

# Upper-limb exercise with Physical and Virtual Robots: Visual Sensitivity Affects Task Performance

Pauline Chevalier, Valentina Vasco, Cesco Willemse, Davide De Tommaso, Vadim Tikhanoff, Ugo Pattacini, Agnieszka Wykowska\*

\*Corresponding Author: Agnieszka Wykowska, Social Cognition in Human-Robot Interaction, Istituto Italiano di Tecnologia (IIT), Genoa, 16152, Italy, [agnieszka.wykowska@iit.it](mailto:agnieszka.wykowska@iit.it)

Pauline Chevalier, Cesco Willemse, Davide De Tommaso, Social Cognition in Human-Robot Interaction, Istituto Italiano di Tecnologia (IIT), Genoa, 16152, Italy, [name.lastname@iit.it](mailto:name.lastname@iit.it)

Valentina Vasco, Vadim Tikhanoff, Ugo Pattacini, iCub, Istituto Italiano di Tecnologia (IIT), Genoa, 16163, Italy, [name.lastname@iit.it](mailto:name.lastname@iit.it)

## ABSTRACT

We investigated the influence of visual sensitivity on the performance on an imitation task with the robot R1 in its virtual and physical forms. Virtual and physical embodiments offer different sensory experience to the users. As each individual responds differently to their sensory environment, their sensory sensitivity may play a role in the interaction with a robot. Investigating how sensory sensitivity can influence the interactions appears to be a helpful tool to evaluate and design such interactions. Here, we asked 16 participants to perform an imitation task with a virtual and a physical robot under conditions of full and occluded visibility, and to report the strategy they used to perform this task. We asked them to complete the Sensory Perception Quotient questionnaire. Sensory sensitivity in vision predicted the participants' performance in imitating the robot's upper limb movements. From the self-report questionnaire, we observed that the participants relied more on visual sensory cues to perform the task with the physical robot than the virtual robot. From these results, we propose that a physical embodiment enables the user to invest a lower cognitive effort when performing an imitation task over a virtual embodiment. The results presented here are encouraging that following this line of research is suitable to improve and evaluate the effects of the physical and virtual embodiment of robots, for applications in healthy and clinical settings.

## KEYWORDS

Socially assistive robot, Virtual agent, Sensory sensitivity, Rehabilitation

## 1 INTRODUCTION

Nowadays, the use of social, interactive robots is increasing in our daily life. How to improve the interaction between one or more users and one or several robots is one of the main questions of the human-robot interaction research field. Inter-individual differences in acceptance of the robot as an interaction partner have often been reported [1-5]. One critical dimension along which users differ is their sensory sensitivity. For example, some individuals would suffer from exposure to background noise in an open-space environment or would be affected by bright lights in a shopping mall, whereas others would be less disturbed. Sensory sensitivity is mainly discussed and defined in individuals diagnosed with autism-spectrum disorder (ASD) as a hypo- or hyper-reactivity to sensory stimuli [6]. However, the sensitivity to sensory stimuli can be found in typical population as well [7], although this too was found to be correlated with sub-diagnostic autistic traits [8]. In [8], the authors discussed that, for auditory inputs, in the range of audible frequencies, individuals' "comfort zones" differ and would be narrower for individuals diagnosed with ASD. We believe that sensory sensitivity is critical in human-robot interaction, as sensory information plays a major role in interactions with the environment, including social interactions [9]. The use of vision enables people to observe the interaction partner's expressions and bodily behavior. Hearing is used to analyze emotion in the speaker's voice or to follow instructions. In addition, robots are a complex source of sensory information, as they vary in embodiment types (e.g., humanoid or not, physical or virtual), display mechanical parts or LEDs, and produce noise from their motors, mechanical joints, and loudspeakers.

To our knowledge, only a few studies investigated the role of sensory sensitivity in the acceptance of social robots. In clinical settings, sensory sensitivity in vision and proprioception was investigated as predictor of performance in an interaction with a robot (in imitation [3], emotion recognition [10] and joint attention [11]). In [11], 11 children diagnosed with ASD were asked to perform a joint attention task with the robot Nao (Softbank Robotics). The results showed a relationship between their reaction times, and their proprioception and visual sensitivity. However, these results were obtained during a single interaction with the robot, with few joint attention trials. Similarly, in [10], 19 children and adults diagnosed with ASD were asked to identify four emotions on different artificial agents (two robots, a virtual agent, and a human). A relationship between correct classification of facial expression of emotion, and proprioception and visual sensitivity was found. Finally, in [3], 12 children diagnosed with ASD did an imitation training with the robot Nao. Sensory sensitivity in vision and proprioception were found to predict the improvement in imitation during the sessions. Overall, these studies show that in individuals diagnosed with ASD, visual and proprioception sensitivity affected the success of the interaction (for example, training imitation) and in emotion recognition [10]. However, little is known about the impact of visual sensitivity in general (healthy) population on engagement with a robot. This information is crucial in order to make comparisons with clinical populations. In typically developed adults, the auditory sensitivity was shown to have an impact on the performance in a Stroop task with the Tiago robot [4]. In this single case experiment, the authors manipulated visual and auditory stimuli from the robot (e.g. movements and speech volume) to make it appear neutral, encouraging or stressing. The participants were screened for their sensory sensitivity in audition and vision. Individuals who generally seek auditory sensation or who do not detect changes in auditory information performed better with a stressing and an encouraging robot, compared with the neutral robot. The results of this work demonstrate an impact of the sensory sensitivity of typically individual in a cognitive task done with a robot, and that manipulating the robot's behavior in auditory and visual cues can help improve participants' performance.

In this context, it is important to further understand the impact of users' sensory sensitivity on their interactions with social robots. These considerations will be impactful for both healthy and clinical populations. Indeed, the more attuned and engaging the robot is, the more benefits the interaction might bring [12-14], and sensory sensitivity might have an impact on engagement. In [12], [14], engagement with the robot was found to be beneficial during stroke rehabilitation therapy. The engagement with the robot helped the users perform exercises for longer periods. In long term interactions with a robot, the engagement with the user is often discussed as a limitation and a point to improve in future work, as observed in case of with diagnosed with ASD (see [15]). Therapists, during interviews reported in [13], highlighted the need of engagement during therapies including a social robot. Therefore, maximizing the engagement and interest of the users in the interaction can improve their performance and avoid premature withdrawal from the training program. We believe that studying the sensory sensitivity of users of social robots, especially in socially assistive robotics, could improve engagement and commitment to interaction.

In the framework of a collaboration between the Italian Institute of Technology and the Don Gnocchi Foundation, we tested the use of the R1 robot [16], see Fig. 1, as a training coach for upper-limb rehabilitation in an imitation task with neuro-typical young adults, in view of using the robot with elderly individuals in later phases of the project. We believe that this is a fundamental preliminary step, which provides a baseline for comparison for tests with non-healthy individuals. In a previous paper [17], we investigated the differences in the participants' engagement towards the robot and performance in the task when the robot was presented in its physical and virtual forms. Virtual agents are often used in stroke rehabilitation therapy, and we investigated which type of embodiment was most engaging for participants. As expected from literature (see [18], [19] for reviews on the impact of the robot embodiment types), we observed that the physical form of R1 improved the engagement of our participants. We believe that real and virtual robots offer a different sensory experience to the user. For examples for visual cues, the flickering of the screen in which is displayed the virtual robot may interfere with the task for some sensitive participants, or similarly with bright lights from the LED display of the robots. In this paper, we aimed to observe the influence of sensory sensitivity in vision on our participants' performance in the imitation task, with regards to the different embodiments. With this study we aim to advance the knowledge of how sensory sensitivity can affect the interactions between a human and a robot partner. Investigating the difference between physical and virtual embodiments in regards of sensory sensibility was not done previously to our knowledge.

