use of exoskeletons has the advantage of enhancing the operator's ability in all the physical interactions with the remote environment, i.e. manipulation, grasping, hand-shaking, etc., and in general, enriches the embodiment experience [26].

As shown in Fig. 3, the kinematic structure of ALeX is composed of a serial chain of rigid links and rotational joints distributed as follows: 3 DoFs for the shoulder (3 revolute joints constituting an equivalent spherical joint), 1 DoF for the elbow, and 3 DoFs for the wrist (spherical joint). Such a kinematic structure can cover about 90% of the natural workspace of the human upper arms without singularities. Consequently, even complex motions of the operator can be completely retargeted to the avatar with a few limitations of the natural workspace. Besides, the characteristic open structure of ALeX is meant to prevent collisions during bimanual tasks, thus strengthening comfort and safety.

Moreover, all the ALeX joints are actuated and hence, force interaction of the avatar can be feedback all over the operator's arm, attempting to target a realistic sensation of interaction with the remote environment. In order to reduce the reflected mass and inertia on the user, ALeX endorses a remote actuation: electric actuators - from the joint to the elbow DoFs - are placed remotely, behind the operator seat, and connected to the joints through in-tension metallic tendons that are routed over several idle pulleys.

All seven actuators are sensorized with digital encoders, while the three shoulder joints and the elbow joint are also equipped with analogical Hall effect sensors (by Honeywell) measuring local rotational displacements.

B. POSITION-FORCE TELEOPERATION ARCHITECTURE AND OPERATOR'S ARMS FORCE FEEDBACK

The teleoperation architecture - implemented for both arms/wrists as in Fig. 6(a) - transforms the operator's arm positions on the avatar arm joints level. The fundamental schematic is shown in Fig. 6(b).

The transformation matrix A_d, containing the 6D end-effector pose of each operator's arm, is estimated through the direct 7 DoFs kinematics of the exos and sent to the avatar as a reference. At this point, an impedance control computes the necessary joint torques through inverse kinematics, based on the actual pose of the Panda end-effector A_m.

The Panda arms are driven by specifying to the controller the joints torque vector τ_d (external controller mode). By default, the Panda control architecture (version 3) embeds gravity compensation for a horizontal configuration of the base. Hence, when the Panda is mounted in the *natural* horizontal configuration, no further compensations are needed to balance the gravity joint vector τ^0_{gra}. Otherwise, when the robot is tilted by an angle $\theta \approx 17\hat{A}°$, a gravity correction τ^θ_{gra} must be computed based on a mathematical gravity model. It follows that the resultant external torque acting on the Panda is:

$$\tau_{ext} = \tau_{env} + \tau^\theta_{gra} - \tau^0_{gra} \tag{1}$$

where: τ_{env} is the torque vector due to the interaction with the environment.

Consequently, the desired joints torque τ_d required to the Panda controller is expressed by:

$$\tau_d = J^T \left(K_p e + K_v \dot{e}\right) + \tau^\theta_{gra} - \tau^0_{gra} \tag{2}$$

where: J stands for the Jacobian matrix of the Panda; $\mathbf{e}$ is derived from the error between the Panda pose A_m and the operator pose A_d; K_p and K_v are the coefficient of the proportional and derivative contributions of the impedance control, respectively.

On the other side, the resultant 6 DoFs external force F_d acting on the Panda end-effector is:

$$F_d = J^{-T} \cdot \tau_{ext} \tag{3}$$

This force is fed back to the exos for the purpose of force rendering. The ALeX exos control scheme, depicted in Fig. 6(c), adopts an open-loop strategy. Consequently, adequate feedforward contributions are needed to counteract the main disturbances appearing during the operation of the exoskeletons. A first contribution is the gravity compensation

