Team JANUS Humanoid Avatar: A cybernetic avatar to embody human telepresence

R. Cisneros[1,*], M. Benallegue[1], K. Kaneko[1], H. Kaminaga[1], G. Caron[1], A. Tanguy[2], R. Singh[1], L. Sun[1], A. Dallard[1,2], C. Fournier[2], M. Tsuru[1], C. Yang[1], Y. Osawa[3], G. Lorthioir[1], F. Kanehiro[1] and A. Kheddar[1,2]

***Abstract*— We present the avatar system developed by Team JANUS, which is finalist in the ANA Avatar XPrize competition. We briefly describe our avatar robot, the operator interface system, and the robot operation system.**

I. Introduction

We recently experienced a very unusual situation worldwide. Our lifestyle has changed dramatically since the start of the COVID-19 pandemic. The latter impacted lifestyles and habits of most daily tasks that were usually done in physical presence. Before, experts were easily able to travel to a very far remote location where their specific knowhow is needed. The ordinary population was able to visit their loved ones at will. Now, the situation is no longer the same. Medical doctors, engineers, researchers... are now in need to find an alternative to deal more efficiently with such situations that are highly likely to show-up again 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. In March 2018, XPRIZE foundation launched a new challenge: the ANA Avatar XPRIZE. The purpose of the competition is to develop an avatar system to deploy senses, actions and presence to a remote location in real time.

Team Janus is among the 20 finalists of the ANA Avatar XPRIZE competition. It is a bi-located team that gathers expertise from the National Institute of Advanced Industrial Science and Technology (AIST), Japan and the French National Centre for Scientific Research (CNRS), France. Particularly, it is formed by members of the CNRS-AIST Joint Robotics Laboratory (JRL) in Tsukuba, and the Interactive Digital Human group at LIRMM in Montpellier.

II. Our vision

In Roman mythology, Janus represents the transition from the past to the future, like the one we are seeing today with new technologies. It also represents bridges and connections, like the ones we are building between humans and robots: two entities into the same *body*.

We are using a humanoid avatar. Among all the existing geminoids, HRP-4CR is the only one with a close-to-human look that can walk, manipulate objects, and realize facial expressions. Therefore, we didn't want to sacrifice the bipedal challenge nor its shape. Instead, we wanted to demonstrate a humanoid avatar that (i) can be easily controllable, (ii) can provide rich sensory feedback, and (iii) can be an elegant solution in many applications. See Figure 1.

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

III. Avatar robot

HRP-4CR is a 38-dof geminoid of 160 cm height, weighting 47 kg. It is the updated version of HRP-4C [1]. HRP-4C had 6 dof arms, no force/torque sensors at the wrists and the hands were not dexterous. This is because it was originally designed for entertainment to imitate human-like motion.

As HRP-4C is already more than ten years old, some of its components became deprecated and the wires became unreliable. Therefore, to develop HRP-4CR we changed the wiring, the electronics, the computing system, the low-level field-bus technology into EtherCAT and the cooling system. Additionally, we improved the reachability of the humanoid robot – by adding one dof at each arm (to have 7 dof arms), as well as the payload capability. We also mounted force/torque sensors at the wrists, speakers, and a helmet featuring a stereo camera (Zed Mini) and a microphone.

The head of HRP-4CR is mounted on a 3-dof neck, allowing for head motion close to that of a human. HRP-4C originally had 8 dof to realize facial expressions; however, as we mounted a small PC with GPU inside (Jetson NANO), it took the place of 4 motor drivers. As a consequence, HRP-4CR can only control 4 face dof: the eyelids, the horizontal and vertical motion of both eyes, and the mouth aperture.

To realize dexterous grasping, our partner DOUBLE R&D[1] developed underactuated dexterous hands capable of power and precision grasps. These hands feature a 1-dof actuated flexion-extension of fingers and thumb, allowing the latter to adapt to the shape of the grasped objects.