## **2 METHODS**

### **2.1 Robot and framework**

We used the robot R1 [16], a 1.2-meter-high humanoid capable of extending the torso height up to 1.35 meters, with two 8-degrees of freedom arms (Fig. 1 left). The robot's head is equipped with an Intel Real Sense RGB-D camera for depth sensing. We developed a realistic simulation of the robot in the environment Gazebo (Fig. 1 right). The robot and its simulated version run on YARP [20], an open-source middleware in which a large variety of modules running on different operating systems can be interconnected.

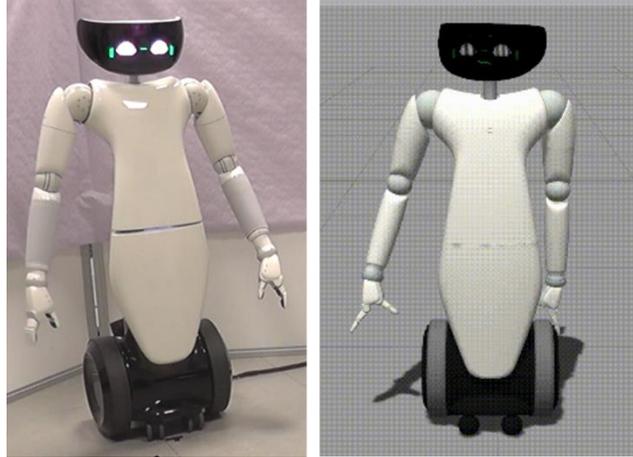


Figure 1. The robot R1 in its physical (left) and virtual (right) embodiments.

The devised framework<sup>1</sup> used in this paper retrieves human skeletons from RGB-D camera to analyze actions and to provide verbal feedback to the user. The framework and its modules are shown in Fig. 2. First, we detect 2D human keypoints on the RGB image provided by the camera, using OpenPose [21], an open-source library for real-time multi-person 2D pose estimation. Keypoints are then reconstructed in 3D fashion, using the depth provided by the camera, with the classical pinhole camera model (i.e. *skeletonRetriever*). The retrieved skeleton undergoes an action recognition layer, which classifies the skeleton data in a pre-defined temporal window, using a Recurrent Neural Network with Long Short-Term Memory cells (i.e. *actionRecognizer*). We trained the network in a supervised fashion offline, with pairs of temporal sequences of fixed length, including 2D joints of the upper body, and the corresponding class labels (abduction and random). At runtime, if the predicted class has the label of the exercise to perform (i.e. *abduction*), we perform a finer analysis, by comparing the observed skeleton with a pre-recorded template, moving coherently with the robot (i.e. *motionAnalyzer*). Observed and template skeletons are aligned spatially, considering the roto-translational offset between the bodies, and temporally, applying the Dynamic Time Warping to the 3D joint positions. We then perform a statistical analysis on the error in position between each component of the candidate and template skeletons and look for positive and negative tails in the error distribution, which reflect to a higher and lower range of motion, respectively. We also evaluate the speed performing the Fourier transform of each component of the joints under analysis: a positive difference in frequency relates to a skeleton moving slower than the template and a negative difference relates to a faster movement than the skeleton. Finally, the participant is associated with a score ranging from “0” (if the movement is completely wrong) to “1” (if no error is detected). Real-Time verbal feedback is synthesized according to the protocol summarized in Table 1. For this paper, the template is a pre-recorded skeleton generated by a person performing the exercise with the robot and is normalized at runtime to deal with differences in the body structure. However, there might be small differences between the template and the robot movement (in terms of amplitude and timing). To avoid such variations in future studies, a new version of the template was developed to now compute at runtime the 3D configuration of the robot limbs directly from the robot’s joints.

---

<sup>1</sup> The code is open-source, available at <https://github.com/robotology/assistive-rehab>

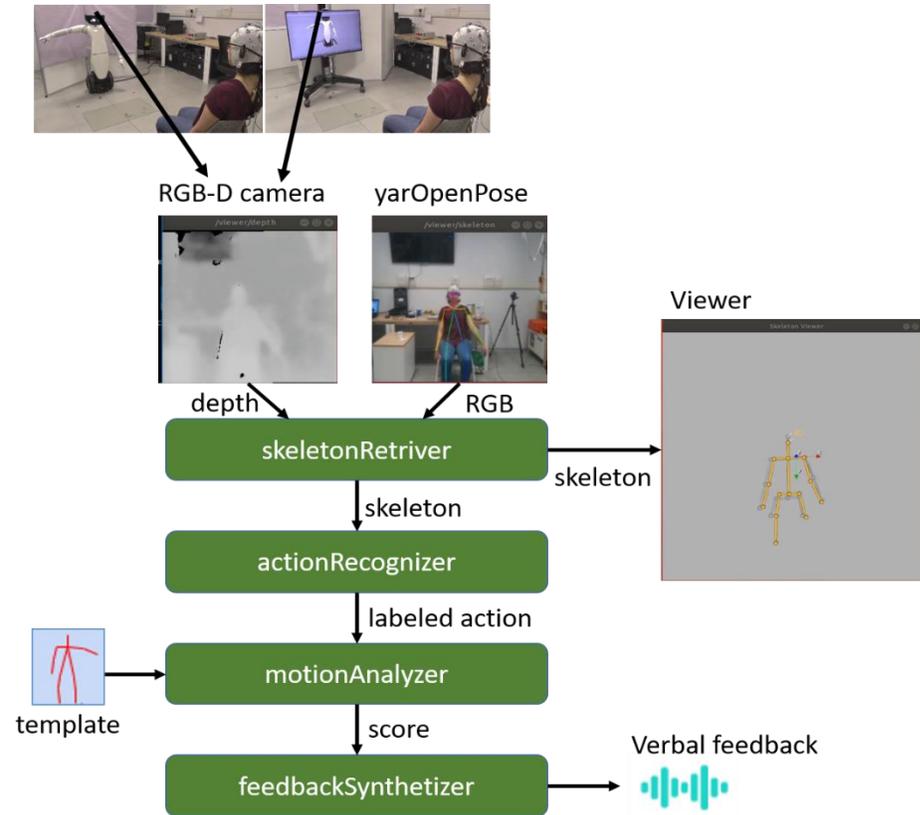


Figure 2. The devised framework

Table 1. Verbal feedback delivered by the robot during the interaction

Detected error	Score	Verbal feedback
Action not recognized	0.0	Please put more effort.
$df > 0$ ( $< 0$ )	0.5	Move the left arm faster (slower).
$\epsilon_x > 0$ ( $< 0$ )	0.5	Move the left arm more on the left (on the right).
$\epsilon_y > 0$ ( $< 0$ )	0.5	Move the left arm further up (further down).
$\epsilon_z > 0$ ( $< 0$ )	0.5	Move the left arm backward (forward).
–	1.0	You are moving very well.

## 2.2 Experimental design

### 2.2.1 Participants

We recruited sixteen participants (age  $M=23.0$ ,  $S.D.=2.81$ ; 7 males) via the web platform “join the science” of the Italian Institute of Technology. We sampled 16 participants due to the limited time of the availability of the robot and the long duration the experiment and installation of the participants (around one hour and a half per participant). Our study adhered to the ethical standards laid down in the Declaration of Helsinki and was approved by the local Ethics Committee (Comitato Etico Regione Liguria). Participants were compensated with 20€ for their participation. Only right-handed participants were recruited for the experiment. Firstly, the participants were asked to imitate a movement with their left arm. Choosing right-handed participants ensure us that dominance differences would not be an issue. Secondly, for the study reported in [17] we also recorded neural activity, for which it is typical to only select right-handed individuals to avoid potential variations in brain-structure and functioning related to left-handedness. Participants did not report any previous familiarity with robots.