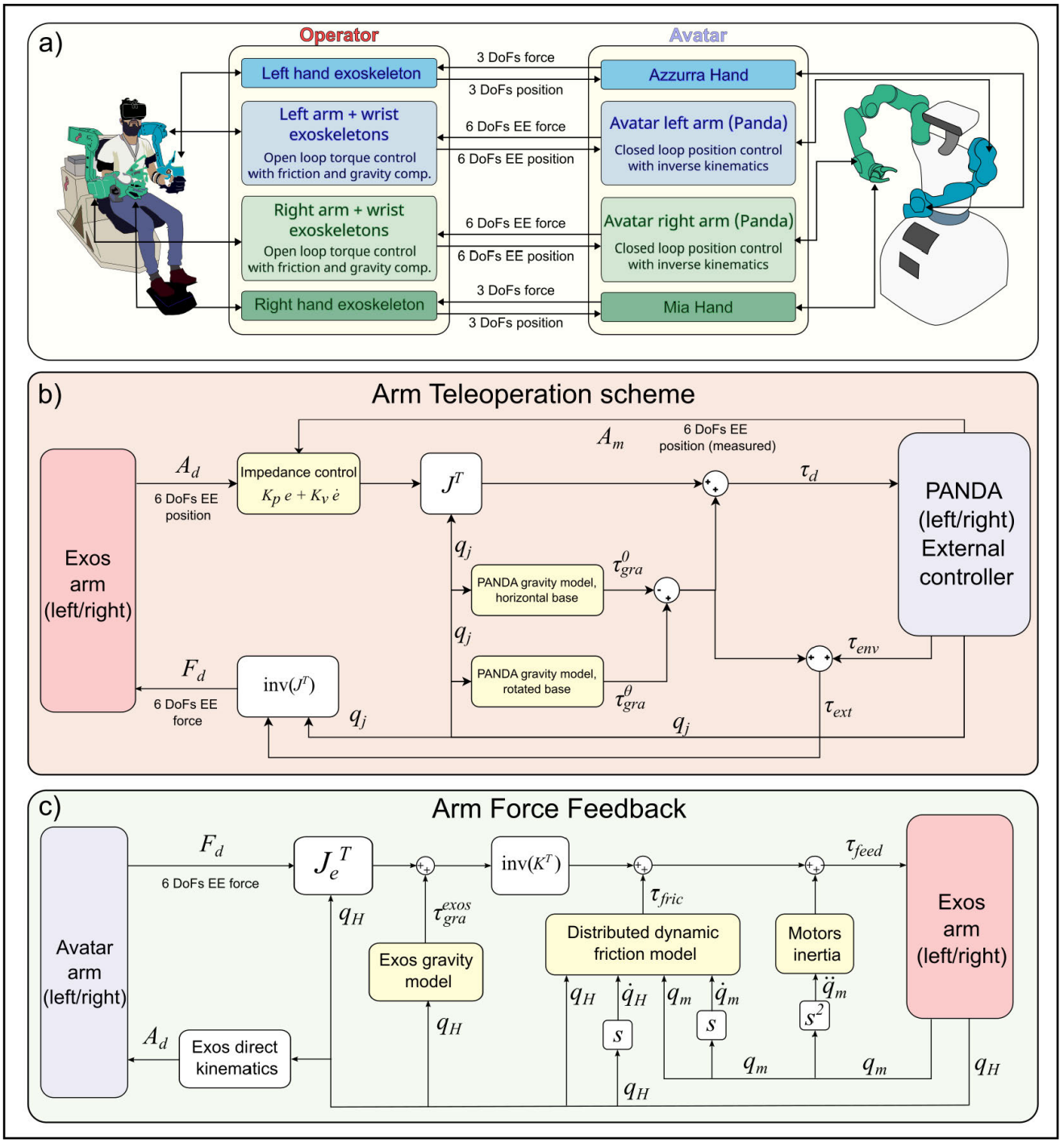

FIGURE 6. (a) An illustration showing the multilateral position-force teleoperation architecture. While hand positions are perceived on the joint level and hand forces are measured for each finger, arm positions, and forces are projected to the end effectors (EEs). (b) Detailed control scheme of the arms teleoperation. (c) Schematic of the force feedback implemented to display the remote force to the operator's arms.

torque τ^{exos}_{gra}, computed based on a mathematical gravity model of the exos, and the current joint positions q_H.

The presence of distributed friction losses throughout the tendon transmission and the effect of coupling between joints is addressed through the friction model, proposed in [27]. This approach allows us to iteratively determine the parameters of a dynamic friction model at each joint, by exploiting only the position and velocity estimates of the exos actuators q_m, $\dot{q}_m$ and joints q_H, $\dot{q}_H$, and to consequently, compute in real time the friction compensation torque vector τ_{fric}. Lastly, actuator inertia is compensated based on the estimate of the motors' acceleration vector $\ddot{q}_m$ and the moment of inertia diagonal matrix I_m. No further dynamic compensations have been introduced due to the

low-acceleration regimes of the avatar tasks. Hence, the overall torque τ_{feed} that is displayed to the exos actuators is:

$$\tau_{feed} = K^{-T} \cdot \left[\tau_{gra}^{exos} + J_e^T F_d\right] + \tau_{fric}\left(q_m, \dot{q}_m, q_H, \dot{q}_H\right) + I_m \ddot{q}_m \quad (4)$$

where: K is the transformation matrix from the joint to the motor position space, constituted by the constant transmission ratios of the tendon actuation, while J_e is the Jacobian matrix of the exos.

From Eq. 3 - 4, it turns out that when τ_{ext} is null, i.e. no force is fed back from the remote environment to the operator, the compensation terms still contribute to generating a weightless and frictionless feeling for moving the arm.

Besides, a common frame is set for all the terms in Eq. 2 - 4, defined in the center of mass of both the operator and avatar hands. In this way, the teleoperation and force feedback dataflow are referred to in this common frame before being transmitted.

The controller runs with an update rate of 1 kHz.

C. HAND-EXOS MODULES AND TELEOPERATION

Two hand exoskeleton modules are used to track the finger motions of the robotic hands and to provide grasping and tactile haptic feedback. For the hand exoskeleton, we adopted a custom device, based on [28] and [29], combining kinesthetic force feedback at finger phalanxes, with wide-bandwidth tactile feedback at fingerpads. For kinesthetic force feedback, the exoskeleton implements underactuated parallel kinematics (1 DoF for each finger) closed on the same finger kinematics and actuated by lead-screw DC linear motors. The advantage of the adopted kinematic design is to transmit only linear forces between exoskeleton links and skin tissues (torque decoupled), hence improving the wearability and stability of the device. A strain gauge force sensor is added for each actuator to obtain transparency through a force-velocity admittance control.

In addition, linear voice-coils actuators are implemented at the index and thumb fingerpads, in order to provide high-dynamic tactile signals to the operator (i.e. contact thresholds, vibrations). These signals are otherwise difficult to render through high-force but low-bandwidth compact actuators with mechanical reduction. Fig. 7 shows a scheme of the transmitted signals and control modules related to the teleoperation of fingers.

D. EXPERIMENTAL ASSESSMENT OF INTERACTION STABILITY

A crucial concern when designing teleoperation interfaces is to achieve stable dynamics during the interaction with the remote environment. In this regard, designing a control strategy is essential to prevent the onset of instabilities. Accordingly, dedicated experiments were performed for both the arms and hands subsystems of our avatar. As a practice, the stability is experimentally assessed through contact

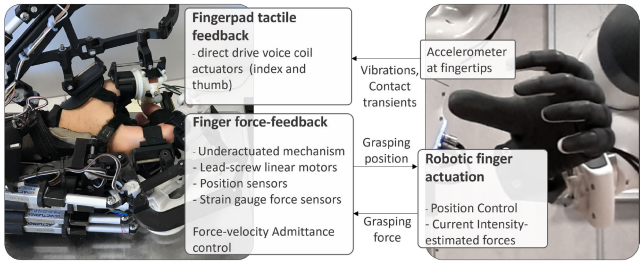

FIGURE 7. Scheme of the hands modules and provided signals during teleoperation.

experiments in which the follower (avatar arm/hand) comes in contact with a stiff remote environment. The effectiveness of the control strategy can thus be evaluated by studying the behavior during interaction transients.

The arms teleoperation scheme embeds a Time-Domain Passivity Approach (TDPA) [30], based on a passivity observer and a passivity controller for both operator (leader) and avatar (follower) parties. The passivity observers estimate the amount of energy injected into the systems due to delay in the communication channel. The passivity controller of the leader prevents instability by dampening the commanded force based on the injected energy estimated by the leader's passivity observer. Similarly, the passivity controller of the follower robot cuts the commanded speed based on the injected energy estimated by the follower's passivity observer.

During stability assessment, the end-effector of the avatar is driven in contact with a table which is admitted to be a stiff environment. Experiments are executed in a teleoperation mode, i.e. the operator is wearing the ALeX exoskeleton. Starting from a no-contact position, the operator is asked to drive the avatar in contact with the table, keep the contact for approximately one second, and then leave. Experiments have been performed separately for spatial directions, as shown in Fig. 8(a)-(b) where we plot the positions of the operator/exos and the avatar along the horizontal and vertical direction, respectively. Linear paths represent the approaching and leaving phases, while constant profiles represent the contact phase. In particular, two consecutive contact transients are reported for the vertical direction. Position trajectories are close, demonstrating the capability of the avatar to track the remote operator's movements before, during, and after the contact.

