Grant agreement No: 101017008

Harmony

Assistive robots for healthcare

Enhancing Healthcare with Assistive Robotic Mobile Manipulation

(HARMONY) | H2020-ICT-2018-20| RIA

Start of the project: 01.01.2021
Duration: 42 months

Deliverable Number	D24
Deliverable Name	Immersive multimodal interface for providing user grasping demonstrations
WP Number	6
Lead Beneficiary	UEDIN
Dissemination Level	Public
Internal Reviewer	CREATE
Due Date	30.06.2022
Date of Submission	30.06.2022
Version	1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017008

Revision History

Version	Date	Author(s)	Comments
0.1	15/06/2022	Mohammadreza Kasaei, Ruoshi Wen, Joao Pousa de Moura, Zhibin Li	First draft
1	24/06/2022	Mohammadreza Kasaei, Ruoshi Wen, Joao Pousa de Moura, Zhibin Li, Mohammad Hossein Hamedani, Fanny Ficuciello	Internal review and revision

Table of Contents

Summary

This document is focused on development of an immersive multimodal interface which is fundamental for in-situ demonstration of tasks. This interface will provide an unprecedented, immersive capability by integrating multimodal sensors, e.g. haptic, force/torque sensors such that the operator can sense the form, stiffness, and interaction forces in addition to seeing visual clues. Indeed, we start by introducing a first of its kind 3D human performance metric in full 3D space for pointing and manipulation tasks with combined translational and rotational movements. Afterward, we will introduce an immersive multimodal interface that has been developed to gather users' grasping demonstrations. This interface integrates a variety of sensors to provide operators with visual cues and the sense of touch of remote side for in-situ task-level demonstrations to realize the safe transfer of human's demonstrations to robotic arms. Furthermore, we developed a ROS-PyBullet interface which allows us to test our shared control methods using haptic devices in a simulation environment as well as on a real robot.

Introduction

Robotic manipulation is a challenging field of research, particularly when the robot is moving in new scenes that have never been seen before. Additionally, existing methods cannot be verified completely in accordance with the well-established safety criteria needed in clinical and medical procedures. Through the use of an immersive multimodal interface, we can overcome these difficulties by learning from the safety-guaranteed and clinically permitted demonstrations. This method offers medical professionals and staff a natural and intuitive way to apply, practice, or even introduce new manipulation skills at aided, semi-autonomous, and autonomous levels.

An immersive multimodal interface is essential to provide operators with visual cues and the sense of touch of remote side for in-situ task-level. It will provide an unprecedented, immersive capability by integrating multimodal sensors, e.g. haptic, force/torque sensors, and visual inputs, such that the operator can observe visual cues, and sense the shape, stiffness, and interaction forces all through the robot sensing suite. All these input-output data will then produce large sensorimotor datasets that encode and represent the human motor control policies.

Assessing the performance of human movements is essential to acquire valid human demonstrations for robots to learn grasping and manipulation skills. Therefore, we designed and setup a Virtual Reality (VR) framework strategically, which allows us to rapidly iterate design and evaluation metrics, and support in-depth investigation on the performance evaluation of human's demonstrations. This has successfully facilitated our novel proposal of a new human performance metric, as the first of its kind in full 3D space for pointing and manipulation tasks with combined translational and rotational movements. To fulfill the requirement of this task, we have set up haptic devices to collect human demonstration dataset. For more effective demonstration, we have designed a force-motion adaptive controller for grasping stabilization, which reduces the average demonstration time.

Besides, we developed a ROS-PyBullet interface which allows us to test our shared control methods using haptic devices in a virtual environment first, and once fully debugged and tested we can easily swap them to the physical hardware. In the reset of this documents, the detail of proposed human performance metric, immersive multimodal interface and our ROS-PyBullet interface will be presented.

Metrics for 3D Object Pointing and Manipulation

System Overview and Apparatus

Figure 1 illustrates the system overview of our Virtual Reality (VR) framework, including the simulation engine Unity3D, all necessary hardware and software. Two input sensors were used to interact with the simulation. The first sensor was the optical hand tracking device – Leap Motion Hand Controller (LMHC) which is equipped with two infrared cameras at 120Hz and with a $135°$ Field of View (FoV), allowing the interface between the user's hand movements and the physics engine PhysX in Unity3D.

