UNIVERSITY OF GENOVA

PHD PROGRAM IN BIOENGINEERING AND ROBOTICS

Motion for cooperation and vitality in Human-robot interaction

by

Fabio Vannucci

Thesis submitted for the degree of *Doctor of Philosophy* (32° cycle)

December 2019

Prof Giulio Sandini	Supervisor
Dr Alessandra Sciutti	Supervisor
Dr Francesco Rea	Supervisor

Thesis Jury:

Prof Selma Sabanovic, *Indiana University*	External examiner
Dr Shelly Levy Tzedek , *Ben-Gurion University of the Negev*	External examiner
Prof Silvio Sabatini, *Università degli Studi di Genova*	Internal examiner

Dibris

Department of Informatics, Bioengineering, Robotics and Systems Engineering

A nonna Gina

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Fabio Vannucci
March 2020

Acknowledgements

All the colleagues and IIT staff from the 7th floor of Erzelli, from Morego and San Quirico.

Prof. Yukie Nagai, Prof. Minoru Asada and their groups that so kindly hosted me in Japan.

All the current and former members of RBCS and CONTACT, especially my supervisors Alessandra, Francesco, Prof. Giulio Sandini, and Ana.

Manfredi la GANG.

My friends.

My family for supporting my choice of starting a PhD, for the passion and interest that they always have in what I do.

Abstract

In social interactions, human movement is a rich source of information for all those who take part in the collaboration. In fact, a variety of intuitive messages are communicated through motion and continuously inform the partners about the timing and the future unfolding of the actions. Also the style of an action, i.e. the way it is performed, has a strong influence on interaction between humans. The same gesture has different consequences when it is performed aggressively or kindly, and humans are very sensitive to these subtle differences in others' behaviors. A similar exchange of implicit information could support movement coordination in the context of Human-Robot Interaction.

During my PhD, I focused on these aspects of human motion and on their potential role in human-robot interaction. In a first study, we investigated how implicit signaling in an interaction with a humanoid robot can lead to emergent coordination in the form of automatic speed adaptation. Moreover, we assessed whether different cultures – specifically Japanese and Italian – have a different impact on motor resonance and synchronization in HRI. Since Japanese culture shows a higher general acceptance toward robots when compared with Western cultures and acceptance, or better affiliation, is tightly connected to imitation and mimicry, we hypothesized a higher degree of speed imitation for Japanese participants when compared to Italians. In the experimental studies undertaken both in Japan and Italy, we observed a clear influence of the robotic speed on human actions. Cultural differences did not impact on the natural predisposition of subjects to adapt to the robot.

The ability to naturally influence the action timing of the human partner with no explicit instructions, could play an important role in all applications that require a tight coordination between humans and robots.

In a second study, we investigated how to endow a humanoid robot with behaviors expressing different vitality forms, by modulating robot action kinematics and voice. Drawing inspiration from humans, we modified actions and voice commands performed by the robot to convey an aggressive or kind attitude. In a series of experiments we demonstrated that the humanoid was consistently perceived as aggressive or kind. fMRI analyses demonstrated that a similar neural activation can be triggered by the observation of a human or a robotic

action, as far as specific properties of the movement kinematics, reflecting action style, are preserved. Moving to an actual interactive context, human action style changed in response to the different robot attitudes and matched the behavior of iCub. In fact participants exhibited aggressive kinematics when the robot was aggressive and kinematic properties typical of gentle behavior when the robot was gentle.

Such behavior might be crucial in emergency and in authoritative situations in which the robot should instinctively be perceived as assertive and in charge, as in case of police robots or teachers.

The opportunity for a humanoid robot to express implicit communication in its behavior, enriches the array of nonverbal signals that can be exploited to foster seamless interaction.

Table of contents

List of figures

List of tables

Part I

Section One

Chapter 1

Introduction

1.1 Movement-based communication

Communication between humans comes naturally, as we are very skilled in coordinating with each other even without the need of verbal instructions. Both explicit and implicit signals contribute to the high efficiency of human-human interaction (1).

Explicit signals, such as gestures, are performed with the specific intention of providing information to someone else. Implicit signals, instead, are not under our complete control but they are equally important to carry information. For instance, when we want to take a tool, our gaze would focus on the object before we even start the motion, helping our partners to predict our goal and intention (2). This example shows that implicit communication is embedded in our way of moving and does not need awareness of the "sender" or the "receiver". In fact, such signals are produced involuntarily also when there is no one else around: a person would look at the object before moving in any case. This is due to the fact that all humans have in common certain motor regularities that characterize action execution, and these regularities are unconsciously interpreted when we observe our peers (3). The ability to understand these latent signals is exhibited in many other situations and contexts, for example when we detect the emotional status of someone from voice features (4) or from style of the body movements (5).

Implicit signals are important because they represent the foundation of "emergent coordination" (1; 6). This is a feature of human interaction that allows efficient collaboration through mutual adaptation, synchronization and anticipation without the need for direct verbal instructions (7; 8; 9). Thanks to "emergent coordination" we exchange covert signals that help us "to deal with the real-time aspects of joint action".

Human-robot interaction could greatly improve if robots were able to both perceive and send such signals. Recognition and understanding of these human cues is fundamental to predict the goals and intentions of the partner, while the ability to express in the same way is very important to achieve an effortless interaction. Both features enhance the relationship with the robot, making it more natural. They would also simplify the exchange of information: for example when we ask a robot to take an object, language interpretation and noise in the environment could make our request quite hard to recognize. Moreover, dealing with the differences among people and languages would add other layers of complexity, while we can easily identify the gaze direction with only a camera. In addition, humans are often persuaded that humanoid robots have their same set of perceptual and motor skills, due to the very similar anthropomorphization (10). For example a person usually takes it for granted that the robot can not only hear and see as we do, but also understand speech and correctly interpret actions (11). Conversely, speech processing is still a quite challenging task, especially in natural noisy and unconstrained scenarios. These are the main reasons why would be important to exploit implicit communication channels.

1.2 State of the art

Human actions are recognized by human brain since our movements differ from any other kind of motion (12). Robots with humanoid shape and presenting human-like behavior could easily trigger these neural mechanisms and facilitate human-robot interaction. The same "emergent coordination" exhibited by humans could be replicated with a robot, leading to a more natural and efficient interaction (13). But there could be also a disadvantage: a robot displaying behaviors and movements that do not follow the human style as we expect, would worsen the interaction. Such mismatch between our expectations and the reality could induce discomfort in the partner, similar to the uncanny valley effect (14).