### 2.2.2 Design considerations

The imitation task has been chosen as it is common during typical rehabilitation therapy. Robots have been used in upper limb rehabilitation for stroke patients (see [22] for a systematic review), for children with physical impairment in pediatric rehabilitation structures [23] or for children diagnosed with ASD to train body movement imitation [3], [24]. In the typically developed population, the use of robots has been discussed to promote sports activities to prevent insufficient physical activity [25]. In our task, participants performed an imitation task with a physically embodied R1 robot and also with its virtual version. Critically, we introduced two additional conditions for both the physical and the virtual robot: one in which the robot was visible to the participants, and one in which it was occluded. We were interested in the accuracy of continuation of the imitation rhythm in the absence of visual feedback [17]. Patients undergoing such therapy often suffer from sensitivity impairments, for example vision deficiency is common after stroke or in the elderly. It is then relevant to investigate the relationship between sensory sensitivity in the context of rehabilitation with artificial agents. However, before recruiting patients in rehabilitation therapy to participate to this study, we wanted to test our setup with young healthy individuals. Indeed, it enables to verify the feasibility of the setup without recruiting vulnerable population.

### 2.2.2 Procedure

In this study, the participants were asked to do an imitation task with the R1 robot, in its *physical* and *virtual* embodiments. After receiving their informed consent, we gave the participants earplugs to wear to attenuate background and actuator noise and we fitted them with an EEG equipment (EEG data are not the focus of this paper, see [17]). The setup is visualized in Fig. 3. The two types of embodiment were presented in counterbalanced order to the participants, so that half of the participants interacted first with the physical robot then the virtual robot, and the other half interacted first with the virtual robot and then with the physical robot. The physical robot and the screen displaying the virtual robot were placed at the same distance from the participant. Before the experiment, the experimenters informed the participants of the procedure (the text provided to the participant can be found in the Supplementary Material). Participants had to perform abduction movements with the left arm following the robot’s instructions, during three conditions. They first had to observe the robot’s arm performing the abduction movement (*observation*, see Fig. 3 first row). This step acted as a training phase, as when a therapist

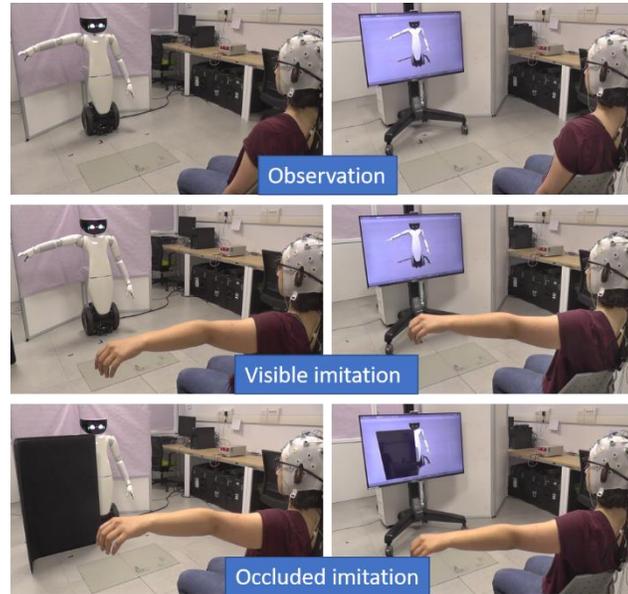


Figure 3. The experimental scenarios during observation, visible imitation and occluded imitation (respectively first, second and third row) for the physical robot and virtual robot (first and second column, respectively).

demonstrate the movement to patients. Then they had to do the movement with the robot (*visible imitation*, see Fig. 3 second row). Finally they had to do the movement with the robot with its arm occluded from the participant's view (*occluded imitation*, see Fig. 3 third row). The occlusion phase was added to evaluate how patients would perform in situations in which perceptual information processing is not complete. Instructions for these three conditions were provided by the experimenters before the experiment commenced. When the participant was interacting with the physical robot, one of the experimenters placed a panel in front of the robot arm at the onset of the occluded imitation condition. With the virtual robot, a virtual panel appeared in front of the robot's arm. For each condition, the participants performed the movement with the robot eight times, and the sequence (i.e. *observation – visible imitation – occluded imitation*) was repeated six times for each embodiment type of the robot. During the visible imitation phase, the verbal feedback (see Table 1) was delivered approximately every three movements and half (around 18 seconds). Participants were allowed breaks between each sequence. When the six sequences were performed with the physical or the virtual robot, the participants were asked to complete a self-report questionnaire. As the virtual robot did not generate any noise, while the physical robot did, and because there was external background noise from the equipment and environment, we made the participants wear earplugs to neutralize most of the noise of the motors of the physical robot. Since auditory sensitivity was not the focus of our study, we excluded the auditory sensitivity from our evaluation. Before starting the experiment, we checked with each participant if wearing the earplugs was not canceling out the vocal feedback of the robot (as described in Section 2). The selected movement (abduction of the left arm) is rather simple, but this movement is at present in use for arm rehabilitation, for example after stroke. Both physical and virtual robot did the movement at an average period of 5.6s. The frequency was the same across all conditions. Trial repetition is fundamental for both the effectiveness of a rehabilitation therapy and the reliability of EEG

measurements. Given the duration of the experiment (around 30 minutes for each participant), the simplicity of the movement further allowed us to keep participants' fatigue under control.

### 2.2.3 Measurement

To evaluate the relationship between sensory sensitivity and the participants' performance, we collected performance metrics and questionnaires described below.

#### 2.2.3.1 Performance metrics

As described in Section 2, the framework recorded the participant skeleton and the "template" skeleton, see Fig. 4. These recordings enabled comparing each participant's performance in imitating the movement to the template during the *visible imitation* and *occluded imitation* conditions. We used the participant's hand position as a predictor of good performance, as the hand's movement is the widest compared to other keypoints involved in an abduction. As shown in Fig. 4, the abduction movements (i.e. the arm is resting alongside the body and is then spread away from the midline of the body) occurs mainly in the axes X and Y (perpendicular to the body and parallel to the body, respectively), this is why we do not consider the Z-axis in the measurements. As our framework does not enable us to guaranty the synchrony in the recording of the participants and the template, we corrected the lag between the signals. We computed the cross-correlation between the participant and template's hand position in X and Y axes for each condition to determine the delay between the two signals. We corrected the lag between the two

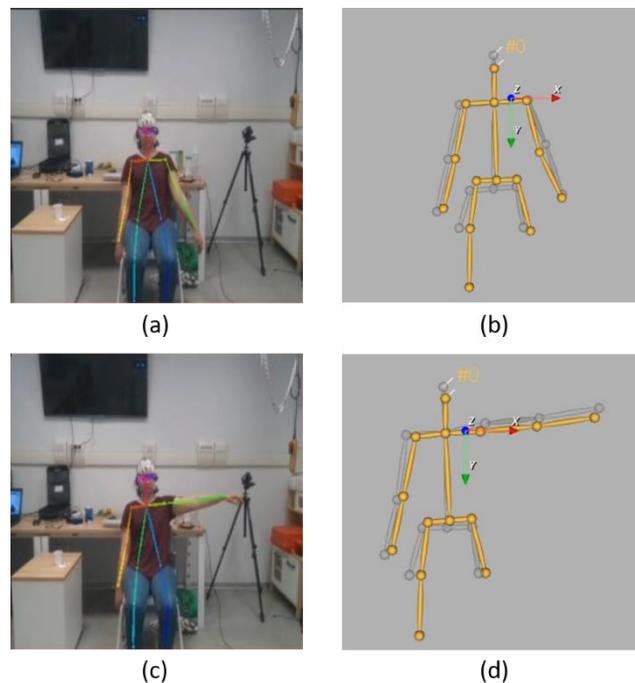


Figure 4. Captures of the participant's and template's skeletons during the abduction movement of the left arm. The movement occurs mainly in the X and Y axis (respectively in red and green in (b) and (d)). The participant's skeleton is tracked in the robot's video feed as seen in (a) and (c). With the depth-sensor data and as explained in Section 2, the 3D-skeleton of the participant, in yellow in (b) and (d), is extracted. The template's skeleton is visible in grey in (b) and (c).