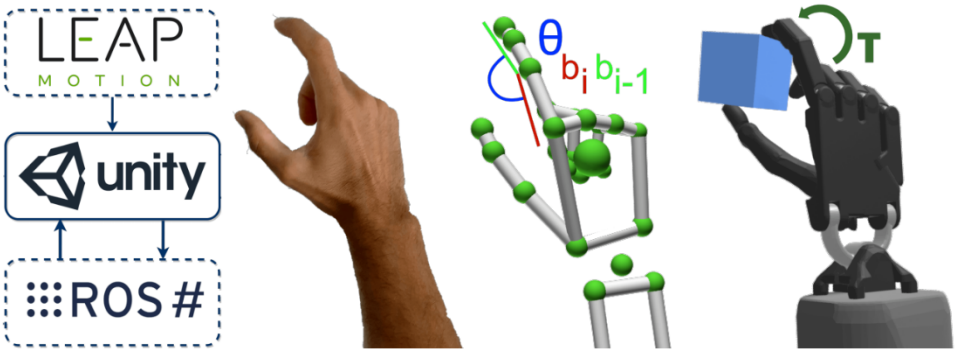


Figure 1: Physics simulation environment and hand setup: the illustration of the user's hand, the LMHC visualization of the joint vectors and the motion retargeting of the robotic hand in Unity3D.

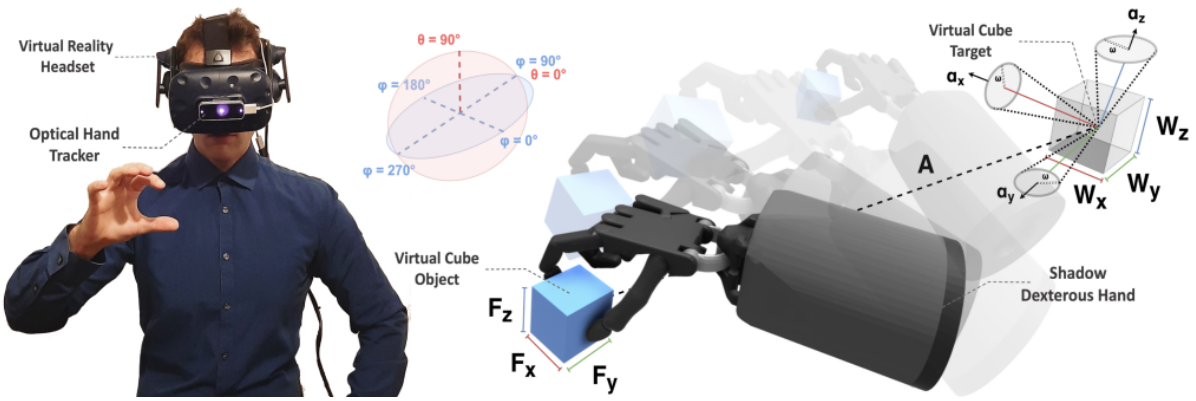
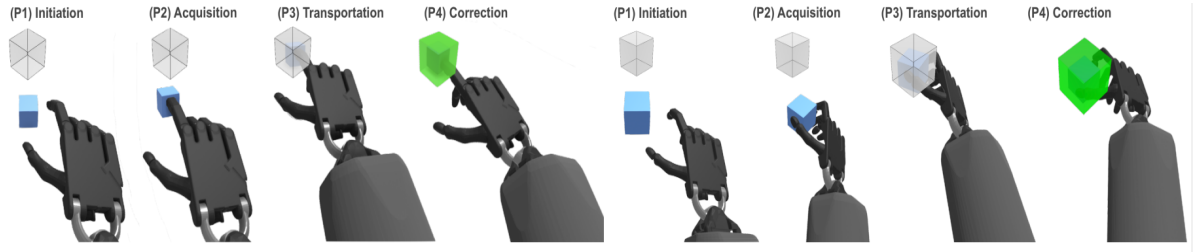


Figure 2: An operator interacts with objects in full 3D VR with all task-related spatial variables.

For visual feedback to the user, high-resolution displays were used to limit distance overestimation and degraded longitudinal control, a known issue in VEs [1].



(a) Pointing tasks (b) Manipulation tasks

Figure 3: Pointing and manipulation tasks: broken down into the four basic phases of interacting with an object. The only difference between (a) and (b) lies in phase P2, where the object is either grasped or attached to the hand depending on the type of interaction.