There is evidence that, in simple scenarios, a humanoid robot displaying human like behavior can convey the aforementioned implicit signals and provoke the same automatic answers of humans. One example is the gaze: paying attention to the eyes of a humanoid robot in a collaborative task, helped reducing the reaction times of the human partner who did not have to wait the full speech instructions indicating the location of an object. The participant in fact, could predict the position of the hidden object thanks to the gaze direction of the robot in the same way of human-human interactions (15) and more in general many studies demonstrated that people is more positively affected by robots that display a consistent social gaze (16). On the motion side, goal-oriented transport actions performed by a humanoid robot seems to be processed similarly to the ones performed by a human actor, triggering the attention of the observer towards the goal with approximately the same degree of anticipation (17). Also lifting actions seem to be processed likewise since participants were able to deduce the weight of unknown lifted objects with the same accuracy when the lifting was performed by the human actor or the robot (18). Another behavioral response related to motor resonance that is comparable between human and humanoid robot is priming: for example this effect was present with similar size when observing the goal-oriented actions of a human or of a robot (19) and also during repetitive game sessions with a robot (20; 21). Motor interference is another process that has been demonstrated to occur for human-human and human-robot interaction (22). In this case, the interference effect generated by observing another agent moving was significant both for a human and robotic partner. Motor contagion in the form of emergent coordination also arises after the observation of hand actions adopting human-inspired kinematics (23). To support this, was also found that the observation of human and robot actions activate the mirror-neuron system, but watching just a robot performing repetitive movements does not trigger the same effect. (24).

All these data confirm that simple cooperative actions like passing, reaching, transporting or grasping, performed by a humanoid robot and by a human can be interpreted in a similar way. But it is also very important to note that for this to happen, it is crucial to respect some rules that characterize human behavior. In fact most of the mentioned studies could find coherence or matching between human-human and human-robot interaction, only when the implicit signals of biological communication were conveyed properly (25).

Considering that present technology allows humanoid robots to easily perform such basic movements, deepening the analysis of implicit signals and embedding them in the robot motions is already possible and could facilitate the interaction of naive users with complex platforms like humanoid robots. Keeping in mind the advantages and limitations of this approach, during the three years of my PhD I tried to address specifically some of the movement-based implicit signals by giving the iCub humanoid robot the ability to express such signals with motion.

1.3 Research focus

As previously mentioned, the human way of moving is very peculiar and interpreted by the brain in a particular way. It is well known that we can recognize biological motion since very early infancy (26). This kind of motion is important also because we can deduce a lot of information from it, like for example the goal of the action or the identity and emotional state of the actor (17; 27). Most of the times, these are all signals that we "send" and "receive" unconsciously, communicating things such time and place without the need for verbal exchange, making the interaction more immediate and efficient. Body motion is in fact important also to convey why and how that particular action is carried out, revealing the intention and affective state of the counterpart.

During my PhD, we worked mainly on two features of this implicit communication: the effect of biological motion on timing of a collaboration and the expression of an emotional status through movement (vitality forms). The first aspect is more general, fundamental to improve the efficiency of collaboration tasks carried out with robots. As already mentioned, biological motion is perceived and processed automatically by our brain, helping anticipation during an interaction. Implementing such kind of motion on a humanoid robot, would greatly increase the pleasure and effectiveness of the interaction. A lot of researchers tend to focus more on making the robot adapt to the human partner and this has indeed been proven to make cooperation more fluid (28; 29; 30). However, the adaptation between humans is often “mutual”, therefore it might be relevant that also the human partner changes his behavior as a function of collaborator’s behavior in a joint task. In current applications, when it is important to maintain a stable pace in action execution during the interaction with a machine (e.g., in chain production), the compliance of the human is often enforced, with no possibility for the user to diverge. The use of humanoid robots might however trigger an automatic adaptation to the robot rhythm, leading to a more fluid and less fatiguing coordination. Indeed, there is evidence that humanoids moving according to biological motion rules can actually trigger automatic imitation, but so far the demonstration has been limited to actions performed in sequence and not embedded in a joint coordination (23). To investigate this point we designed a study in which a participant had to execute a joint task with the humanoid robot iCub. The two had to fill a box with Lego blocks provided in their hands. We altered the speed with which the robot iCub accomplished the action from relatively fast to very slow, in order to assess the natural tendency of the human partner to adapt in order to maintain synchronization. Our goal was to investigate whether the human partner would adapt to the speed of the robot, in line with results demonstrating that humans tend to exhibit speed adaptation when interacting with an embodied robotic partner, (23; 31) showing that the phenomenon generalizes to simultaneous actions in a semi-shared space. We also investigated whether this adaptation can be influenced by the cultural context in which the interaction takes place, performing the same experiment in Italy and Japan. These two cultures differ substantially in their acceptance toward robots (32). Considering that a strong link exists between acceptance and other mechanisms such as imitation (33) we hypothesized a higher degree of speed adaptation in the Japanese sample, when compared to the Italian one.

Going more in detail regarding movement-based implicit signals, the second aspect we focused on involves the communication of an affective status through motion (vitality forms). It can be important in the scope of human-robot interaction to communicate intuitively the attitude of the other and enable a fast and immediate adaptation. The same action, in fact, acquires a different meaning, also in terms of urgency or importance, when associated with different vitality forms (34). Robots, specifically humanoid robots, might need to express different attitudes in particular contexts such as elderly care or law enforcement. In the former, for example, they could be more gentle displaying slow and fluid motion, to avoid fear or rejection from the users, while in the latter could probably be better to have an assertive robot with the ability to communicate imperative commands and quickly convince a person to follow its instructions during an emergency situation. Research in robotics has often focused on generation and execution of human-like movements in the attempt of creating communicative actions (35; 36; 37). Although the emotional aspect has been reproduced mainly through the use of facial expressions (38; 39) there are also several attempts to communicate affective states with motion (5; 40). For example different authors proposed to generate humanoid motions on the basis of the Laban Movement Analysis (35), that describes the emotion conveyed by movement using features such as velocity, curvature and acceleration (38). However, there are no studies addressing the issue of expressing vitality in the behavior of a humanoid. The challenge we address in this study is to create robot behaviors that can achieve a goal while communicating various vitality forms, by exploiting a modulation of the kinematics of the motor act or of the robot's voice. This studies were carried on in collaboration with the research group of Prof Giacomo Rizzolatti from Universita' di Parma, in particular with Dr Giuseppe Di Cesare. In fact the main inspiration for these studies comes from some of the past publications by Dr Di Cesare regarding how vitality is perceived and processed by the brain, and how this influences our movements during a human-human interaction. In human-human interaction, fast and efficient collaboration is promoted by non-verbal communication (40; 41). Implicit signals such as details in the motion of the body, gaze direction or voice features are typically exchanged during a joint task between two humans. The correct exchange of these signals, greatly enhances the quality of the interaction by revealing the partner's goals, intentions, desires (42; 43) or effort (18) and even discloses interactant's emotional status (44; 45). Actions can be performed gently, vigorously, or rudely, in general, as a function of the mood driving them (40; 46). These different ways of communicating have been called vitality forms by Daniel Stern and play a crucial role in social relations. Taking inspiration from these studies, we designed a behavioral experiment to investigate whether human participants could be influenced by a robot expressing rude and gentle vitality forms.