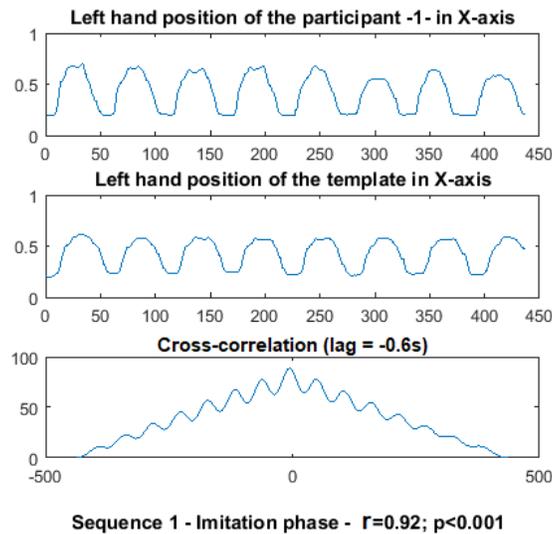


Figure 5. The template's (top) and participant's (middle) hand positions in the X-axis during the imitation phase of the first sequence with the virtual robot. The cross correlation between the two signals is computed (bottom) and the lag between the two signals is corrected to compute the correlation coefficient ( $r$ ) between the two signals. The correlation coefficient is taken as an indicator of good performance.

signals and computed their correlation coefficients, which indicates the “closeness” from one signal to another (see Fig. 5). The correlation coefficient depends on differences in the amplitude and the frequency of signals: a hand moving at a different velocity, frequency or amplitude than the robot's hand would produce a lower correlation coefficient.

We computed the mean correlation score of each participant for the whole experiment per embodiment type (*virtual robot*; *physical robot*) and per condition (*visible imitation*; *occluded imitation*), resulting to each participant obtaining nine scores in both X and Y axes.

#### 2.2.3.2 Sensory Perception Quotient (SPQ)

To screen the participants' sensory sensitivity, we used the *Sensory Perception Quotient* (SPQ)<sup>2</sup> [26]. This self-reported questionnaire enables determining one's sensory sensitivity in five modalities: Hearing, Vision, Touch, Taste, and Smell. The questionnaire contains 92 items, and scoring them by subscales enables obtaining one's sensitivity for each modality. The higher the score is, the more sensitive the person is. All participants filled out the questions only for the Vision, Hearing and Touch subscales of the SPQ (20 questions each). Each question can be answered by “strongly agree”; “agree”; “disagree”; and “strongly disagree”. The scoring key give a table for each question corresponding to an answer, each question can score “0” or “1”. We compute for each subscale (Hearing; Visual; and Touch) the score corresponding to the sum of each question of a subscale. Therefore, each participant had a score between 0 and 20 for Visual, Hearing and Touch sensibility. The authors of the SPQ indicated that a higher score

<sup>2</sup> The Sensory Perception Quotient and its scoring key can be found at [https://www.autismresearchcentre.com/arc\\_tests](https://www.autismresearchcentre.com/arc_tests)

corresponds to a higher sensitivity, but they did not indicated ranges. The SPQ was originally developed for a population with ASD, to screen their sensory perception, however, the authors of the SPQ state that the questionnaire can be used for population with and without ASD [26]. We did not ask our participants to complete the Taste and Smell subscales, as these senses were not related to the experiment. However, as we used earplugs in our setup to cancel the high background and R1's actuator noise, we did not use the Hearing subscale in our analysis as we are unsure on how it affected the participants' hearing sensitivity.

#### 2.2.3.3 Reported strategy questionnaire

After completing the six sequences of movements with the virtual and the physical robot, the participants were asked to describe what strategies they used to imitate the robot when the arm was occluded.

#### 2.2.4 Data analysis

A 2 by 2 between-groups analysis of covariance (ANCOVA) was conducted to assess the effect of the different robots on the participants' performance in visible and occluded imitations. The independent variables were the type of robot (virtual, physical) and the imitation conditions (visible, occlusion). The dependent variable was the performance in the X and Y-axis of the participants. The mean-centered visual sensitivity score from the SPQ was used as a covariate. We ran a similar ANCOVA but with mean-centered sensitivity to touch as a covariate instead of sensitivity to vision. In view of the results of the ANCOVA, we then computed bivariate (Pearson) correlations between the participants' performance metrics in X and Y axes and their visual sensitivity score.

### 3 RESULTS

#### 3.1 Sensory Sensitivity Effects

On the X-axis, there was a significant main effect of the robot's embodiment ( $F(1, 14)=7.81, p=.014$ , with a large effect size ( $\text{partial eta squared} = .358$ )) and of the imitation conditions ( $F(1, 14)=8.50, p=.011$ , with a large effect size ( $\text{partial eta squared} = .378$ )). There was also a significant interaction effect between the robot's embodiment and the imitation conditions ( $F(1, 14)=7.75, p=.015$ , with a large effect size ( $\text{partial eta squared} = .356$ )). These results suggest that both the robot's embodiment and the imitation conditions impacted the participants' performance on the X-axis. See Figure 6 for a visual representation of the interaction effect with visual sensitivity. We followed up this analysis with correlations to observe the effect of the covariate (visual sensitivity score).

We found significant relationships between the X-axis of the participants' performance and the visual sensitivity of the participants. Firstly, we found that a higher visual sensitivity was positively correlated with a better performance of the participants over the whole experiment ( $r=.606, p=.0128, 95\%CI [0.158, 0.847]$ ) (see Fig. 7).

Regarding the embodiment types, we found that a higher visual sensitivity was positively correlated with a better performance of the participants when interacting with the *virtual* robot ( $r=.605, p=.0129, 95\%CI [0.156, 0.847]$ ), but we did not found any relationship in the *physical* robot condition ( $r=.188, p=ns, 95\%CI [-0.339, 0.626]$ ) (see Fig. 8).

Regarding the condition of the task, we found that a higher visual sensitivity was positively correlated with a better performance of the participants in *occluded* imitation ( $r=.614, p=.0114, 95\%CI [0.170, 0.851]$ ), but we did not find any relationship for *visible* imitation ( $r=.294, p=ns, 95\%CI [-0.236, 0.689]$ ) (see Fig. 9).

Regarding the interactions between the embodiment types of the robot and the condition of the interaction, we only found that a higher visual sensitivity was positively correlated with better performance of the participants when interacting with *the virtual robot when its arm was occluded* ( $r=.607, p=.0126, 95\%CI [0.160, 0.848]$ ) (see Fig. 10).

We did not find any significant relationships between visual sensitivity and participants' performance along the Y-axis of the motion. The ANCOVA with sensitivity to touch yielded no significant main effect for sensitivity to touch in either X or Y-axis. There were also no significant interactions involving touch. We did not find any significant relationships in the correlations between touch sensitivity and participants' performance. The full results can be found in the Supplementary Materials.