Consequently, the Virtual Reality Head-Mounted Display (VRHMD) HTC Vive Pro was used, with a 2880 x 1600 pixel resolution display and $110°$ FoV at 90Hz.

The photosensors on the VRHMD also represented our second sensor, allowing for head rotations in the VE. Furthermore, ROS was used to import physics models of robots and objects (Unified Robot Description Format). For all experiments, the LMHC was fixed on the front of the VRHMD. The physics simulation time-step was set at 1000Hz to ensure robust and stable performance with realistic forces and frictions. Finally, to ensure optimal hand tracking performance, lightning conditions were consistent and operational space was limited to about 100cm as the upper maximum reaching bounds from the chest of users. In addition, a low-pass filter with a cutoff frequency of 10Hz was applied to the LMHC to reduce noise during retargeting, ensure continuity and robustness.

Metrics for Assessing Human Performance

To propose a higher dimensional metric for assessing human performance, we investigate Paul Fitts' original predictive model, short for Fitts' law [2], [3]. The original formulation has been extensively used in human-computer interaction (HCI) and ergonomics research and still represents the gold standard as a performance metric[4]. We investigated the most widely used extensions of Fitt's law through four designed experiments in a simulated teleoperation VR setting which is shown in Figure 2. This allowed us to evaluate the applicability of each model with various spatial complexities entailing translational and rotational movements.

A total number of 20 participants (N = 20) were recruited in this study (4 females and 16 males), with ages ranging from 19 to 46 ($\mu = 27.35$, $\sigma = 5.43$). The selection criteria we set during recruitment were that each participant was (i) right-handed, (ii) had healthy hand control with (iii) normal/corrected vision and (iv) was familiar with either video games or VR. Participants that did not meet all of these criteria were excluded from the experiment. Participants were asked to find a balance between minimizing errors and selecting the targets as quickly as they could during target selection and placing.

In these experiments, participants were asked to move a cube from a start to a target location. The use of a cube allowed us to assess rotational variations in our experiment. Due to its identifiable orientation and as one of the most basic 3D shapes, the cube presented a suitable choice to assess both translational and rotational tasks. Regarding rotation, we instructed participants to match the sides of the cube with that of the cube target, as parallel as possible. While using a cube introduces in essence four “correct” rotations and limits to some extent the range of rotations one can investigate (e.g. to a maximum of 45 degrees), it still represents the dominant and most widely used 3D shape in current work [5]. Our approach included rotational tasks, but was limited to 2D movements only following straight lines without directions and inclinations. The targets were arranged in spherical coordinates with the object at the center. Each experiment included pointing and manipulation tasks which are the most widely used types of interactions in collaborative VEs, VR simulators and robot teleoperation [1], [5], [6]. These tasks are shown in Figure 3.

This study showed that in the most basic form of 3D object target pointing and manipulation along a one-directional line only. However, when complexity increased by including directional and inclination angles, these models had reduced performance at predicting the results of our experiments, which dropped slightly in pointing tasks but significantly more in manipulation tasks. Though Fitts’ law has been extended towards rotation, studies so far have limited their findings in 2D space [7], [8], [9]. Furthermore, combining rotational and translational movements under one setting in 3D remains largely unexplored. While Kulik et al. [8] and Stoelen & Akin [9] studied combined movements, these were still limited to 2D space and only following movements across one line.

These essential investigations led us to our novel proposal of a new metric to overcome these limitations and the derived model outperformed other extensions. We present a first of its kind 3D human performance metric based on Fitts’ law, which

extends beyond current work by modelling full 3D space better than existing formulations.

More specifically, our metric is able to capture 3D human motion entailing combined translational and rotational movements, with varying degrees of directions and inclinations in both object pointing and manipulation under a single formulation. This metric can be used to assess human performance by modelling the complex motions, Degrees of Freedoms (DoFs), and dimensions associated with VEs entailing Virtual Reality (VR) [1], [5] as well as teleoperation [2], [6], [10]. Consequently, the effects of different user interfaces, devices, and robotic systems on user performance can be modelled and assessed with such a metric with the added advantage of also combining time and spatial based metrics under one model. The applicability of the proposed metric has been validated across all the experimental results, showing improved modelling and representation of a human performance in combined movements of the 3D object pointing and manipulation tasks than existing work.