Interaction forces are plotted in Fig. 8(c)-(d). The red line stands for the interaction force between the avatar arm and the table, while the interaction force between the exoskeleton and the operator is plotted with the blue line. More interestingly, we also represent the contribution of the Passivity Controller (PC) fed to the exoskeleton (dot-dashed line). Such a force is computed as the sum of the interaction force between the avatar's arm and the remote environment and a damping term, proportional to the difference of the observed energies. Accordingly, after some unavoidable oscillations at the contact transients, the avatar can keep

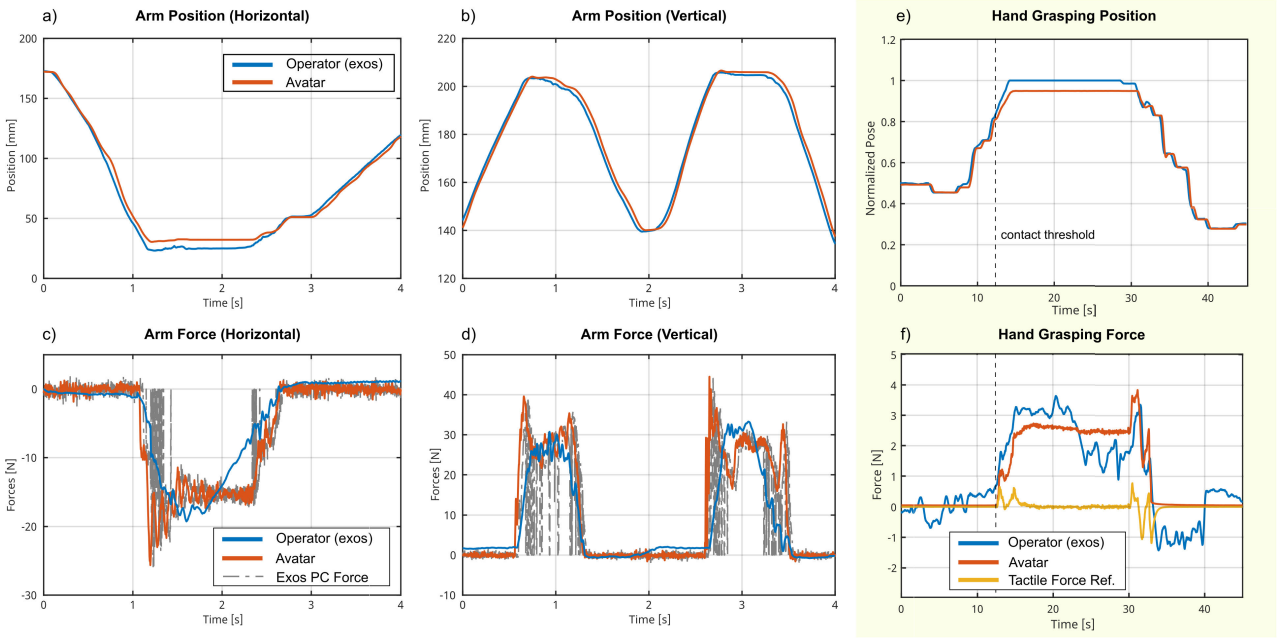

FIGURE 8. Measured signals during remote contact teleoperation experiments: contact transients at the arm level in the horizontal direction (a)-(c), in the vertical direction (b)-(d), and contact transition during hand grasping (index finger) (e)-(f).

the contact (step regions where the force is approximately constant) and continue the interaction with the environment, while the operator can even slightly exceed the target position (penetration of the virtual wall) without the occurrence of divergent and undesired bounces/tremors.

Contact experiments were also performed for the hands subsystem. Fig. 8(e) shows a simple grasping transient measured during the teleoperated grasping of a plastic cup. For simplicity, signals refer to the index finger of the HandExos and of the teleoperated robotic hand. In Fig. 8(e), the position is normalized between the fully open pose and the fully closed pose. After the contact threshold with the object, the reference position followed by the avatar hand starts to diverge from the reference position generated by the exoskeleton, due to the stiffness of the object. The increase of the measured position signal after the contact threshold is due both to the adjustment of the object between the thumb and other fingers, and to the compliance of the robotic hand. Regarding forces, in Fig. 8(f), the signal measured by strain gauges at the hand exoskeleton side (blue line) sums up the contributions of the actuators and the operator's hand. Residual forces felt by the user when moving fingers in non-contact conditions are below 0.5 N, measured at the linear actuator of the hand exoskeleton mechanism. With the given teleoperation parameters, no oscillations were noticed after the contact threshold, although the overall dynamics of the hand's subsystem were kept lower than the arm subsystem, in line with the limited dynamics exhibited by the linear actuators. Similar behaviors were observed for the other fingers.

To enrich tactile information, dedicated voice-coil tactile units, capable of high dynamics and wide bandwidth, were used for rendering contact transients and vibrations. The tactile units were driven in feed-forward from the measured robotic hand force through a high-pass filter (yellow signal in Fig. 8(f)). Since tactile units were grounded at the same finger dorsum, such high dynamic components could not interfere with the hand teleoperation loop involving the whole finger kinematic.

V. FIRST EVALUATION: TELEPRESENCE AND REMOTE SOCIAL INTERACTION EXPERIMENTS

This section reports on experimental evaluations meant to demonstrate the system's capabilities and describe application examples where Sully is used in realistic social contexts. Namely, we take into consideration two larger experiments, concerned with:

- a user study in which we demonstrated the suitability of the robotic system to access the semi-finals;
- the ANA Avatar XPRIZE semi-finals.

The pictures and photo sequences presented in this section are extracted from the video attached to this article. In the following, we will refer to the avatar Sully piloted by the human operator by HOS (Human-Operated Sully), and to the human recipient interacting with Sully by R. The experimental procedures were approved by the Ethical Review Board of Scuola Superiore Sant'Anna (approval number 152021).

A. USER STUDY: THE MAGIC TRICK

The user study mission is composed of an integrated session of tasks that proved the suitability of the Sully avatar to access the semi-finals of the XPRIZE competition, including mobility, auditory, speaking, and fine manipulation.

To this aim, we conceived a magic trick interaction, shown in Fig. 9, in which the HOS is involved in the following sequence of tricks by a magician recipient:

- **Task 1**: Approaching (A)
- **Task 2**: Disappearing ball (D)
- **Task 3**: Guess the card (C)
- **Task 4**: The Necklace (N)
- **Task 5**: Leaving (L)

In the first task, starting from one station, the avatar approaches a meeting room, passing a door (A1) and moving through a constrained environment (A2). Once reached the designed table, the HOS maneuvers its orientation up to reach a frontal position with respect to the magician. At this point, the HOS and R verbally and gesturally greet each other (A3) and introduce themselves by handshaking (A4).