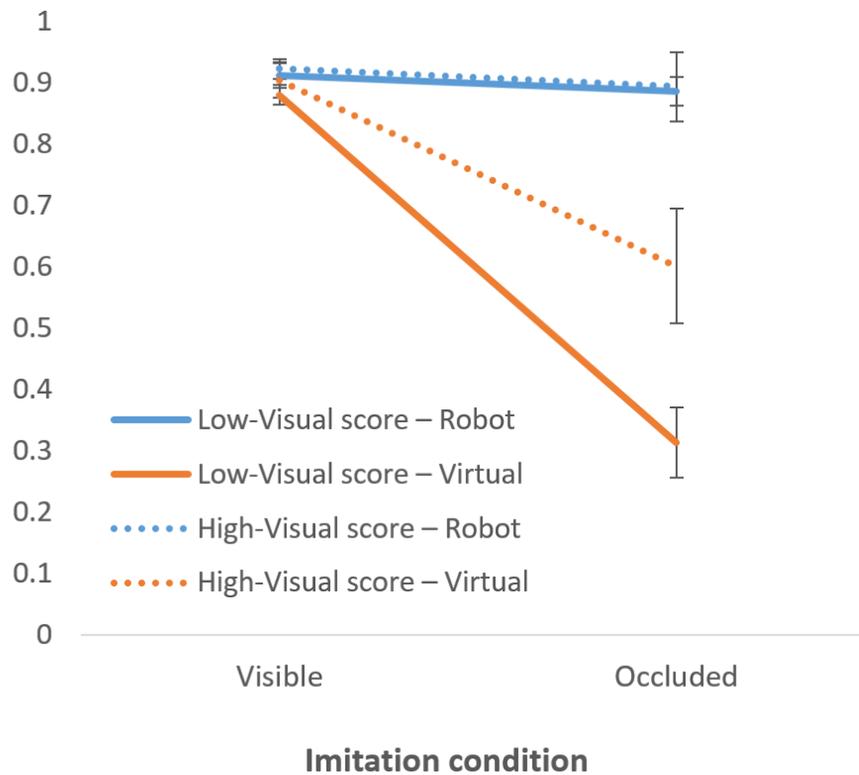


Figure 6. Visualization of the interaction effect (Robot embodiment x Imitation condition) of the participants' performance in the X-axis with vision sensitivity. We used the median split to create a group of Low-visual sensitivity ( $n = 6$ ) and a group of High-visual sensitivity ( $n = 5$ ; the remaining participants scored the median of 9, and therefore could not be grouped). Within-subjects standard errors [31] were used for the error bars. This graph is purely for visualization purposes. Because of the small remaining group sizes, only the between-subjects main effect of visual sensitivity group was significant, but none of the interactions anymore. We believe this graphs really shows how the participants' performance in X-axis drops only if the virtual, but not if the physical robot is occluded.

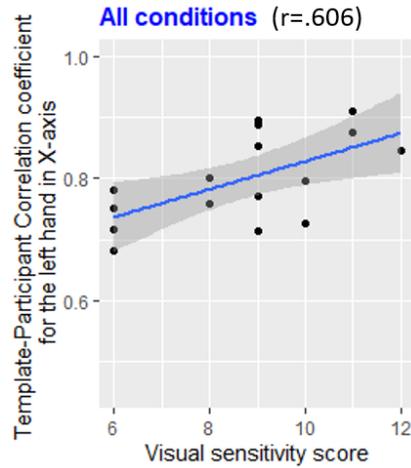


Figure 7. Scatter plot of the Template-Participant correlation coefficient for the left hand in the X-axis in function of the participants visual sensitivity score, for the whole experiment. The more a participant has a High-visual sensitivity score, the better they performed during the experiment.

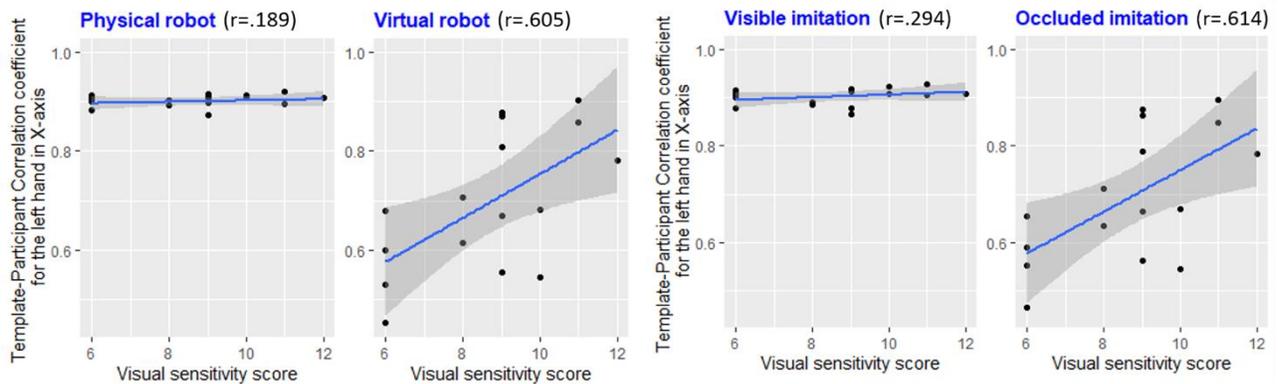


Figure 8. Scatter plot of the Template-Participant correlation coefficient for the left hand in the X-axis in function of the participants visual sensitivity score, for both embodiment types (physical – left; and virtual - right). The more a participant has a High-visual sensitivity score, the better they performed when interacting with the virtual robot. This relationship was not found with the physical robot.

Figure 9. Scatter plot of the Template-Participant correlation coefficient for the left hand in the X-axis in function of the participants visual sensitivity score, for both conditions (visible imitation – left; and occluded imitation - right). The more a participant has a High-visual sensitivity score, the better they performed when performing the occluded imitation. This relationship was not found when they performed the visible imitation.

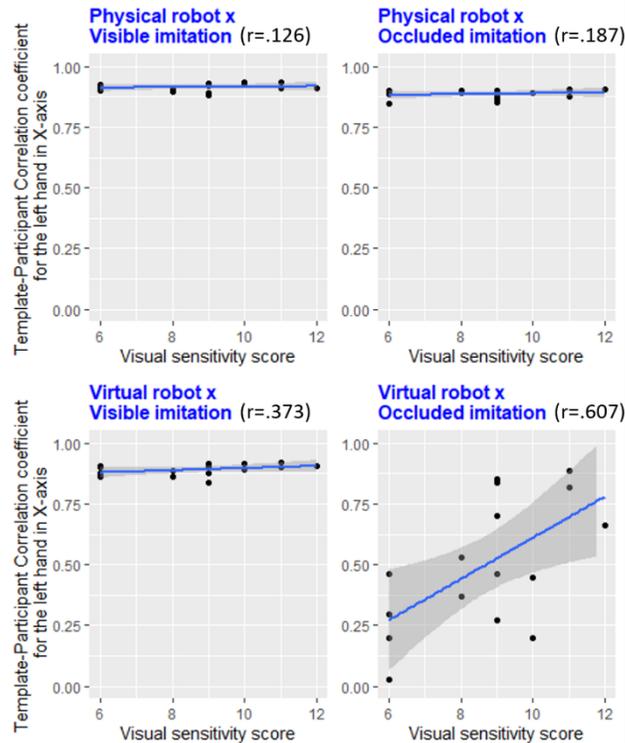


Figure 10. Scatter plot of the Template-Participant correlation coefficient for the left hand in the X-axis in function of the participants visual sensitivity score, for the interactions between the embodiment types (virtual and physical) and the conditions (visible imitation and occluded imitations). The more a participant has a High-visual sensitivity score, the better they performed when performing the occluded imitation with the virtual robot. The participants' performance was not affected by their visual sensitivity in the other conditions.

### 3.2 Reported strategy questionnaire

Participants used different strategies to continue following the robot's movement during the *occluded imitation* conditions. They reported a combination of strategies, involving the auditory and visual cues delivered by the robot or internal processes such as counting or following their body rhythm (e.g., breathing). See in Table 2 some of the strategies reported by our participants. We categorized our participants' strategies by classifying them as "auditory", "visual" or "internal" strategies (see Table 3 for a contingency table of the participants' strategies). When interacting with the physical robot, five participants out of 16 reported to rely on visual cues, 10 reported relying on auditory cues and eight on internal processes. Ten of the participants used auditory cues even if they were wearing earplugs, as it did not neutralize all the sounds from the robot's motors. During the interaction with the virtual robot, four participants reported having relied on visual cues, and all of them reported having relied on internal processes, such as counting or synchronizing with their breath to perform the task.