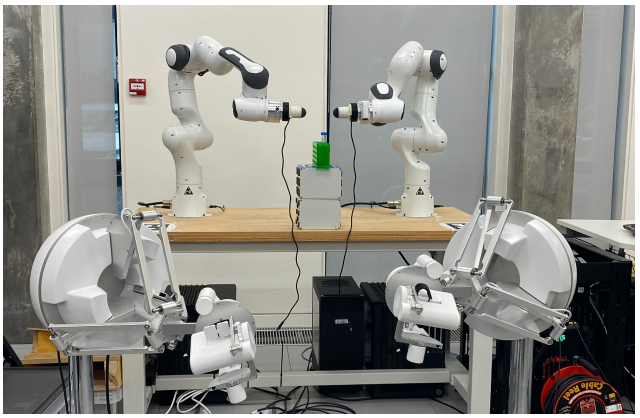
Immersive multimodal interface for demonstrating in-situ operations

We have set up the hardware consisting of two Sigma. 7 haptic devices, two Franka Emika arms and the vision system with three cameras for bimanual teleoperation to collect a human demonstration dataset. To realize the safe transfer of human's demonstrations to robotic arms, we investigated the performance of a multi-contact motion retargeting method. For more effective demonstration, we have designed a force-motion adaptive controller for grasping stabilization, which reduces the average demonstration time to 7 seconds. The formulation of a grasping controller that adjusts the relative poses between two end-effectors was implemented to gather valid demonstration data from different end-effectors' poses. The robustness and safe motion transfer performance of the teleoperation-based multimodal demonstration interface has been validated through a few repetitive experiments where human operators succeeded in completing the pick & place task of test tubes. Using the bimanual teleoperation setup in Figure 4, we have collected human demonstration data in the task of picking and placing test tubes.

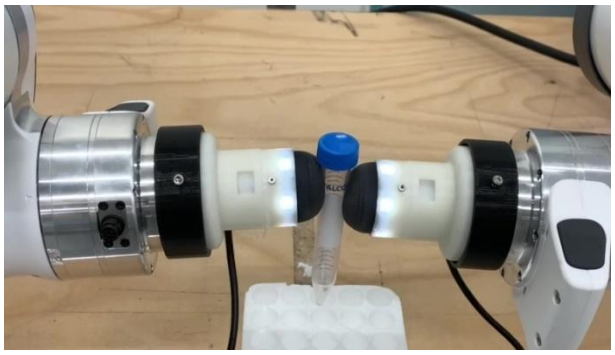
ROS-PyBullet interface for reliable contact simulation

The co-piloting between the human operator and the robot will be mapped at the end-effector level, i.e., map human hands to robot hands. The resulting joint motions of the robot arms will be resolved by the whole-body planning and control module by the outcomes of Work Package 7 (WP7), which enhances collision-free requirements

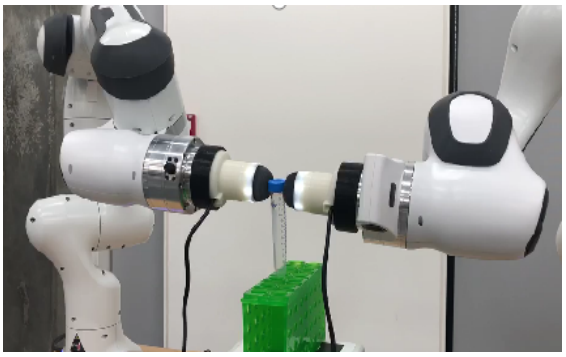
for safety. To address this point, we propose solutions for manipulating objects without grasping. This non-prehensile manipulation by pushing and sliding objects entails the robot having contact-rich interactions with the object and/or the environment. Developing methods for dealing with contact-rich manipulation tasks is inherently complex due to the uncertainty and unpredictability of the task.



(SEQ (a) * alphabetic a) Overview of the tele-operation setup consisting of two Sigma. 7 haptic devices, two Franka Emika arms and the vision system with three cameras for bimanual teleoperation.



(b) Grasping test tubes via human demonstrations.



(c) Collecting the real human demonstration.

Figure 4: Collecting the real human demonstration data in the task of picking and placing test tubes using the bimanual teleoperation setup via two Sigma 7 haptic devices.

From a software development point of view, dealing with the uncertainty introduced by contacts requires careful testing of the software. Typically, robotics manipulation with contacts, such as teleoperation with haptics, tends to be developed and tested directly on the robot, which is significantly more time consuming than developing and testing using a simulator.