In task D, the magician (R) invites the HOS to verbally choose one ball among two and to confirm coherently the choice by pointing to it (D1). The HOS takes the chosen ball from the magician's hand and then grasps it in the fist (right hand in Figure) (D2), while the remaining ball is held in the magician's fist. At this point, the magician shows that no ball is present in his hand (D3), and invites the HOS to open the fist, proving that actually both balls are in the HOS's hand (D4).

In task C, the magician (R) invites the HOS to choose one card from a deck (C1). The HOS grasps the card so that it can see and memorize the figure (C2). Then the card is replaced in the deck on the hidden side (C3). After shuffling the cards, the magician guesses which was the chosen card from the deck and the HOS confirms (C4).

The last trick is the necklace. The HOS holds a necklace between two fingers (thumb and index) of the left hand (N1). The magician is thus able to insert a pendant on the necklace through a trick (N2). The HOS grasps the pendant with the right hand to extract the necklace from the first hand while keeping the thumb and the index open (N3) and brings the completed necklace to the magician's hand (N4).

Lastly, the HOS verbally thanks the magician, waves hello (L1), handshake (L2), leaves the set (L3), and returns to the starting location (L4).

The entire sequence was performed on two trials with the same participants: in particular, the operator was an expert in teleoperation, while the magician/recipient had no substantial relation to telepresence robotics. The completion times and the success rates are listed in Tab. 3.

The proposed sequence proves the ability of the avatar in terms of vision and auditory capabilities, locomotion, and fine manipulation. While all the tasks were accomplished with no errors, in the first trial, the card and the necklace fell from the avatar's hand during tasks (C2) and (N3), respectively. The fall of the card was due to imperfect coordination in the closure of the index and the thumb which caused the card to rotate and fall. The reader should note that the manipulation of the necklace (N3) is a hard task in terms of manipulation - as confirmed by the highest value of the completion time - due to the thinness of the pendant ring and the length of the lace. Indeed, grasping the ring and extracting the lace with the fingers open requires synchronization between visual and kinematic feedback and also the synchronization between the movements of the two arms and the fingers. Again, the failure of the task was due to ineffective grasping of the pendant.

TABLE 3. User study: completion times and success rates.

Magic Trick		
Task	Correct	Compl. Time (mean) [min:sec]
Approaching	100%	00:53
Dis. Object	100%	00:38
Guess the card	50% (fallen card in trial 1)	00:53
Necklace	50% (fallen necklace in trial 1)	01:27
Leaving	100%	01:27

B. CHALLENGE MISSION

In this section, we report on the challenge missions, inspired by the ANA Avatar XPRIZE semifinal that we attended in March 2022. The scenarios domains are representative of the anticipated areas in which avatars will provide benefits to humans in the coming years. These include cultural exchanges, healthcare activities, and social interactions. Namely, the three scenarios chosen for the semifinal capture aspects of these domains to reflect real-world situations, and they are:

- social, cooperative interaction
- business dial
- culture, travel, teaching activity.

Each scenario is designed to have six distinct tasks that present the opportunity to accomplish or demonstrate the required capabilities listed in Sec. III. Fig. 10 shows an interpreted and synthesized tasks list inspired by the official challenge.

In the first scenario, the HOS helps to finish a partially completed puzzle with the participation of the R, as in Fig. 11(a). The setup is composed of a cleared table with a simple, toddler-type puzzle with images on each piece and wood grips. The recipient is seated on the far side of the table, while the avatar is positioned in front on the opposite side.

The second scenario mimics the final stage of a business deal. The HOS joins the meeting with a business partner acting as the host at their office to celebrate the closing of an important deal. The R is seated at a table with a variety of non-breakable beverage containers (coffee mugs and plastic wine glasses) set up on the table.

The last scenario, Fig. 11(b), is concerned with culture, travel, and teaching activities: the HOS, as a visitor, explores a museum of antiquities and interacts with the museum host, as the recipient. In the setup, the recipient is seated behind a table with representative objects from the museum. The HOS is positioned on the other side of the table to start.

FIGURE 9. Photo sequence and operator view of the five tasks included in the user study. The avatar is involved in the execution of magic tricks to assess the suitability of the system to verbally, gesturally, and physically interact with a human recipient.

Two slots were planned on the schedule to accomplish the contest. Each slot was two hours in duration with one hour for equipment setup and operator training, and one hour for the scored trial. Over the course of the semifinal, the operator and the recipient were meant to interact in discrete spaces, i.e. the operator control room and the avatar scenario room. No communication between the operator and the recipient other than through the avatar was allowed.

C. RESULTS: QUESTIONNAIRE AND EVALUATIONS

A judging panel has been selected from the XPRIZE foundation within a wide range of experts in the technology domains that were expected to be integrated into the avatar competition. During the semifinal, two judges participated in each testing run. Namely, one judge served as the operator, while a second judge served as the recipient at the remote location. Each judge was supported by an assistant in their room.

During testing, the judges evaluated qualitatively the avatar system by filling out a questionnaire based on the following four categories:

1) operator experience, i.e. the operator judge:
   - felt present in the remote space with the recipient
   - was able to clearly see and hear what was happening in the remote space
   - was able to move around to complete the tasks
   - was able to manipulate remote objects effectively and felt able to gesture effectively
   - felt safe, and the avatar easy and comfortable to use
2) recipient experience, i.e. the recipient judge:
   - was able to identify and understand the operator, feeling a sense of shared experience
   - felt safe while the avatar was navigating the environment and manipulating objects
   - felt that the avatar's aesthetics were adequate
3) avatar ability to complete each task (pass/fail)
4) overall system:
   - the operator judge felt that the avatar operated reliably in terms of hardware and software (power source, network stability etc.)