[1] CNRS-AIST JRL (Joint Robotics Laboratory), IRL, AIST, Japan.
[2] CNRS-UM, LIRMM, France.
[3] Industrial Cyber-Physical Systems Research Center, AIST, Japan.
* Corresponding author E-mail: rafael.cisneros@aist.go.jp

[1] https://j-d.co.jp

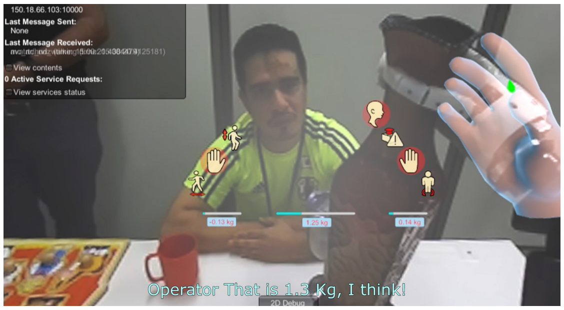

Fig. 2. HUD showing tracked end-effectors and detected weight

IV. Operator interface system

The human interface system consists of a PC running Unity connecting an HTC Vive ProEye head-mount display (HMD) with facial tracker, two Valve Index controllers in the hands, and three Vive trackers: two on the forearms and one on the torso. Each of these elements allows to track in real-time the position and the orientation of the following body parts respectively: head, wrists, elbows and chest. The hands tracking allows an easy completion of object grasping tasks whereas the elbow tracking helps to maintain a natural posture between the operator and the robot during teleoperation. The chest tracking is used as the reference frame to guarantee posture integrity between the robot and the operator if the latter moves during the teleoperation. The head tracking can be activated/deactivated with a button of the controller. The hand tracking can be activated with variable stiffness depending on how hard the index triggers are pressed. In that way, the operator can "rest" and move freely without unintentionally operating the robot. The system in Unity sends these postures as well as the controller commands to the robot operating system.

To transmit the facial expressions of the operator, we use the eye-motion tracking system embedded into the HMD and the facial tracker. As the eyes of the robot cannot move independently (same pitch and yaw for both), the gaze direction is calculated for a "virtual" eye placed between the operator's eyes, and the resulting angles are sent. As for the mouth, the facial tracker detects the lip motion, from which the aperture of the mouth is extracted and also sent.

The operator can also receive the stereo images from the robot's stereo camera and hear the sounds from the robot's microphone by using the HMD and its speaker. The HMD also integrates a Heads-Up-Display (HUD) that allows the operator to visualize which end-effectors are currently being tracked, as well as if the robot is walking or not. Additionally, the HUD displays the weight of the objects as measured by the wrist force/torque sensors, and the trackpad of the Valve Index controllers transmits vibration whose amplitude is proportional to that weight. See Figure 2.

The communication between Unity and the robot operating system is made using ROS# (provided by Siemens) which can establish a bridge between ROS and the Unity environment by using WebSocket communication.

V. Robot operating system

The robot is controlled using the open-source `mc_rtc` framework[2] co-developed by CNRS and AIST. The controller consists of a task-space quadratic programming (QP) controller that computes joint commands that realize to the robot capabilities a set of tasks under constraints (self-collision avoidance, joint torque and kinematic limits, etc).

The QP tasks and constraints are managed by a finite state machine (FSM) that receives inputs from the operator side (the information sent by Unity) and accordingly executes the appropriate states to achieve the desired behavior. For example, when the operator moves the analog thumb joystick, the FSM switches to a walking mode that computes a feasible footstep plan and corresponding dynamic trajectory of the robot's CoM, and executes the walking. Multiple states can be executed in parallel, allowing complex behaviors. For instance, when the operator enables the tracking of head and hands, end-effector task targets are computed, therefore allowing the operator to control them, even while walking.

To realize facial expressions, we use the 4 dof of the face. The eyelids are programmed to blink at random intervals automatically, whereas the desired angles for the corresponding joints of the eyes and the mouth are received from Unity.

VI. Safe interaction

During the semifinals, the robot was required to interact with objects on a table. In the lack of haptic feedback, an operator controlling the robot hand pose could easily overforce on contacts, which could lead the robot to losing balance. Supposing that the robot hand end-effector can only hit the table from above, we derived an admittance control scheme to constraint the effort of the robot hand end effector in one direction such as: in the case of a potential collision with the table, the admittance control will limit the effort applied on the table whereas if the robot's hand is lifting an object, the effort on the hand would not be regulated.

VII. Experimental results

The accompanying video shows a compilation of short clips that demonstrate the operation of our system, as described in this short paper. These clips correspond to the semifinals video, as well as further demonstrations showing some of the capabilities required for the finals.

VIII. Future work

We are currently working on making the robot tetherless, by embedding appropriate batteries and improving our wireless communication. Additionally, we are integrating haptic feedback and devising a suitable haptic interface for the operator. Concurrently, we are improving the locomotion, the retargeting and the visual feedback.

References

[1] K. Kaneko et al. Cybernetic Human HRP-4C. In *IEEE-RAS Humanoids*, 2009.

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