On the other hand, in recent years, the robot learning community has been relying on a set of reliable simulation libraries for contacts, such as PyBullet [11] and MujoCo [12] to train their policies. Besides all the questions related to the Sim2Real gap, there is also the additional problem of translating the code developed using physical

simulators to communicate with the physical hardware, through for instance ROS. Note that although ROS integrates well with Gazebo, this is known to be less reliable for contact-rich tasks.

In this work, we developed a ROS-PyBullet interface to address this problem. In this way, we can develop all our code base using the familiar ROS environment with access to a multitude of existing libraries for communication, signal processing, estimation, control, etc., and test that code with PyBullet. Once confident in the reliability of the methods being developed/tested, we can easily swap the virtual environment by the physical setup.

Figure 5 shows a typical experimental setup where we control a dual-arm robot using our Model Predictive Controller (MPC) method for non-prehensile manipulation, use Inverse Kinematics (IK) algorithm for getting the robot joint angles, use Vicon to track the object pose, and use a force sensor to obtain the contact forces. We use ROS to manage all those processes. We can also now stream all that information to PyBullet, for enhanced visualization of the task being executed.

Figure 6 shows a different scenario where we want to test our shared control framework in a physical simulation before attempting that on physical hardware. Through our interface, we can now read the forces being simulated from PyBullet and stream them through ROS, so that the operator can have force feedback from the task without having to be testing it on the actual hardware. Therefore, this interface improves the development of our control methods both in terms of integration time and safety.

Figure 5: Robot pushing a box using a Model Predictive Controller (MPC) for non-prehensile manipulation. We track the box pose using the Vicon system which streams the pose through ROS and displays that information in PyBullet alongside other useful information such as the target pose.

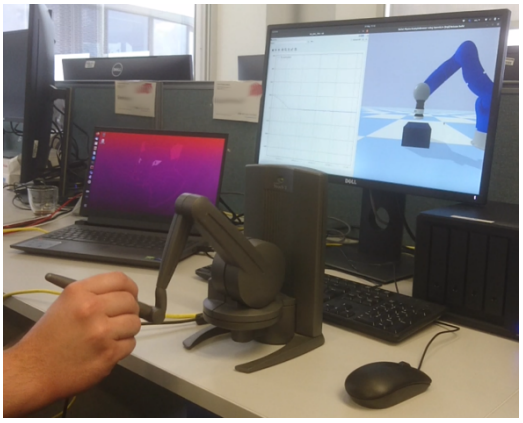


Figure 6: Teleoperation of a virtual robot with haptic feedback being originated by the Pybullet simulator and communicated to the haptic device through ROS.

Conclusions

We have investigated the fundamentals of the metrics to evaluate the manipulation performance, and proposed our novel metrics for 3D tasks, which led to a flagship journal publication and laid a foundation for the task evaluations for future research outcomes. To achieved so, we have investigated the design of an immersive multimodal interface to provide operators with visual cues and the sense of touch of remote side for in-situ task-level. The essential outcomes are concluded as follows.

Virtual Reality Framework. We developed a VR framework including the simulation engine Unity3D along with all required hardware and software. This framework allowed us to perform massive testing to evaluate the performance of human's demonstrations, and to iterate rapidly a wide range of candidate metrics to quantify the human manipulation skills. We designed four experiments in a simulated teleoperation VR setting to evaluate the applicability of existing models with various spatial complexities entailing translational and rotational movements. Also, we recruited 20 participants and asked them to perform these tasks for benchmarking human performance versus robot performance. In all experiments, the participants were asked to move a cube. This investigation found out that instead of conducting the experiments using many different objects, we can employ the same standard cubic object to benchmark, measure, and evaluate the performance in order to standardize the benchmarking protocol. Such a benchmarking protocol is very important to quantify the autonomous robot manipulation in the stage of performance evaluation.

A new human performance metric. We have investigated the performance evaluation of human's demonstrations through our VR framework. We proposed a new human

performance metric (a first of its kind) in full 3D space for pointing and manipulation tasks. More specifically, this metric can capture 3D human motion entailing combined translational and rotational movements, with varying degrees of directions and inclinations in both object pointing and manipulation under a single formulation. This metric forms the guideline to design effective human-robot interfaces.