**Table 2. Examples of the participants' strategies to continue following the robot's movement during the occluded imitation conditions**

Visual strategies	<i>Physical robot</i>
	"[I used the] reflection of the arm on its face"; "[I used] the vibrations of the right arm during the movement"
	<i>Virtual robot</i>
	"I used the shadow and the little movement of R1's left arm"; "I used first the oscillation of [R1's] body, that was lining towards the direction of the arm that was raising"
Auditory strategies	<i>Physical robot</i>
	"I was not seeing the robot's arm but I was hearing its movement, so I followed its rhythm"; "I heard its mechanism"
Internal strategies	<i>Physical robot</i>
	"I counted the seconds when I was still able to see the movements, and I continued with the same rhythm"; "I was counting my breath and replicate the rhythm and number in the occluded part"
	<i>Virtual robot</i>
	"Creating a kind of metronome in my head for the movement and counting the repetitions"; "I remembered the rhythm between movements"

**Table 3. Classification of the participants' (N=16) strategies as "auditory", "visual" or "internal" strategies when asked to imitate the physical and virtual robots during the Occluded Imitation condition. Participants may have used a combination of strategies to perform the task**

Agent	Strategies		
	Auditory	Visual	Internal
Physical Robot	10	5	8
Virtual Robot	0	4	16

## 4 DISCUSSION

In this study, we evaluated the relationship between sensory sensitivity and the performance on an imitation task with a virtual and a physical robot. From the reported strategy questionnaire, we confirm that sensory strategies based on visual and hearing cues were used by the participants of our study alongside more internal strategies, like counting or synchronizing the movement with internal processes, like breathing. We observed that to perform the task with the virtual robot, the participants might have invested more cognitive effort as they all reported using counting or other internal processes to follow the movement, while only half of them used these strategies with the physical robot. Indeed, with the physical robot, participants reported to rely more on auditory and visual cues to perform the task, suggesting that the physical embodied robot provides indeed more physical presence than its virtual counterpart.

The participants' performance in following the trajectory of the robot's hand during the task showed a number of relationships with their visual sensitivity as scored with the SPQ questionnaire. Participants with a higher score in visual sensitivity performed the task in the X-axis better across the entire experiment. However, this relationship between visual sensitivity score and performance was observed for interactions with the virtual robot, but not with the physical robot. We also found that participants with a higher sensitivity to vision performed better than those with lower sensitivity in the occluded imitation. Visual sensitivity did not play a role in the visible imitation condition. In sum, the visual sensitivity of participants was of benefit when performing occluded imitation with the virtual robot. Following on from these results, we propose that participants with a higher visual sensitivity might have been better able to detect subtle movements of the robot than participants with a lower sensitivity. For example, they could have been better able to detect the vibrations of the agent's body, which were more subtle to detect on the virtual agent. These types of behaviour are indeed similar to the SPQ items that relate to visual screen sensitivity. This result is in line with previous findings [3], where children diagnosed with autism spectrum disorder who rely more on visual cues were found to perform better in an imitation task.

Contrary to the X-axis, we did not find similar results in the Y-axis. Prior to the analysis of the data, we expected to find similar patterns in the X and Y axes, as the abduction movement of the arm is present and similar in both axes. We did not find an explanation to this difference in the participants' performance in x and y-axis in the literature. However, this behaviour can be explained by basic physics, as it is related to the forces and activation of the muscles applied to the arm during the abduction movement. Indeed, at the initial position (against the body) and the goal position (i.e. away from the body), the arm has a null velocity and therefore a null momentum (i.e. quantity of movement). At the goal position (i.e. away from the body), the arm is in extension and the muscles are activated. We hypothesize that the hand suffers less variation when the arm is stiffer, causing no difference in the variation in X- and Y- axes. Contrary, in the initial position, the arm is close to the resting position (i.e. the arm against the body). Thus, the participants had their relevant muscles passing from an active to a resting state in this position, inducing a wider variation of movement in the X-axis than in the Y-axis. Furthermore, when the arm goes down, the Y component moves in favor of gravity, which might justify the lower variation too.

Finally, we did not find any significant relationship between the participants' performance and touch sensitivity. The tangible nature of the robot's embodiment did not appear to influence the performance of our participants depending on their touch sensitivity. However, touch was not involved in the task itself, so lack of relationship with sensitivity to touch is not surprising.

Interestingly, visual information was discussed to be primordial in synchronization between two interaction partners. Indeed, during a repetitive task, two humans will synchronize [27,28]. As discussed in [27], visual information of the movement enables partners to increase the synchrony. The authors asked participants to perform a puzzle while holding a pendulum, and observed that when the visual stimuli was visible, the synchrony increased. Similarly, [28] observed that people will synchronize with a stimulus displayed in the environment. However, in [29], the authors discussed that in human-robot interaction, the synchronization does not emerge naturally. The robot needs to adapt to the human movements during the interaction, to retrieve this natural phenomenon. In our paper, we did not record the difference of phase between the participants and the robots, but we nonetheless observed that visual cues (from the robot vibration for example) could help in the imitation in frequency and amplitude in the occluded imitation phase, as there was no impact of occlusion on participants' performance.

As participants' performance in the physical robot condition was practically at ceiling, with no impact of occlusion, we suggest that a physical embodiment requires less cognitive strategies in continuing an imitation task under conditions with lack of sensory information, compared to its virtual counterpart. This is perhaps not surprising due to the larger amount of sensory cues that embodiment provides, even when the most relevant cue – the arm in our case – is unavailable. The interesting finding of our study is that people with higher visual sensitivity are able to keep a rhythm of imitation in the occluded condition (with virtual robot) better than those with lower visual sensitivity. This might be due to a better consolidation of sensorimotor representation of the task. Finally, it is also important to note that overall, the performance in the virtual occluded condition was substantially lower than performance in the other conditions (physical robot and visible imitation; physical robot and occluded imitation; and virtual robot and visible imitation). This is due to lack of the additional sensory cues that the virtual agent's embodiment provides, even in the absence of the most relevant sensory input. Hence, this speaks in favour of the idea that physical embodiment provides physical presence to a larger extent and allows for compensatory strategies, such as picking up additional cues and signals in the absence of the relevant channel.

Given the relationship we found, the absence of visual input might detrimentally affect a typical rehabilitation therapy. However, the quality of the therapy could benefit from the use of an embodied platform, specifically designed to enhance different sensory channels. Therefore we believe that our findings are important when devising strategies for delivering rehabilitation. In the rehabilitation contexts this should be taken into account when considering whether physical embodiment can be substituted by only virtual agents. For example, virtual agents are widely used to deliver rehabilitation therapies to elderly individuals, however this population shows a degradation in their visual functions [30]. The use of physical robot may be beneficial for them in view of our results.

A limitation of our work is the difference in size of both robots. The virtual robot was indeed smaller than the real robot, as it is shown in the Figure 3. It was already discussed that the size of virtual agent against real artificial agent could play a role in the perception and interaction with the agent (see [19]). Indeed, this review suggests that people respond more favorably to real agent than virtual agents. Results that we also obtained in our previous paper [17]. We cannot exclude this effect to happen in our experiment. However, we made sure that during the task, the movement of the robot was clearly visible in both condition. Another limitation of our work is the subject pool number, as we recruited only 16 participants. This was due to the limited time of the availability of the robot and the long duration the experiment and installation of the participants.

To our knowledge, this is the first study investigating the relations between sensory sensitivity and the performance in an imitation task with a real or a virtual robot. Taken together, we show that visual sensitivity predicts performance in an imitation task, especially in the absence of complete sensory input (as was the case for our virtual robot condition, with occlusion). Even though these results are only preliminary and need to be replicated with larger samples, including patients, they can be informative with respect to designing human-robot interaction for rehabilitation. These results should be followed up with studies with clinical populations, such as children diagnosed with autism spectrum disorder (ASD), who show wider and more pronounced inter-individual variability in sensory sensitivity [6] and elderly people suffering from sensitivity impairments (for examples, hearing or vision loss) due to aging conditions.