More specifically, participants were required to observe an humanoid robot (iCub) making a gestural or verbal requests to grasp a ball. All requests were expressed with rude and gentle vitality forms. After the iCub' requests, participants performed the required actions (grasping the ball bottle with the goal to give or place it). Our goals were to verify whether participants were influenced by the motion of the iCub in both brain activation and kinematic features.

In the following chapter 2 and 3 we will describe the two main studies in detail and in chapter 4 we will discuss the results, their impact, limitations and possible future developments.

Figure 1.1 *Understanding in human-human interaction is often mutual. Human-robot interaction should resemble it. Photo by Laura Taverna — Istituto Italiano di Tecnologia.*

Part II

Section Two

Chapter 2

Coordination

2.1 HandProfiler: a biological motion module for the iCub

There are many features that allow us to coordinate properly with a partner during an interaction, the "way we move" is for sure one of these and has been deeply investigated. One of the most famous rules that regulate our human motion, especially the arm movements, is the Two-Thirds Power Law (47; 48). This is a well-known feature, that relates the speed and curvature of a continuous movement. In particular, when the radius of curvature is high, our motion is slower while on the contrary when the radius of curvature is low, we have the tendency to speed up. We designed a module that enables the robot to execute curve movements compliant with this law (49), leveraging on the existing Cartesian controller of the iCub (50). Given a specific trajectory in 3D space, the module can convert it to a smooth human-like movement, which is then executed by the robot through the original Cartesian controller. Particular care was devoted to the generation of biologically plausible motion for the robot since, as it has been previously demonstrated, such motion is crucial in eliciting motor resonance and automatic imitation, two key features of in human-human interaction (23; 51).

The module takes as input some parameters to build the trajectory and establish a speed of execution. These parameters are:

- A, B, C, O : four points of the 3D space that are used to shape the elliptical trajectory.
- θ : an angle to set the range of motion across the trajectory.

- G : a value to set the speed gain of the movement.

Through the first 4 parameters, an ellipse is generated. This represents the trajectory that the end-effector of the robot will follow. The θ angle parameter is crucial to decide the boundaries of the movement, at which points of the trajectory the end-effector will start and stop. The elliptical trajectory is generated as a collection of points that are reached with the same timing and the value of gain G is exploited to calculate the distance between two consecutive points, keeping into account the curvature: the lower the curvature, the closer the points (see Figure 2.1).

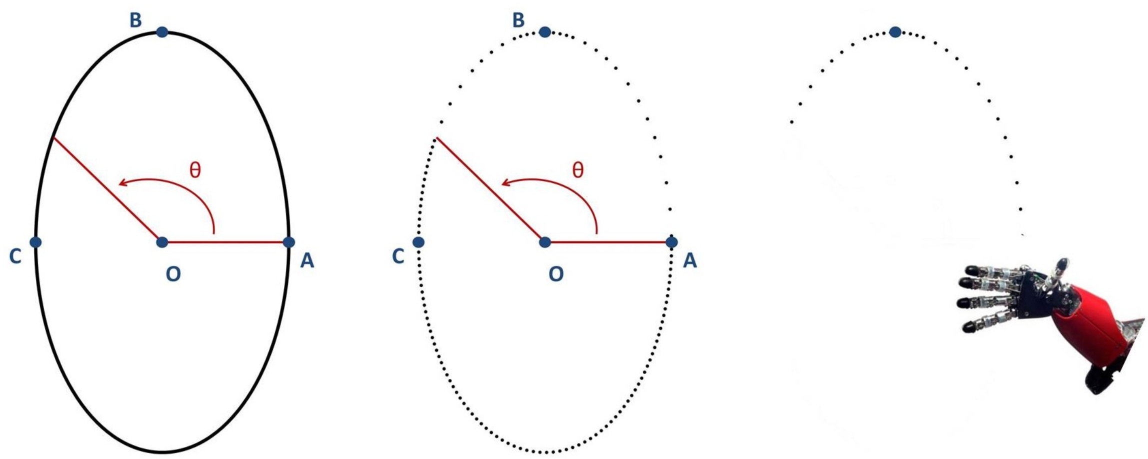

Figure 2.1 *Phases of the trajectory generation performed by the HandProfiler module.*

Since the final motion is achieved through the original cartesian controller, the robot reaches the initial position for the movement in the most efficient way with respect to the previous position of the arm. This causes some variability on the initial configuration of the joints, sometimes resulting in a strange initial posture. To avoid this undesired situation, we implemented the possibility to "play" movements saved in text files. Such files contain the value of all the arm joints with a precise timestamp and can be generated externally e.g. with Matlab, or "recorded" directly with our module, during the execution of a movement.

2.2 Experiment in Italy

The HandProfiler module was used in an experiment to understand whether the temporal adaptation usually observed during human-human interaction occurs also during human-robot cooperation. Mutual synchronization plays a decisive role in effective collaborations in human joint tasks. Interaction between humans and robots need to show similar emergent coordination. To this aim models of human synchronization have recently been ported on collaborative robots with success (52). However, it is also important to consider under which conditions the human partner is willing to adapt to the robot while performing a joint task. We designed a collaboration engaging both human participants and the humanoid robot iCub in pursuing an identical common goal: putting blocks into a box. We examined human action speed, evinced from motion capture data, in order to investigate whether humans adapt their behavior to the robot.

2.2.1 Methods

Experimental Design

We designed an experiment in which our participants had to perform a joint task with the humanoid robot iCub (53). Human and robot were sitting face to face and their goal was to fill a box with Lego blocks (see Figure 2.2). Each trial started when an experimenter put simultaneously one block in the open hand of the robot and another in the open hand of the participant. The following blocks were passed only when both had dropped the previous one into the box, and after they both put their hand in the initial position. Thanks to this, the robot and human always started at the same moment. The experimenter explained, at the beginning of the experiment, that the robot and the participant both had the same goal of filling the box that was closer to them with the Lego blocks that he would have provided them. He indicated that that they could only get a new block after both had dropped the previous one. No instructions were provided regarding the synchronization with the robot. The robot was pre-programmed to transport the block on its open palm, drop it into the box, and then immediately go back to the initial position. Each participant performed 6 sessions of 10 repetitions, i.e. they transported 10 blocks into the box for each session. During different sessions two factors were manipulated: Robot Speed, which could be slow, medium or fast and Box Number, which could be one – corresponding to a shared target space, or two – corresponding to individual space since both the participant and the robot had a smaller personal box in front of them. The order of conditions was randomized and unique for every participant. We set as "fast" a speed that we considered reasonable to complete the task after a few pilot tests, then we selected the "mid" and "slow" speeds by trying to maximize the differences among conditions, without making the "slow" motion seem unnatural to the participant. The average speeds in all conditions are reported in Table 2.1. The motions of the robot were inspired by biological human-like movements, as detailed in section 2.1.