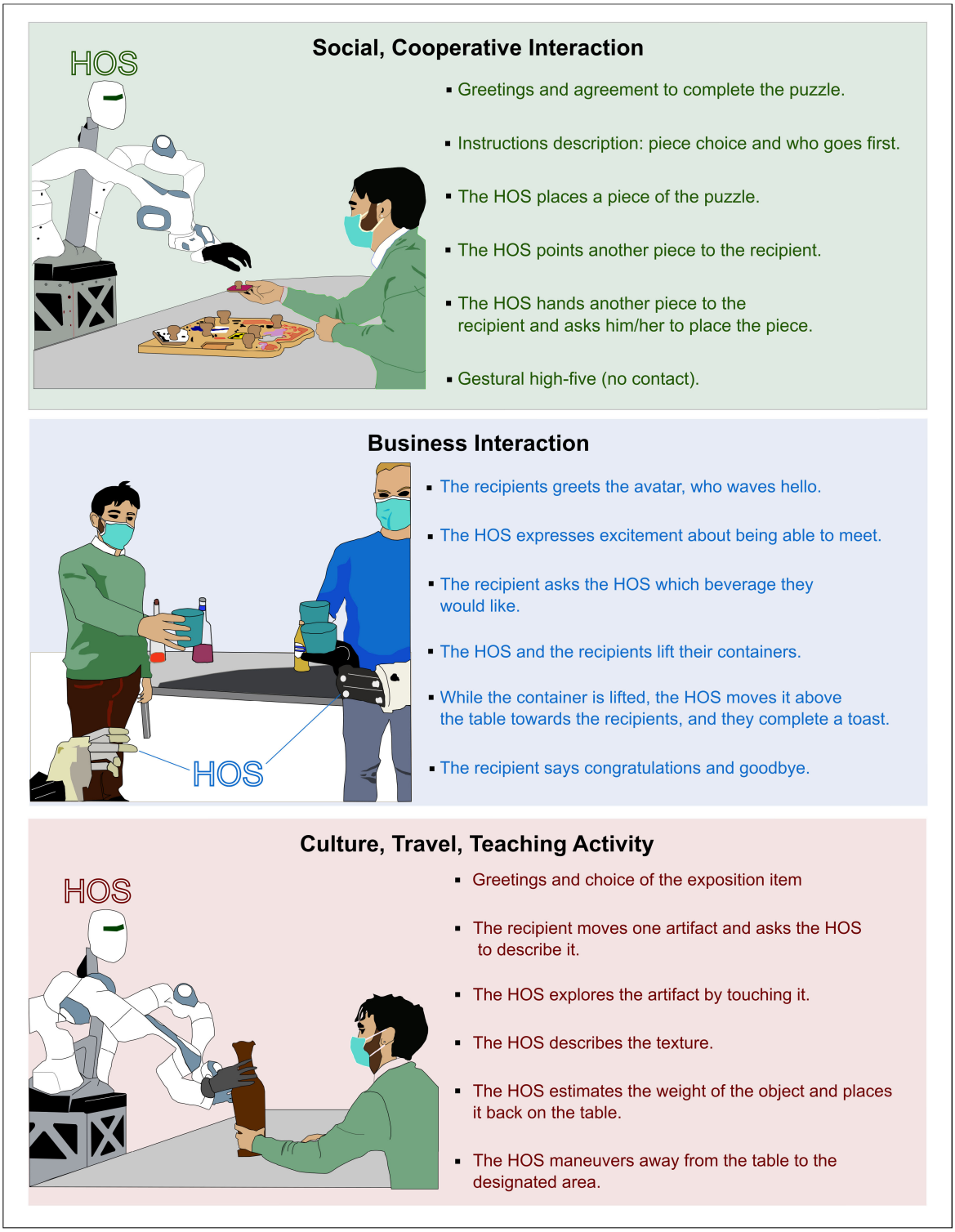

FIGURE 10. **Illustrative description inspired to the XPRIZE semifinal tasks showing the interaction of the avatar system (Human-Operating Sully - HOS) with human recipients.**

- the recipient judge felt that the avatar was physically stable.

Also, a quantitative evaluation was attributed: each scenario was given a maximum of 30 points, while a maximum of 10 points were given to the user study video. During the challenge trials, no multimedia capture was allowed and besides, the judges scoring forms were kept secret, except for the final mark (91/100).

Hence, after the challenge, we replicated the three scenarios to derive further considerations about the system's

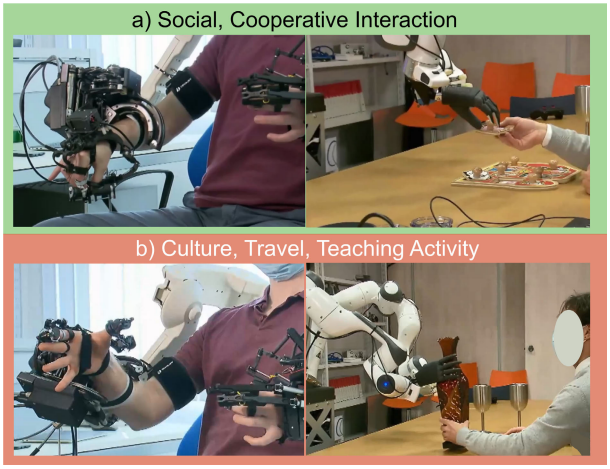

FIGURE 11. Detail of the operator (left) and avatar (right) operation during some fine manipulation tasks inspired by the ANA Avatar X-PRIZE semifinal challenge. (a) The avatar interacts with the human recipient in the completion of a puzzle as an example of a social task. (b) The avatar allows the operator to explore a museum of antiquities integrating visual, manual, and tactile perception of objects.

performance and to take the illustrative pictures and videos shown in the following. Five repetitions of the challenge trials were performed internally, involving five people as operators (males, age: 32 - 42): two were experts in teleoperation, one with partial experience in teleoperation, and two with no experience. Five people were involved as recipients (4 males, and 1 female, age: 26 - 35), all roboticists. In this way, we could compare the detailed questionnaires of these persons with the synthetic judges's evaluation. In the following, we report on the qualitative considerations that emerged from both the operator and the recipient parties.

1) OPERATOR

On average, the operator felt a sense of presence in the remote space: he/she was able to understand the recipient's emotions throughout all scenarios. On the other side, the operator was able to express his/her own emotions even though some difficulties were encountered in the business interaction scenario since no gestures were allowed.

The system's ability was judged excellent in terms of all the perception and feedback capabilities: the operator could receive the necessary tactile, haptic, and force feedback to manipulate and control the scenario objects. In this regard, the puzzle scenario is the most significant, due to the small dimensions of the piece and the required precision to insert the piece in the slot. Success in this task was enhanced through the force feedback capability: all the persons who interpreted the operator stated that feeling the contact threshold between the edge of the piece and the puzzle was a substantial help in completing the task. In the same fashion, a description of the antiquity surface in the last scenario was allowed by vibrotactile feedback on the fingertips that targeted the reliefs. Besides, the operator judged excellent the hearing and visual feedback, being able to accurately sense his/her position and movements.

Safety and easiness were considered quite high, while comfort level was average. Generally, the overall system was considered reliable enough to succeed in all the tasks in terms of hardware, power, software, and network.

2) RECIPIENT

The recipient felt a sufficient emotional connection and sense of shared experience with the operator, due to the absence of a direct visual connection to the operator.

On the other side, the human factors of the avatar were satisfying: the recipient felt safe during the interaction with the robot, whose aesthetic appearance was not threatening, and he/she felt safe during avatar navigation. The ability to understand gestures was average. Also for the recipients, the ability of the avatar was excellent, since the HOS was able to greet, verbally state, and coherently point and describe objects.

As an overall system, the avatar remained safe and stable when not actively controlled and completed the scenarios without needing to be repositioned.