## ACKNOWLEDGMENTS

The project leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754490 – MINDED project and by the Joint Lab between IIT and Fondazione Don Carlo Gnocchi Onlus (Valerio Gower and Furio Gramatica). The authors thank Nina Hinz and Serena Marchesi for their assistance with data collection.

The content of this work reflects the author's view only and the EU Agency is not responsible for any use that may be made of the information it contains

## REFERENCES

- [1] S. Andrist, B. Mutlu, and A. Tapus, "Look Like Me: Matching Robot Personality via Gaze to Increase Motivation," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, New York, NY, USA, 2015, pp. 3603–3612, doi: 10.1145/2702123.2702592.
- [2] L. Robert, "Personality in the Human Robot Interaction Literature: A Review and Brief Critique," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3308191, Dec. 2018. Accessed: Sep. 18, 2019. [Online]. Available: <https://papers.ssrn.com/abstract=3308191>.
- [3] P. Chevalier, G. Raiola, J.-C. Martin, B. Isableu, C. Bazile, and A. Tapus, "Do Sensory Preferences of Children with Autism Impact an Imitation Task with a Robot?," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA, 2017, pp. 177–186, doi: 10.1145/2909824.3020234.
- [4] R. Agrigoroaie and A. Tapus, "Influence of Robot's Interaction Style on Performance in a Stroop Task," in *Social Robotics*, 2017, pp. 95–104.
- [5] D. Shepherd, M. Heinonen-Guzejev, M. J. Hautus, and K. Heikkilä, "Elucidating the relationship between noise sensitivity and personality," *Noise Health*, vol. 17, no. 76, pp. 165–171, 2015, doi: 10.4103/1463-1741.155850.
- [6] APA, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.
- [7] J. Ward, "Individual differences in sensory sensitivity: A synthesizing framework and evidence from normal variation and developmental conditions," *Cogn. Neurosci.*, vol. 10, no. 3, pp. 139–157, Jul. 2019, doi: 10.1080/17588928.2018.1557131.
- [8] A. E. Robertson and D. R. Simmons, "The relationship between sensory sensitivity and autistic traits in the general population," *J. Autism Dev. Disord.*, vol. 43, no. 4, pp. 775–784, 2013.
- [9] E. N. Aron and A. Aron, "Sensory-processing sensitivity and its relation to introversion and emotionality," *J. Pers. Soc. Psychol.*, vol. 73, no. 2, pp. 345–368, 1997, doi: 10.1037/0022-3514.73.2.345.
- [10] P. Chevalier, J.-C. Martin, B. Isableu, C. Bazile, and A. Tapus, "Impact of sensory preferences of individuals with autism on the recognition of emotions expressed by two robots, an avatar, and a human," *Auton. Robots*, pp. 1–23, 2016.
- [11] P. Chevalier, J.-C. Martin, B. Isableu, C. Bazile, D.-O. Iacob, and A. Tapus, "Joint Attention Using Human-Robot Interaction: Impact of Sensory Preferences of Children with Autism, to appear," *RO-MAN 2016*, 2016.
- [12] R. Wilk and M. J. Johnson, "Usability feedback of patients and therapists on a conceptual mobile service robot for inpatient and home-based stroke rehabilitation," in *5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*, Aug. 2014, pp. 438–443, doi: 10.1109/BIOROB.2014.6913816.
- [13] K. Winkle, P. Caleb-Solly, A. Turton, and P. Bremner, "Social Robots for Engagement in Rehabilitative Therapies: Design Implications from a Study with Therapists," in *Proceedings of the*

- 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, Feb. 2018, pp. 289–297, doi: 10.1145/3171221.3171273.
- [14] A. Tapus, C. Țăpuș, and M. J. Mataric, “User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy,” *Intell. Serv. Robot.*, vol. 1, no. 2, pp. 169–183, 2008.
- [15] B. Scassellati, H. Admoni, and M. Mataric, “Robots for Use in Autism Research,” *Annu. Rev. Biomed. Eng.*, vol. 14, no. 1, pp. 275–294, 2012, doi: 10.1146/annurev-bioeng-071811-150036.
- [16] A. Parmiggiani *et al.*, “The design and validation of the R1 personal humanoid,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2017, pp. 674–680, doi: 10.1109/IROS.2017.8202224.
- [17] V. Vasco *et al.*, “Train With Me: A Study Comparing a Socially Assistive Robot and a Virtual Agent for a Rehabilitation Task,” presented at the International Conference of Social Robotics, 2019.
- [18] E. Deng, B. Mutlu, and M. J. Mataric, “Embodiment in Socially Interactive Robots,” *Found. Trends® Robot.*, vol. 7, no. 4, pp. 251–356, Jan. 2019, doi: 10.1561/23000000056.
- [19] J. Li, “The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents,” *Int. J. Hum.-Comput. Stud.*, vol. 77, pp. 23–37, May 2015, doi: 10.1016/j.ijhcs.2015.01.001.
- [20] G. Metta, P. Fitzpatrick, and L. Natale, “YARP: Yet Another Robot Platform,” *Int. J. Adv. Robot. Syst.*, vol. 3, no. 1, p. 8, Mar. 2006, doi: 10.5772/5761.
- [21] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” *ArXiv181208008 Cs*, Dec. 2018, Accessed: Sep. 18, 2019. [Online]. Available: <http://arxiv.org/abs/1812.08008>.
- [22] R. Bertani, C. Melegari, M. C. De Cola, A. Bramanti, P. Bramanti, and R. S. Calabrò, “Effects of robot-assisted upper limb rehabilitation in stroke patients: a systematic review with meta-analysis,” *Neurol. Sci.*, vol. 38, no. 9, pp. 1561–1569, Sep. 2017, doi: 10.1007/s10072-017-2995-5.
- [23] J. C. Pulido, J. C. González, C. Suárez-Mejías, A. Bandera, P. Bustos, and F. Fernández, “Evaluating the Child–Robot Interaction of the NAOTherapist Platform in Pediatric Rehabilitation,” *Int. J. Soc. Robot.*, vol. 9, no. 3, pp. 343–358, Jun. 2017, doi: 10.1007/s12369-017-0402-2.
- [24] B. Robins, P. Dickerson, P. Stribling, and K. Dautenhahn, “Robot-mediated joint attention in children with autism: A case study in robot-human interaction,” *Interact. Stud.*, vol. 5, no. 2, pp. 161–198, 2004.
- [25] S. Schneider, B. Wrede, C. Cifuentes, S. S. Griffiths, and S. Wermter, “PREC 2018: Personal Robots for Exercising and Coaching,” in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA, 2018, pp. 401–402, doi: 10.1145/3173386.3173566.
- [26] T. Tavassoli, R. A. Hoekstra, and S. Baron-Cohen, “The Sensory Perception Quotient (SPQ): development and validation of a new sensory questionnaire for adults with and without autism,” *Mol. Autism*, vol. 5, no. 1, p. 29, Apr. 2014, doi: 10.1186/2040-2392-5-29.
- [27] M. J. Richardson, K. L. Marsh, and R. C. Schmidt, “Effects of Visual and Verbal Interaction on Unintentional Interpersonal Coordination,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 31, no. 1, pp. 62–79, 2005, doi: 10.1037/0096-1523.31.1.62.
- [28] R. C. Schmidt, M. J. Richardson, C. Arsenault, and B. Galantucci, “Visual tracking and entrainment to an environmental rhythm,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 33, no. 4, pp. 860–870, Aug. 2007, doi: 10.1037/0096-1523.33.4.860.
- [29] T. Lorenz, A. Weiss, and S. Hirche, “Synchrony and Reciprocity: Key Mechanisms for Social Companion Robots in Therapy and Care,” *Int. J. Soc. Robot.*, vol. 8, no. 1, pp. 125–143, Jan. 2016, doi: 10.1007/s12369-015-0325-8.