Table 2.1 *Robot hand speed in the different experimental conditions: average and standard deviation.*

Condition	Mean Speed (m/s)	SD
Slow	0.100	0.008
Mid	0.125	0.007
Fast	0.154	0.009

Participants & Data collection

The experiment was performed in Italy by 17 participants. We excluded 2 participants for technical problems with data acquisition. Participants had different working backgrounds, from university students, to lab technicians or administrative staff. The regional ethics committees approved the protocol and all participants gave informed consent before participating (Mean age 30 years ± 5 SD, 6 males, 9 females, 1 Left handed, 14 right handed.)

For each participant, we acquired video recordings and kinematic data using a motion capture system. Videos were recorded from two different points of view in order to monitor participants' behavior in detail. Motion capture data, including 3D trajectory and speed of the hand and arm for each time frame (100 Hz), were gathered using four markers: three on the hand as shown in Figure 2.3, and one on the elbow. The motion capture system is a VICON System of infrared cameras, capturing at a 100 Hz rate. Data about the motion of the robot were gathered with the same motion capture system using 4 markers (one on the elbow) as shown in Figure 2.3. Our analysis focused mainly on the speed of the two agents. From the kinematic data, we extracted the starting and ending moment of each action. We considered the former as the last minimum of the speed when the hand is in the start zone, while the latter is the first minimum of the speed when the hand is in the box zone (see Fig. 2.2). Using these two time landmarks, we analyzed the most relevant part of the action and calculated the average speed for each repetition.

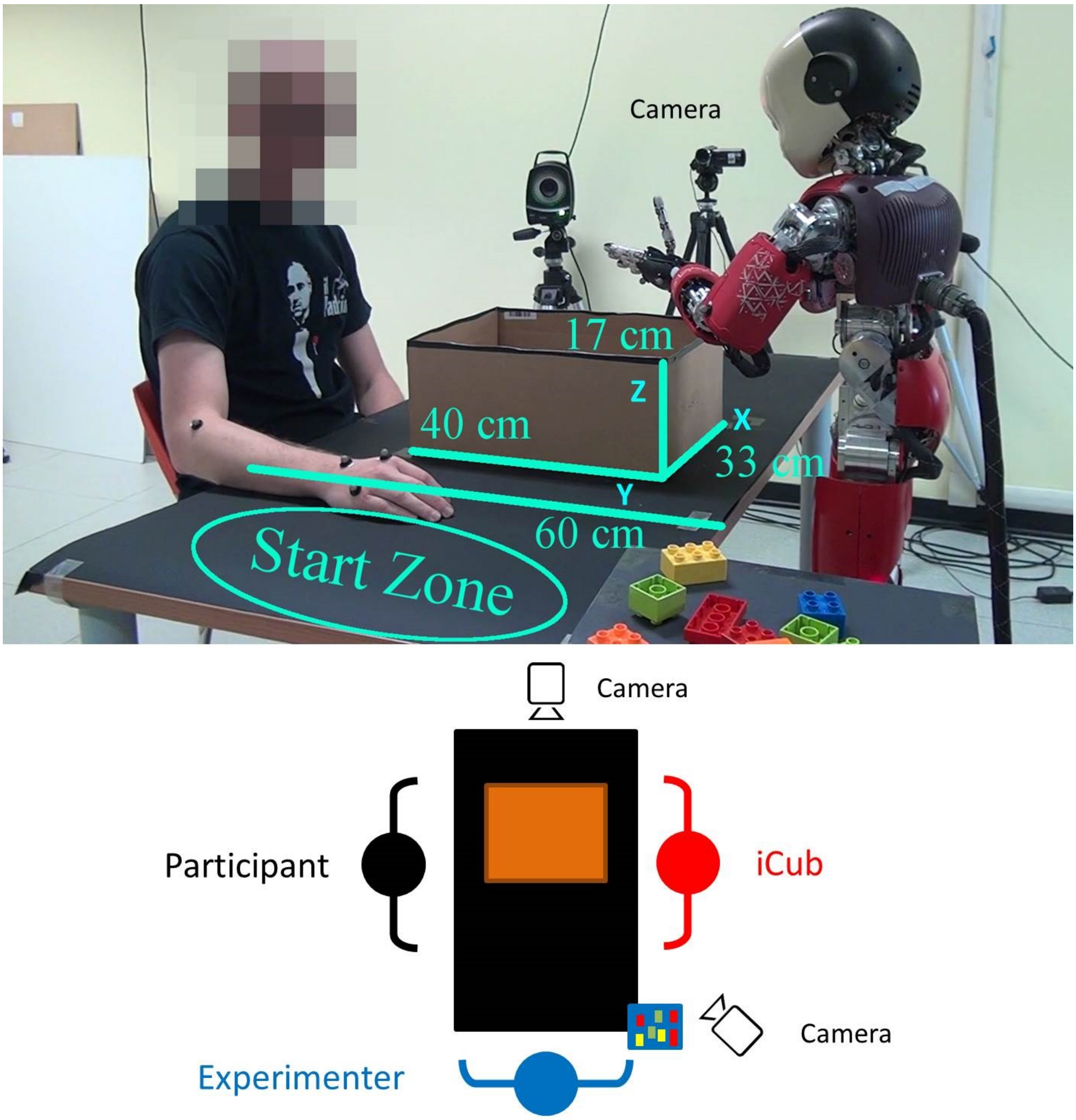

Figure 2.2 *Setup of the experiment. Blocks are passed by an experimenter in the start zone to both the participant and the robot at the same time. Their goal is to drop these blocks in the box. The picture is a snapshot of the "one box" condition.*

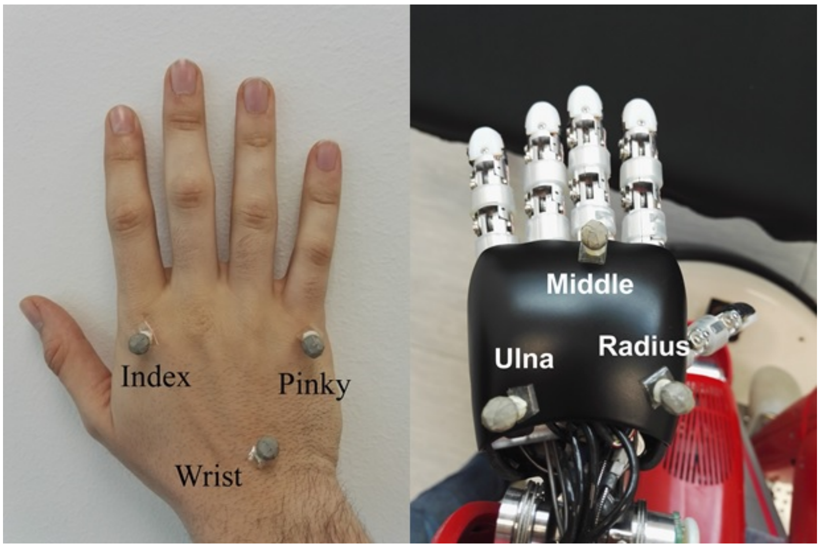

Figure 2.3 *Motion capture markers. Positioning is different between participant and robot in order to distinguish them better during the data analysis phase.*