Nonetheless, some recipients reported that the scenario was mostly silent and that sometimes the time lag on the audio signals made speaking/conversation halted. Some recipients perceived that the noise from the left hand augmented the sensation of interacting with a robot more than a human. In general, the system performed successfully the tasks but offered limited social interaction.

D. DISCUSSION

Based on the ANA Avatar XPRIZE experience, we can state that coexistence with avatars in the near future is not a mere pioneering vision. Indeed most of the presented telerobotics technologies revealed mature enough to accomplish the social interaction and assistance tasks.

In this regard, vision technologies can be considered an established field with respect to the purpose of the challenge. Like most of the teams, Sully adopted stereo cameras and HMD to achieve an immersive experience with enhanced depth perception. Differences among groups can only be found in the choice of resolution, field-of-view, and frame rate which anyway did not affect performance, since no relevant failures were registered due to visualization concerns, even in the presence of low-contrast backgrounds. Among the finalists, only Team Northeastern opted for monocular cameras, using a large screen on the operator's side to improve depth perception.

Navigation in a constrained environment is necessary for avatars, even if it was not crucial for the semifinal. Sully, like the majority of teams, adopted an omnidirectional platform. Other proposed approaches were legged robots and differential drive. Bipedal locomotion, implemented by iCub, Janus, and SNU/Tocabi in Tab. 1, attempted to resemble the human leg through a 6 DoFs kinematics, but it is prone to fall risk. As an example, the iCub avatar fell during one of the final test trials after an accidental collision. Differential drive represents a compromise between

human aesthetics and stability, but still, it does not allow instantaneous lateral motion, and besides it needs balance stability control. To prevent unintended tilting, during finals, the iBotics avatar was secured through a wire to a gantry frame. For obstacle avoidance, additional cameras were implemented by some teams. TRINA, for example, was equipped with ultrasonic proximity sensors, while Sully was equipped with LiDAR to create a map of the environment. Navigation control was managed through handheld (VR controllers, joysticks) or foot-operated devices. As in the Sully avatar, foot pedals were the most common choice, mainly because of their intuitiveness and since they free the operator's hands. Alternatives were 3d rudder (teams NimbRo and SNU), or sensorized shoes (iCub).

Manipulation was a primary concern for the challenge. Two main approaches for robot arms were chosen: commercial manipulators or custom robotic arms. The advantage of commercially available systems relies on their certified precision, accuracy, and durability. More importantly, *cobots* imply safety features (collision detection, speed, and torque limits) that can significantly mitigate the risk for humans while interacting with the avatars such as in the puzzle scenario. As evident from Tab. 1, the 7 DoFs Franka Emika Panda was the most adopted manipulator, also chosen by the Sully team, in a tilted configuration meant to mimic the human upper body. On the other side, custom robotic arms (AlterEGO, SNU, UNIST) allow shaping the design to better emulate the human-arm appearance and kinematics, but at the expense of engineering burden.

Also, robotic hands play a key role in tele-manipulation. Main hand designs can be divided into parallel jaws and anthropomorphic hands. While the former is in general more robust, precise, and less expensive, anthropomorphic hands ensured greater dexterity, compliance towards different object shapes, and intuitiveness for remote operators. Among the top-scored teams, the anthropomorphic hands differ in DoFs, ranging from three to twenty DoFs of the Shunk hand (NimbRo team). The Prensilia Azzurra and Mia hand adopted by our group can mimic up to 6 finger elementary movements, still providing good performance also in fine manipulation tasks, like mounting the necklace in the user study, or grabbing and placing the puzzle-piece knob during the semifinal.

Hence the most crucial differences among avatar systems concerned the haptic and force feedback rendering capabilities which also revealed essential for accessing the semifinal stages of the competition.

To display grasp forces many teams adopted brake-type actuators, endorsed in commercial haptic gloves, and able to render only unidirectional grasping forces. Conversely, Sully and Northeastern avatars among some others utilized custom haptic interfaces, able to provide bi-directional static cues. In particular, our 5 DoFs HandExos achieved adequate dynamic response during transient contacts, as quantitatively shown in the experimental validation of Sec. IV, and also realistic rendering as documented by operators during manipulation tasks. To render oscillatory forces vibrotactile transducers (eccentric motors, electromagnetic actuators) were used for texture identification. In our design, linear voice coil actuators were implemented on the index and thumb. Alternatively, the Northeastern team added microphones on the operator's wrist to enrich the sense of texture with audio cues, while TRINA mounted a LiDAR camera on the right gripper to scan surfaces and determine their texture.

Differently from other capabilities, force feedback was implemented by a few teams due to the design complexity of the associated devices. Actually, all the teams that rendered force feedback obtained better performance in the manipulation scenarios and in general in the overall score. As an example, among the final five teams, three of them displayed force to the operator's wrist. While NimbRo and iBotics teams utilized commercial manipulators - the Panda and the Virtuose 6D, respectively - team Northeastern and SNU developed custom solutions.

In this regard, the main advantage of our system is concerned with the feedback level. Through the use of ALeX, our custom upper limb exoskeleton, we obtained distributed force feedback on the operator's arm joints that revealed essential to assist the operator during remote manipulation. Lastly, the use of the ALeX exoskeleton revealed effectiveness in providing a sense of interaction, despite using exoskeletons inevitably introducing a fatigue concern. Nonetheless, the task duration was short enough not to excessively burden the operator. A further advantage of the Sully avatar was the use of variable impedance by feedback control. This is realized by introducing a force/torque sensor into the control loop to estimate the true contact forces. Some teams, such as AlterEGO, also included physical variable impedance actuators, even though they did not address the force feedback capability.

All these considerations are reflected in the quantitative comparison, reported in Tab. 4, where we list the manipulation performance of our system and two other finalists for comparison (NimbRo and iBotics, classified first and fifth respectively). Tab. 4 is sectioned into three parts as the three scenarios presented in Sec. V-B. Metrics were extracted from video records, available online at [31] and [32].

It can be seen that the performance of Sully is comparable to other systems. In particular, in the puzzle task, the operator could finely perceive the contact threshold and thus complete the task in almost the same time as other avatars, despite having a reduced number of hands DoFs, compared to the 20 DoFs of the Shunk hand equipped by the NimbRo team. Besides, both iBotics and NimbRo utilized the same strategy: after placing the piece on the edge of the puzzle surface, the puzzle was held with one hand (the avatar hand itself for NimbRo, the judge hand for iBotics), and lastly, the avatar pushed the piece within the slot in horizontal motion with the fingers, so without touching the knob. Instead, the Sully avatar oriented the piece with respect to the slot, trying to insert the piece at the first attempt, and lastly pushed vertically the piece from the knob, up to

TABLE 4. Manipulation performance during the semifinal tasks.