- [30] N. S. Majid, N. E. Badarudin, and N. A. Yahaya, "The awareness of visual function deterioration in the elderly: a review," *J. Optom. Eye Health Res.*, vol. 1, no. 1, 2019, Accessed: Aug. 25, 2020. [Online]. Available: <http://mymedr.afpm.org.my/publications/79842>.
- [31] G. R. Loftus and M. E. J. Masson, "Using confidence intervals in within-subject designs," *Psychon. Bull. Rev.*, vol. 1, no. 4, pp. 476–490, Dec. 1994, doi: 10.3758/BF03210951.

## Supplementary Materials

### Instruction to the participants

The following instructions were provided by the experimenters to the participants:

“The robot will demonstrate a simple exercise. Please just observe and try to sit still. When the robot says “Now do the same with me” try to copy the exact same motion together with the robot, using your left arm. Then, we will put a panel between you and the robot’s arm, so you can’t see it anymore. It is very important that you keep looking at the robot: try not to be distracted by the placement of the panel. Even when you can’t see the robot’s arm anymore, try to keep copying the same motion.”

“After this exercise, please take a break for as long as you need. When you indicate that you’re ready, we will repeat the exercise, followed by another break until we have completed the exercise six times. During each break, you can take as long as you need. If you feel any discomfort, please stop immediately. This has no consequences for your participation or your payment.”

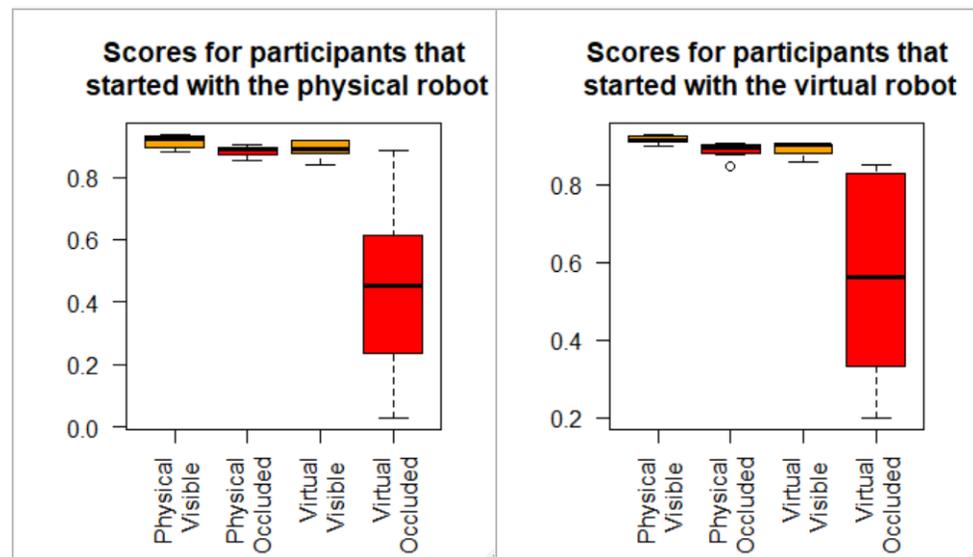
“After these exercises, you can have a rest while we change the set-up. When you’re ready, we will do the same, but this time via the screen / physical robot”

“To reiterate:

- Try to remain still when you’re just observing the robot. When you’re copying the robot, try to only move your left/right arm.
- You decide how long your breaks are.
- You can stop at any time.”

### Additional results

1. Plots of the participants’ performance (i.e. the cross-correlation between the participant and template’s hand position in X axes, R) depending of their first interaction partner (physical or virtual robot)



2. Sensory Sensitivity Effects on the cross-correlation between the participant and template's hand position in the observed axes

a. ANCOVA with sensitivity to vision as covariate in X-axis

Condition	F(1,14)	p	partial eta squared
Robot	7.81	0.014*	0.358
Imitation	8.50	0.011*	0.378
Robot x Imitation	7.75	0.015*	0.356

b. Correlations between the X-axis of the participants' performance and the visual sensitivity of the participants.

Condition	R	p	95%CI
All conditions	0.606	0.013*	[0.158;0.847]
Physical Robot	0.189	0.484	[-0.339;0.626]
Virtual Robot	0.605	0.013*	[0.156;0.847]
Visible Imitation	0.294	0.268	[-0.236;0.689]
Occluded Imitation	0.614	0.011*	[0.170;0.851]
Physical Robot x Visible Imitation	0.126	0.641	[-0.394;0.585]
Physical Robot x Occluded Imitation	0.187	0.488	[-0.340;0.625]
Virtual Robot x Visible Imitation	0.373	0.154	[-0.150;0.733]
Virtual Robot x Occluded Imitation	0.607	0.013*	[0.150;0.848]

c. ANCOVA with sensitivity to vision as covariate in Y-axis

Condition	F(1,14)	p	partial eta squared
Robot	2.59	0.130	0.156
Imitation	3.52	0.082	0.201
Robot x Imitation	3.71	0.750	0.209

- d. Correlations between the Y-axis of the participants' performance and the visual sensitivity of the participants.

Condition	R	p	95%CI
All conditions	0.310	0.243	[-0.219;0.698]
Physical Robot	-0.392	0.133	[-0.743;0.129]
Virtual Robot	0.353	0.180	[-0.173;0.722]
Visible Imitation	-0.141	0.602	[-0.595;0.381]
Occluded Imitation	0.375	0.152	[-0.148;0.734]
Physical Robot x Visible Imitation	-0.288	0.280	[-0.686;0.243]
Physical Robot x Occluded Imitation	-0.473	0.064	[-0.785;0.030]
Virtual Robot x Visible Imitation	-0.073	0.788	[-0.549;0.439]
Virtual Robot x Occluded Imitation	0.401	0.124	[-0.118;0.748]

- e. ANCOVA with sensitivity to touch as covariate in X-axis

Condition	F(1,14)	p	partial eta squared
Robot	0.032	0.862	0.020
Imitation	0.000	0.990	0.000
Robot x Imitation	0.013	0.912	0.001

- f. Correlations between the X-axis of the participants' performance and the Touch sensitivity of the participants.

Condition	R	p	95%CI
All conditions	-0.086	0.751	[-0.558;0.428]
Physical Robot	-0.214	0.426	[-0.642;0.315]
Virtual Robot	-0.068	0.803	[-0.545;0.443]
Visible Imitation	-0.363	0.167	[-0.728;0.162]
Occluded Imitation	-0.045	0.869	[-0.529;0.461]
Physical Robot x Visible Imitation	-0.330	0.211	[-0.710;0.198]
Physical Robot x Occluded Imitation	-0.048	0.860	[-0.531;0.459]
Virtual Robot x Visible Imitation	-0.342	0.194	[-0.716;0.185]
Virtual Robot x Occluded Imitation	-0.042	0.877	[-0.527;0.463]

g. ANCOVA with sensitivity to touch as covariate in Y-axis

Condition	F(1,14)	p	partial eta squared
Robot	0.070	0.795	0.005
Imitation	0.047	0.832	0.003
Robot x Imitation	0.031	0.863	0.002

h. Correlations between the Y-axis of the participants' performance and the Touch sensitivity of the participants.

Condition	R	p	95%CI
All conditions	-0.008	0.978	[-0.501;0.490]
Physical Robot	-0.436	0.092	[-0.766;0.077]
Virtual Robot	0.031	0.910	[-0.472;0.519]
Visible Imitation	-0.160	0.554	[-0.607;0.365]
Occluded Imitation	0.022	0.937	[-0.479;0.512]
Physical Robot x Visible Imitation	-0.436	0.091	[-0.766;0.076]
Physical Robot x Occluded Imitation	-0.394	0.131	[-0.745;0.126]
Virtual Robot x Visible Imitation	-0.044	0.872	[-0.528;0.462]
Virtual Robot x Occluded Imitation	0.041	0.881	[-0.464;0.526]