2.2.2 Results

The aim of this study was the analysis of adaptation to a humanoid robot in a collaborative joint task when both agents were tasked to put Lego blocks into a box. In Figure 2.4 there is an example of the trajectories of both the robot (right) and a representative participant (left) in the "Robot Fast" and "Robot Slow" conditions. The path followed by the robot is very precise and only few variations can be noticed, while the trajectory of the participant is much more heterogeneous (see Figure 2.4). From visual inspection of the participants transport trajectories, it emerged that during box reaching they were quite similar across different conditions. During the return to the starting zone, a lot of variability emerged within and between participants since they did not receive any instruction regarding this phase, to keep the experiment as natural as possible. Participants chose either to stop near the box or at a different position before getting back to the start zone for the next trial (see blue lines in Figure 2.4 for an example). The first part of the action, from the moment the experimenter gives the block to the moment that the block is dropped into the box, is the only phase of the movement which presented a constraint: the robot and the participant started their action at the same moment. We consider this as the most interesting part on which to focus our subsequent speed analysis since it is the phase in which the robot and the participant perform the action at the same time. In the following sections we introduce the results extracted from the analysis of the experiment.

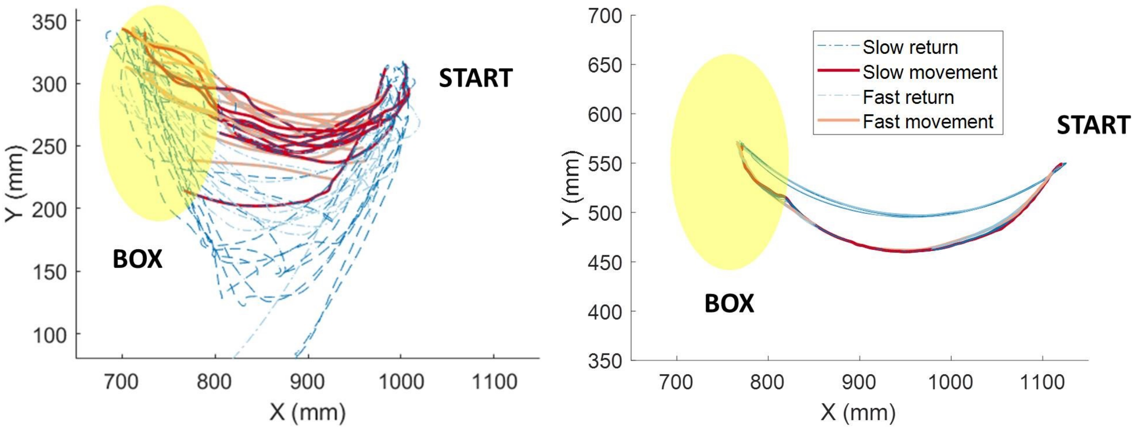

Figure 2.4 *Trajectories of one representative participant (left) and the robot (right) during the "Robot Fast" and "Robot Slow" conditions. X and Y axis are a projection of the 3D space in the frame of reference of the motion capture system.*

Figure 2.5 shows the mean speed of all participants, for each of the different conditions. It is clear that, on average, participants' motion was significantly faster than the motion of the robot (two sample t-tests between participants' and robot velocity in the corresponding "robot speed" conditions, all p's<0.05). From this chart also a form of adaptation can be noticed: even if participants were faster than the robot, their movement speed varied according to the three different speeds of the robot. Conversely, performing the task with one shared box or two different ones does not seem to trigger a different behavior in the participants. A two-way repeated measures ANOVA with Greenhouse-Geisser correction, on participants' speed with factor "Robot speed" (three levels: Slow, Medium, Fast) and factor "Number of Boxes" (two levels: 1 shared box, 2 separate boxes) followed by Tukey post hoc tests shows a significant change in participants' movement speed as a function of robot velocity ($F(1.95, 27.25)=17.43$, $p<0.01$), in particular between the "Robot Fast" and "Robot Slow" conditions (post hoc Tukey test: $t(28)=5.79$, $p<0.01$), and between the "Robot Mid" and "Robot Slow" conditions (post hoc Tukey test: $t(28)=3.6$, $p<0.05$). No significant effect is evidenced as a function of the number of boxes $F(1,14)=0.23$, $p=0.6$).

After the analysis of the mean from a general point of view, we looked for a possible effect of adaptation during the multiple repetitions. The panels of Figure 2.6 represent the mean across all participants, for each of the ten actions. We could not distinguish any form of adaptation with the progression of the repetitions, not even any particular trend. To statistically verify this observation, we ran two Two-Way repeated measures ANOVA with Greenhouse-Geisser correction, with factors "Robot Speed" (3 levels) and "Repetition number" (10 levels) on the "One box" and "Two boxes" conditions respectively. The analysis

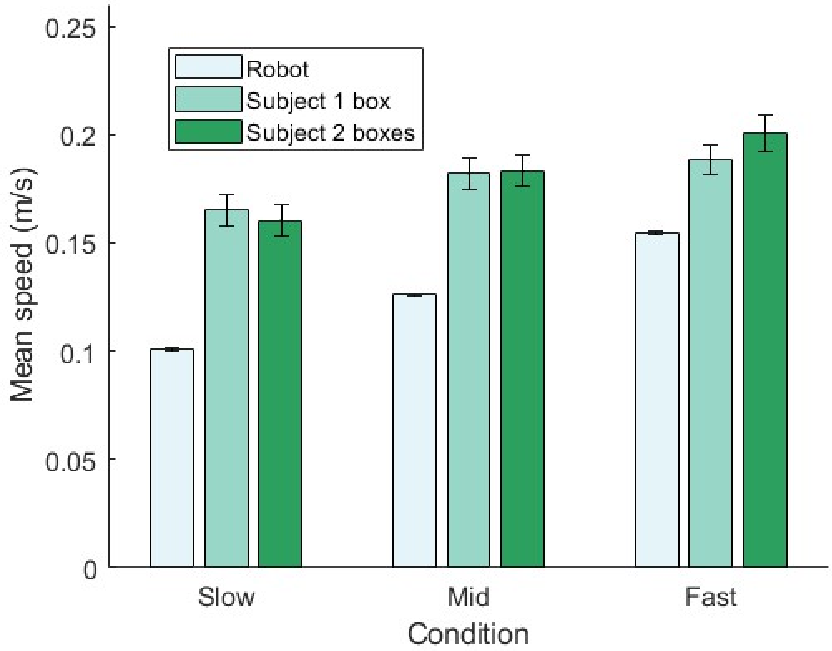

Figure 2.5 *Mean speed, and standard error (error bars) of the robot and participants for all conditions.*