Social, Cooperative Interaction			
	iBotics [31]	NimbRo [32]	Sully
Grabbing the piece [s]	6	8	7
Placing the piece [s]	12	14	16
Completing the puzzle [s]	11	13	12
Handing the piece [s]	5	5	6
N. Fails	4	1	1
Business Deal Interaction			
	iBotics [31]	NimbRo [32]	Sully
Grabbing the glass [s]	9	5	9
Laying the glass [s]	6	7	7
N. Fails	0	1	0
Cultural Activity			
	iBotics [31]	NimbRo [32]	Sully
Grabbing the antiquities [s]	13	2	12
Shaking the antiquities [s]	5	8	6
Laying the antiquities [s]	11	8	12

complete insertion. Failures occurred for all teams in grasping the piece which required trained coordination from the operator.

Instead, the NimbRo avatar outperformed the others in grasping the glass and the antiquities due to the higher manipulability of the hand which endorsed many more DoFs. More importantly, the Sully avatar succeeded in providing a sense of weight to the operator while shaking the antiquities, and also perceiving the texture of the object during surface exploration.

Yet, tasks of the semifinal did not investigate extensively the force interaction capabilities and consequently, the advantage of exoskeletons was comparable to other aspects of social interaction such as emotional rendering, in which our system was less tailored compared to others. In this sense, the head of NimbRo carried a telepresence screen displaying a reconstructed facial video of the operator.

VI. CONCLUSION AND LESSON LEARNED

In this paper, we have presented a complex platform for an avatar system integrating teleoperation and retargeting layers. The system was designed and tested for visual, speaking, auditory, manipulation, haptics, and locomotion capabilities in the context of everyday life interaction and cooperation with humans. At the State-of-the-Art the most arduous challenge is to feed back to the operator the sense of force, and vibrotactile information. To this end, we successfully integrated an upper-arm exoskeleton that allowed the system to accomplish all the fine manipulation tasks with an excellent rendering of physical interaction. Such a fine rendering capability represents the main contribution of our solution to this challenge.

From the experience of our group and the participating teams, it emerges that most of the technologies are mature enough to consider avatar robotics a tangible chance for the next future. Still, the engineering community has to improve several aspects. For example, we have learned that the basic technical skills of a robotic avatar might not be enough to achieve a complete sense of acceptance from human recipients. As an example, certain aspects, like electric noise, that might seem secondary at a design stage, are revealed to generate a sense of artificial interaction instead of the natural humanoid sensation.

Since the impedance of the avatar arms is more than 10 times the impedance of the exoskeleton, the operator can feel inertia oscillation at the point that collision contacts cannot be perceived when the parameters are not adequately tuned. Hence, an accurate choice of the control parameters is crucial for the quality of force feedback. What is more, introducing a visual connection to the operator - through a streaming video - can be more effective than any attempt to shape the robot's aesthetics.

ACKNOWLEDGMENT

Project funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.1 - Call for tender No. 104 published on 2.2.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union - NextGenerationEU – Project Title "AVATAR: Enhanced AI-enabled Avatar Robot for remote telepresence" – CUP J53D23000860006, D53D23001490008 - Grant Assignment Decree No. 960 adopted on June 30, 2023 by the Italian Ministry of University and Research (MUR).

REFERENCES

[1] *ANA Avatar XPRIZE*. Accessed: Feb. 26, 2024. [Online]. Available: https://www.xprize.org/prizes/avatar

[2] S. Behnke, J. A. Adams, and D. Locke, "The $10 million ANA avatar XPRIZE competition: How it advanced immersive telepresence systems," *IEEE Robot. Autom. Mag.*, vol. 30, no. 4, pp. 98–104, Dec. 2023.

[3] E. Ackerman, "Human in the loop: what the avatar XPrize revealed about the future of telepresence robots," *IEEE Spectr.*, vol. 60, no. 5, pp. 38–44, May 2023.

[4] A. D. Dragan, K. C. T. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *Proc. 8th ACM/IEEE Int. Conf. Human-Robot Interact. (HRI)*, Mar. 2013, pp. 301–308.

[5] 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*.

[6] S. Dafarra, "ICub3 avatar system: Enabling remote fully immersive embodiment of humanoid robots," *Sci. Robot.*, vol. 9, no. 86, Jan. 2024, Art. no. eadh3834. Accessed: Feb. 26, 2024. [Online]. Available: https://www.science.org/doi/abs/10.1126/scirobotics.adh3834

[7] R. Cisneros-Limón, "A cybernetic avatar system to embody human telepresence for connectivity, exploration, and skill transfer," *Int. J. Social Robot.*, pp. 1–28, Jan. 2024.

[8] B. Park, J. Jung, J. Sim, S. Kim, J. Ahn, D. Lim, D. Kim, M. Kim, S. Park, and E. Sung, "Team SNU's avatar system for teleoperation using humanoid robot: Ana avatar XPRIZE competition," in *Proc. RSS Workshop Towards Robot Avatars, Perspect. ANA Avatar XPRIZE Competition*, 2022, pp. 1–2. Accessed: Feb. 26, 2024. [Online]. Available: http://dyros.snu.ac.kr/wp-content/uploads/2022/08/RSS-workshop_TEAM_SNU.pdf

[9] S. Park, J. Kim, H. Lee, M. Jo, D. Gong, D. Ju, D. Won, S. Kim, J. Oh, H. Jang, and J. Bae, "A whole-body integrated AVATAR system: Implementation of telepresence with intuitive control and immersive feedback," *IEEE Robot. Autom. Mag.*, early access, Nov. 20, 2023, doi: 10.1109/MRA.2023.3328512.

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

[11] P. Oh, K. Sohn, G. Jang, Y. Jun, and B.-K. Cho, "Technical overview of team drc-hubo@ unlv's approach to the 2015 darpa robotics challenge finals," *J. Field Robot.*, vol. 34, no. 5, pp. 874–896, 2017.

[12] J. M. Marques, J.-C. Peng, P. Naughton, Y. Zhu, J. S. Nam, and K. Hauser, "Commodity telepresence with team AVATRINA's nursebot in the ANA AVATAR xprize finals," in *Proc. 2nd Workshop Toward Robot Avatars IEEE Int. Conf. Robot. Autom. (ICRA)*, London, U.K., Jun. 2023, pp. 1–3. Accessed: Feb. 26, 2024. [Online]. Available: https://www.researchgate.net/publication/370214148_Com-modity_Telepresence_with_the_AvaTRINA_Nursebot_in_the_ANA_Avatar_XPRIZE_Finals

[13] M. Schwarz, C. Lenz, A. Rochow, M. Schreiber, and S. Behnke, "NimbRo avatar: Interactive immersive telepresence with Force-Feedback telemanipulation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 5312–5319.