Immersive multimodal demonstration interface. The presented results are very concrete scientific results that support us in developing the hardware later. According to the findings of our research conducted in a virtual reality and physics simulator environment, we have set up the hardware consisting of two Sigma. 7 haptic devices, two Franka Emika arms and the vision system with three cameras for bimanual teleoperation to collect a human demonstration dataset. The performance of the teleoperation-based multimodal demonstration interface has been validated through a few repetitive experiments in which human operators successfully completed the pick and place task of test tubes.

A new interface between PyBullet simulation and ROS. This interface allows us to develop our methods for contact-rich manipulation tasks, from full autonomous planning and control to shared control with haptic interfaces, directly in the ROS environment and use PyBullet to test their reliability in handling contact interactions. In this way, we can reduce the transition time from testing in a simulation environment to deploying in the physical hardware, because most of the code (developed in ROS) is directly reused.

In addition to the conclusion, furthermore, the outcomes from this deliverable underpin our future work planned in the next stage. Within the deliverable D6.1, the immersive multimodal demonstration interface will support the deliverable from Task 6.3 at the end of year 2: D6.3 Learning from demonstration framework at Month 24. The new interface between PyBullet simulation and ROS are related to the aspect of human-robot interface and can be adapted and utilized to support the development and delivery of D6.4 Human-robot social implications of learning from demonstration at Month 36. Moreover, the communication protocols and software infrastructure stemmed from the new interface between PyBullet simulation and ROS can be naturally synergized with Task 7.4 in work package 7, regarding the manual guidance of contact points and safe reaction in case of unintentional collisions.

References

[1] M. D. Barrera Machuca and W. Stuerzlinger, “The effect of stereo display deficiencies on virtual hand pointing,” in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ser. CHI ’19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3290605.3300437

[2] P. M. Fitts, “The information capacity of the human motor system in controlling the amplitude of movement.” Journal of experimental psychology, vol. 47, no. 6, p. 381, 1954.

[3] P. M. Fitts and J. R. Peterson, “Information capacity of discrete motor responses.” Journal of experimental psychology, vol. 67, no. 2, p. 103, 1964.

[4] I. S. MacKenzie, “Fitts’ law as a research and design tool in human-computer interaction,” Hum.-Comput. Interact., vol. 7, no. 1, p. 91–139, Mar. 1992. [Online]. Available: https://doi.org/10.1207/s15327051hci0701 3

[5] E. Triantafyllidis, W. Hu, C. McGreavy and Z. Li, "Metrics for 3D Object Pointing and Manipulation in Virtual Reality: The Introduction and Validation of a Novel Approach in Measuring Human Performance," in IEEE Robotics & Automation Magazine, vol. 29, no. 1, pp. 76-91, March 2022, doi: 10.1109/MRA.2021.3090070.

[6] J. M. O’Hara, “Telerobotic control of a dextrous manipulator using master and six-dof hand controllers for space assembly and servicing tasks,” Proceedings of the Human Factors Society Annual Meeting, vol. 31, no. 7, pp. 791–795, 1987. [Online]. Available: https://doi.org/10.1177/154193128703100723

[7] G. V. Kondraske, “An angular motion fitt’s law for human performance modeling and prediction,” in Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 1, 1994, pp. 307–308 vol.1.

[8] A. Kulik, A. Kunert, and B. Froehlich, “On motor performance in virtual 3d object manipulation,” IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 5, pp. 2041–2050, 2020.

[9] M. F. Stoelen and D. L. Akin, “Assessment of fitts’ law for quantifying combined rotational and translational movements,” Human Factors, vol. 52, no. 1, pp. 63–77, 2010, pMID: 20653226. [Online]. Available: https://doi.org/10.1177/0018720810366560

[10] S. Fani, S. Ciotti, M. G. Catalano, G. Grioli, A. Tognetti, G. Valenza, A. Ajoudani, and M. Bianchi, “Simplifying telerobotics: Wearability and teleimpedance improves human-robot interactions in teleoperation,” IEEE Robotics Automation Magazine, vol. 25, no. 1, pp. 77–88, 2018.

[11] E. Coumans and Y. Bai. Pybullet, a python module for physics simulation for games, robotics and machine learning. https://pybullet.org, 2016-2020.

[12] E. Todorov, E. Erez, and Y. Tass. Mujoco: A physics engine for model-based control. In IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012.