showed that in neither case there was a significant effect of the interaction (F(5.32)=1.17, p=0.33 and F(6.46)=1.04, p=0.41). Moreover, there was no significant difference among the repetitions for the “One box” case (F(3.71)=2.16, 5 p=0.91), whereas in the “Two boxes” case a slight reduction in speed was observed between the beginning and the end of the task (F(4.16)=3.9, p<0.01). In particular a post hoc Tukey test highlighted a significant difference only between the first and four last trials (all p<0.05).

Further evidence of an automatic adaptation to the robot’s velocity comes from the analysis of individual participants’ behaviors. The graphs in Figure 2.7 show that almost all participants lie above the dashed identity line, meaning that their speed in the “Robot Fast” and “Robot Mid” conditions was higher than their speed in the “Robot Slow” condition. If a participant maintained the exact same speed, for example, in the “Robot Fast” and ”Robot Slow” condition, the corresponding marker would be on the dashed line. Similar results derive from the analysis of the “Two Boxes” condition.

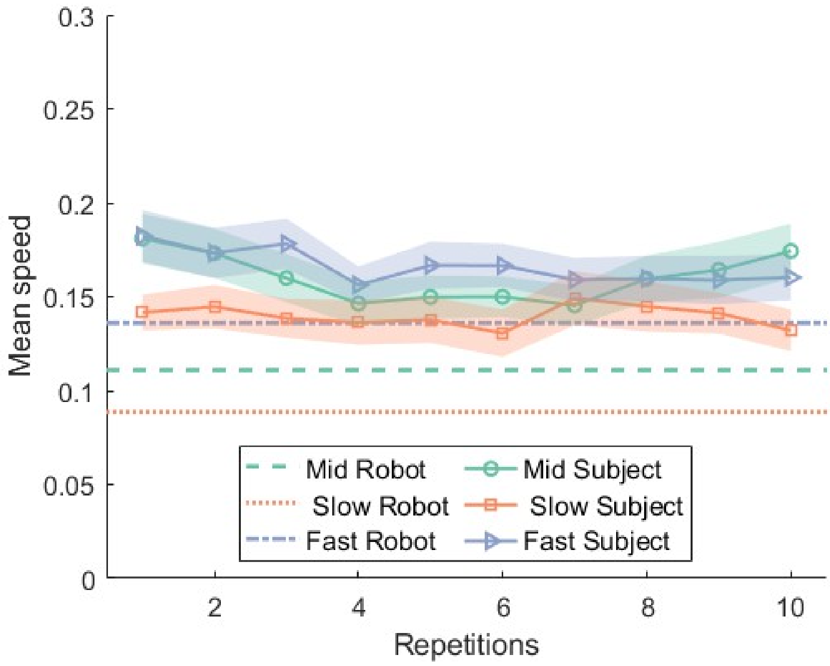

Figure 2.6 *Mean participants' speed for each of the ten repetitions ("One Box"), with shaded areas representing standard error of the mean. The dashed lines represent the speeds of the robot, equal across all the ten repetitions.*

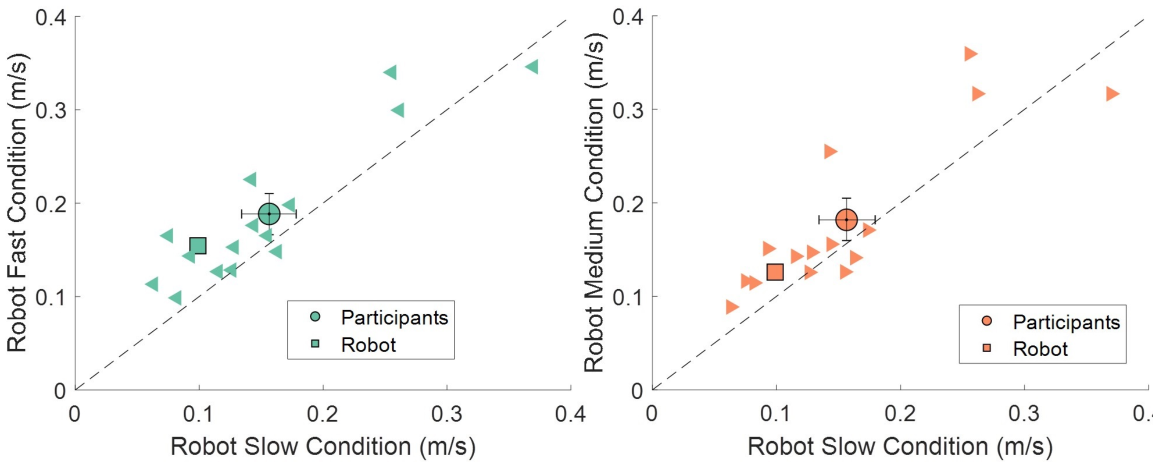

Figure 2.7 *Individual speed of participant in the "Robot Fast" (left) and "Robot Mid" conditions (right), in relation to their speed in the "Robot Slow" condition in the "One Box" sessions. The bigger circles with error bars represent the sample mean and standard error.*

2.3 Experiment in Japan and comparison

We then moved on to investigate whether the automatic imitation of the robot is affected by cultural differences. In particular we focused on the comparison between Italy and Japan. Indeed, converging evidence shows that top down modulation of the activation of the neural substrates possibly supports emergent coordination. Amoruso et al. (54; 55) have for instance demonstrated that motor resonance is not an entirely automatic process, but it can be modulated by high-level contextual representations. Also social aspects might affect motor resonance. Recent neuroimaging studies show that mirror system activation is modulated by social group membership, with higher activation during action observation when the action is performed by an in-group rather than an out-group member (56; 57). However, it is not yet known whether cultural differences might have an influence on emergent coordination, leading to different patterns of behavior across countries, also when the joint action is performed with a partner of the same culture. If cultural differences impacted emergent coordination, we would expect to find differences in the degree of adaptation between the Italian and Japanese population. Japanese people showed a higher general acceptance toward robots when compared with Western cultures (32). Since acceptance, or better affiliation, is tightly connected to imitation and mimicry(33), we hypothesized a higher degree of speed imitation for the Japanese sample, when compared to the Italian one. Moreover, given the difference in the perception of personal space between the two cultures (58), we hypothesized the Japanese to move differently to avoid violating the personal space of the robot in the "shared box" condition.