[14] J. B. F. Van Erp, C. Sallaberry, C. Brekelmans, D. Dresscher, F. Ter Haar, G. Englebienne, J. Van Bruggen, J. De Greeff, L. F. S. Pereira, A. Toet, N. Hoeba, R. Lieftink, S. Falcone, and T. Brug, "What comes after telepresence? Embodiment, social presence and transporting One's functional and social self," in *Proc. IEEE Int. Conf. Syst. Man, Cybern. (SMC)*, Oct. 2022, pp. 2067–2072.

[15] M. Schwarz, T. Rodehutskors, D. Droeschel, M. Beul, M. Schreiber, N. Araslanov, I. Ivanov, C. Lenz, J. Razlaw, S. Schüller, D. Schwarz, A. Topalidou-Kyniazopoulou, and S. Behnke, "NimbRo rescue: Solving disaster-response tasks with the mobile manipulation robot momaro," *J. Field Robot.*, vol. 34, no. 2, pp. 400–425, Mar. 2017.

[16] T. Klamt, "Remote mobile manipulation with the centauro robot: Full-body telepresence and autonomous operator assistance," *J. Field Robot.*, vol. 37, no. 5, pp. 889–919, Aug. 2020.

[17] *Prensilia' Mia Hand*. Accessed: Feb. 26, 2024. [Online]. Available: https://www.mia-hand.com/

[18] M. Controzzi, F. Clemente, D. Barone, A. Ghionzoli, and C. Cipriani, "The SSSA-MyHand: A dexterous lightweight myoelectric hand prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 5, pp. 459–468, May 2017.

[19] *Prensilia' IH2 Azzurra*. Accessed: Feb. 26, 2024. [Online]. Available: https://www.prensilia.com/it/ih2-azzurra/

[20] C. Cipriani, M. Controzzi, and M. C. Carrozza, "The SmartHand transradial prosthesis," *J. NeuroEng. Rehabil.*, vol. 8, no. 1, p. 29, 2011.

[21] N. A. Dodgson, "Variation and extrema of human interpupillary distance," *Proc. SPIE*, vol. 5291, pp. 36–46, Sep. 2004.

[22] M. Labbé and F. Michaud, "RTAB-map as an open-source LiDAR and visual simultaneous localization and mapping library for large-scale and long-term online operation," *J. Field Robot.*, vol. 36, no. 2, pp. 416–446, Mar. 2019.

[23] *ALeX Exoskeleton*. Accessed: Feb. 26, 2024. [Online]. Available: http://www.wearable-robotics.com/kinetek/products/alex/

[24] E. Pirondini, M. Coscia, S. Marcheschi, G. Roas, F. Salsedo, A. Frisoli, M. Bergamasco, and S. Micera, "Evaluation of the effects of the arm light exoskeleton on movement execution and muscle activities: A pilot study on healthy subjects," *J. NeuroEng. Rehabil.*, vol. 13, no. 1, pp. 1–21, Dec. 2016.

[25] M. Mallwitz, N. Will, J. Teiwes, and E. A. Kirchner, "The capio active upper body exoskeleton and its application for teleoperation," in *Proc. 13th Symp. Adv. Space Technol. Robot. Automat. ESA/Estec Symp. Adv. Space Technol. Robot. Autom. (ASTRA)*, May 2015, pp. 1–8.

[26] V. Prasad, R. Stock-Homburg, and J. Peters, "Human-robot handshaking: A review," *Int. J. Social Robot.*, vol. 14, no. 1, pp. 277–293, Jan. 2022.

[27] S. Marcheschi, A. Frisoli, C. A. Avizzano, and M. Bergamasco, "A method for modeling and control complex tendon transmissions in haptic interfaces," in *Proc. IEEE Int. Conf. Robot. Autom.*, Aug. 2005, pp. 1773–1778.

[28] M. Sarac, M. Solazzi, E. Sotgiu, M. Bergamasco, and A. Frisoli, "Design and kinematic optimization of a novel underactuated robotic hand exoskeleton," *Meccanica*, vol. 52, no. 3, pp. 749–761, Feb. 2017.

[29] M. Palagi, G. Santamato, D. Chiaradia, M. Gabardi, S. Marcheschi, M. Solazzi, A. Frisoli, and D. Leonardis, "A mechanical hand-tracking system with tactile feedback designed for telemanipulation," *IEEE Trans. Haptics*, vol. 16, no. 4, pp. 594–601, May 2023.

[30] B. Hannaford and J.-H. Ryu, "Time-domain passivity control of haptic interfaces," *IEEE Trans. Robot. Autom.*, vol. 18, no. 1, pp. 1–10, Jun. 1989.

[31] *IBotics Team Semifinals*. Accessed: Feb. 26, 2024. [Online]. Available: https://www.youtube.com/watch?v=MdUOgW1dWsQ

[32] *Nimbro Team Semifinals*. Accessed: Feb. 26, 2024. [Online]. Available: https://www.youtube.com/watch?v=VeVVt1vxXMw

GIANCARLO SANTAMATO Post Doc Researcher, Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: giancarlo.santamato@santannapisa.it

DANIELE LEONARDIS Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: daniele.leonardis@santannapisa.it

SIMONE MARCHESCHI Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: simone.marcheschi@santannapisa.it

SALVATORE D'AVELLA Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: salvatore.davella@santannapisa.it

TOMMASO BAGNESCHI Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: tommaso.bagneschi@santannapisa.it

CRISTIAN CAMARDELLA Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: cristian.camardella@santannapisa.it

DOMENICO CHIARADIA Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: domenico.chiaradia@santannapisa.it

MASSIMILIANO GABARDI Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: massimiliano.gabardi@santannapisa.it

ANGELA MAZZEO The BioRobotics Institute, Scuola Superiore Sant'Anna. Email: angela.mazzeo@santannapisa.it

MARCELLO PALAGI Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: marcello.palagi@santannapisa.it

FRANCESCO PORCINI Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: francesco.porcini@santannapisa.it

MASSIMILIANO SOLAZZI Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: massimiliano.solazzi@santannapisa.it

LUCA TISENI Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: luca.tiseni@santannapisa.it

PAOLO TRIPICCHIO Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: paolo.tripicchio@santannapisa.it

MARCO CONTROZZI The BioRobotics Institute, Scuola Superiore Sant'Anna. Email: marco.controzzi@santannapisa.it

CLAUDIO LOCONSOLE Faculty of Technological and Innovation Science, Universitas Mercatorum. Email: claudio.loconsole@unimercatorum.it

ANTONIO FRISOLI Institute of Mechanical Intelligence, Scuola Superiore Sant'Anna. Email: antonio.frisoli@santannapisa.it

Open Access funding provided by 'Scuola Superiore "S.Anna" di Studi Universitari e di Perfezionamento' within the CRUI CARE Agreement