2.3.1 Methods

Participants & Data

The experiment was performed in Japan, in collaboration with Osaka University, by 9 participants with different working backgrounds. We excluded 1 participants for technical problems with data acquisition. The regional ethics committees approved the protocol and all participant gave informed consent before participating (Mean age 29 years ± 9 SD, 3 males, 5 females, all right-handed.)

We collected the same data for Japanese and Italian participants. Motion capture data, including 3D trajectory and speed of the hand and arm for each time frame (200 Hz), were gathered using four markers: three on the hand as shown in Figure 2.3, and one on the elbow. The motion capture system is a Motion Analysis MAC3D capturing at a 200 Hz rate. Data about the motion of the robot were gathered in a different way in Japan. We took advantage of "yarpdatadumper", a module created to record and save in files different kind of information from the robot (59). We took advantage of this to record the values of the joints of the robot arm at specific time instants. We then transformed the joints data into 3-dimensional trajectories and speeds with the specific tools supplied by the designers of the robot (60). For the analysis we extracted the same data and followed the same pattern as for Italian participants.

The average speeds in all conditions are reported in Table 2.2. It is slightly different between the two countries. The reason for this discrepancy is that two different iCub robots were used and the movement speed is influenced by different factors, out of which some were not under our control, including the age of the robot and low-level settings of the electronics. However the values are comparable between the two experiments and the relative variation in robot's speed between the fast and slow conditions was very sim ilar (54% velocity increase vs. 57% velocity increase).

Table 2.2 *Robot hand speed in the different experimental conditions: average and standard deviation (Japan).*

Condition	Mean Speed (m/s)	SD
Slow	0.080	0.008
Mid	0.108	0.004
Fast	0.126	0.006

2.3.2 Results

In Japan we performed the same analysis that we did in Italy for a direct comparison between the two countries. From the bar chart in Figure 2.8, we can see that participant in Japan are always faster than the robot and that they tend to adapt to the different robot speeds. Concerning the 'shared space' condition there seem to be no differences across the "number of boxes" variable, for all the three speeds of the robot. These observations are confirmed by a two-way repeated measures ANOVA with Greenhouse-Geisser correction, on participants speed with factors "Robot speed" and "Number of boxes", which shows a significant change in participants' speed as a function of robot movement velocity $F(1.68, 11.74)=7.74$, $p<0.01$) and no change as a function of number of boxes $F(1,7)=2.73$, $p=0.14$. Even though there is no significant difference between the numer of boxes, a trend seem to appear since the speed is slower in the condition with two boxes, but this result needs further investigation.

Similarly to Italian participants, there is a significant difference for Japanese participants between the “Robot Fast” and “Robot Slow” conditions (post hoc Tukey test: t(14)=4.99, p<0.01), and between the “Robot Mid” and “Robot Slow” conditions (post hoc Tukey test: t(14)=4.55, p<0.05).

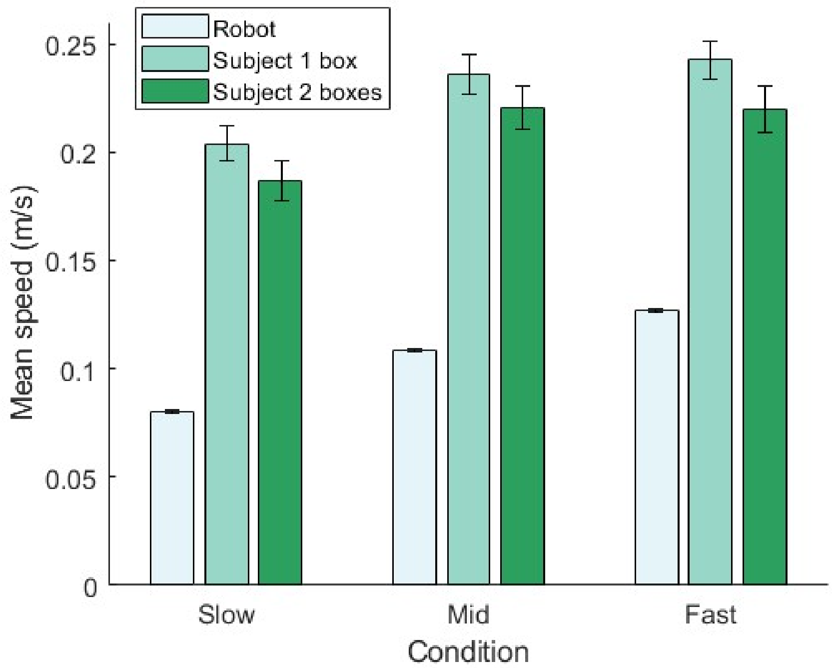

Figure 2.8 *Mean speed, and standard error (error bars) of the robot and participants for all conditions (Japan).*

As shown in Figure 2.9 there is no clear effect of adaptation with the progress of repetitions. 2 Two-Way repeated measures ANOVA with Greenhouse-Geisser correction, with “Robot Speed” and “Repetition number” as factors, did not highlight any significant difference among repetitions for both the “One-box” (F(2.62)=2.7, p=0.08) and the “Two-boxes” (F(3.83)=1.39, p=0.26) conditions, nor any interaction (F(5.05)=1.05, p=0.4 and F(4.42)=1.09, p=0.38).

Finally, figure 2.10 illustrates another validation of the aforementioned results. Here each marker represents a single participant and since the majority of them are above the dashed identity line, this means that they changed their speed according to the change of the robot velocity, adapting to it.

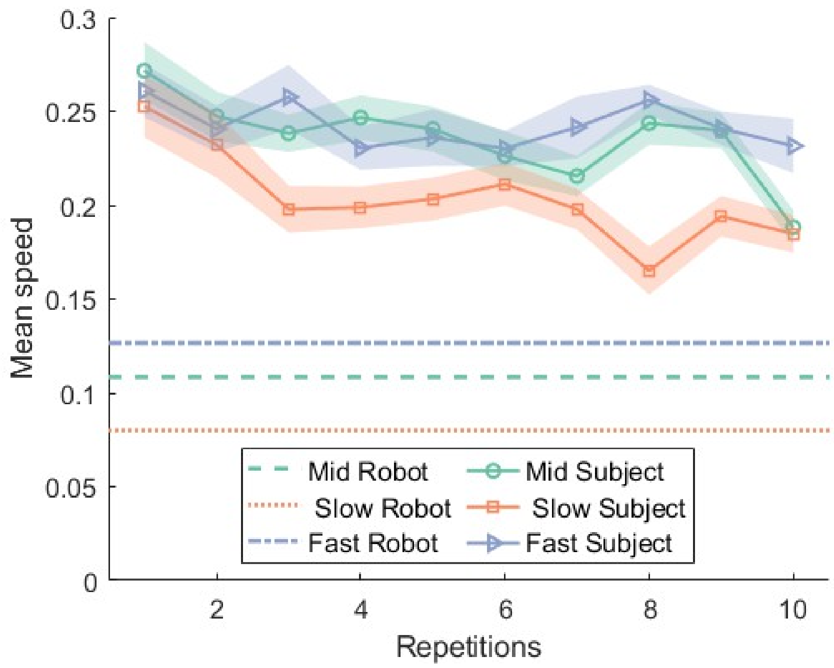

Figure 2.9 *Mean participants' speed for each of the ten repetitions ("One Box"), with shaded areas representing standard error of the mean. The dashed lines represent the speeds of the robot, equal across all the ten repetitions (Japan).*

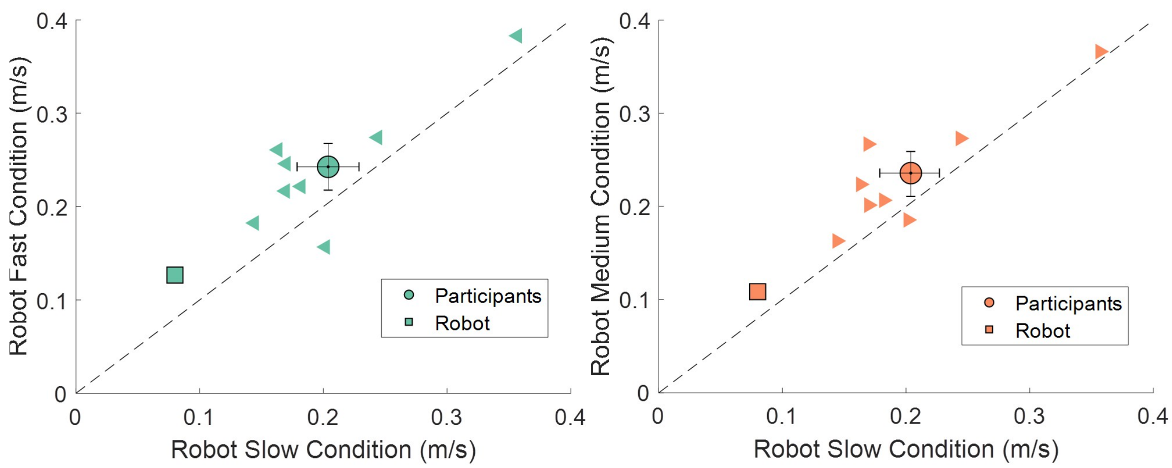

Figure 2.10 *Individual speed of participants in the "Robot Fast" (left) and "Robot Mid" conditions (right), in relation to their speed in the "Robot Slow" condition in the "One Box" sessions. The bigger circles with error bars represent the sample mean and standard error (Japan).*

2.3.3 Cross-cultural comparison

From the previous sections, it appears that there is a high similarity in the behaviors of the participants from Italy and Japan during the interaction with the robot: both groups were influenced by robot's speed, with no modulation of their behavior due to the presence of a shared box or two individual target spaces. To directly compare the level of adaptation between the two countries, we plotted the individual speeds of all participants in the "Fast Robot" and "Slow Robot" conditions on the same graph (Figure 2.11). In the figure, it can be noticed that, on average, the Italian participants tend to be slower than the Japanese, even though the robot speed was slightly faster in the Italian experiment than in the Japanese one (see squares in Figure and Tables). Besides this effect, both the Italians and the Japanese tend to change their speed similarly, adapting to the changes of the robot velocity, as shown by the similar relative position of the average markers with respect to the identity line.

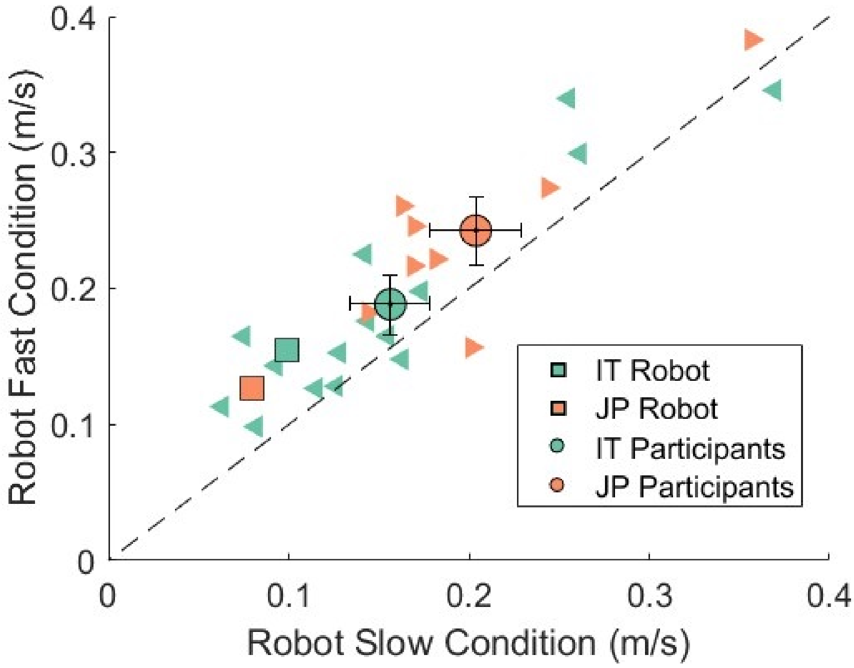

Figure 2.11 *Individual participants' speed in the "Robot Fast" speed condition, displayed in relation to the corresponding speed in the "Robot Slow" speed condition (in the "One Box" condition) for both countries. The bigger circles with error bars represent the mean and standard error. If a participant had the exact same speed in the two conditions, the respective marker would be on the dashed line.*

To quantify the degree of adaptation, and directly compare it between the two countries, we computed for each participant the relative variation of her speed as a function of the variation in the robot speed across conditions. To calculate this so called "slope", we regressed each participants' average speeds in the three speed conditions with respect to the corresponding robot's velocities and extracted the slope of the resulting line. A number close to one would correspond to a relative change in participants' speed comparable with that exhibited by the robot across conditions, implying a high level of adaptation. The stem plots in Figure 2.12 represent the computed individual slopes, which are similar between the two groups. Bars represent the "Mean return difference" (MRD), that is the mean difference in timing between the instants in which the robot and the participant returned their hand to the starting position to get a new block. A Mixed Two-Way repeated measures ANOVA with Greenhouse-Geisser correction, on participants slope with "Number of Boxes" as within factor and "Nationality" as between factor: shows that the adaptation was not significantly different between the two groups tested (F(1,21)=0.267, p=0.61), nor was affected by the presence of a shared target space (F(1,21)=0.005, p=0.94).

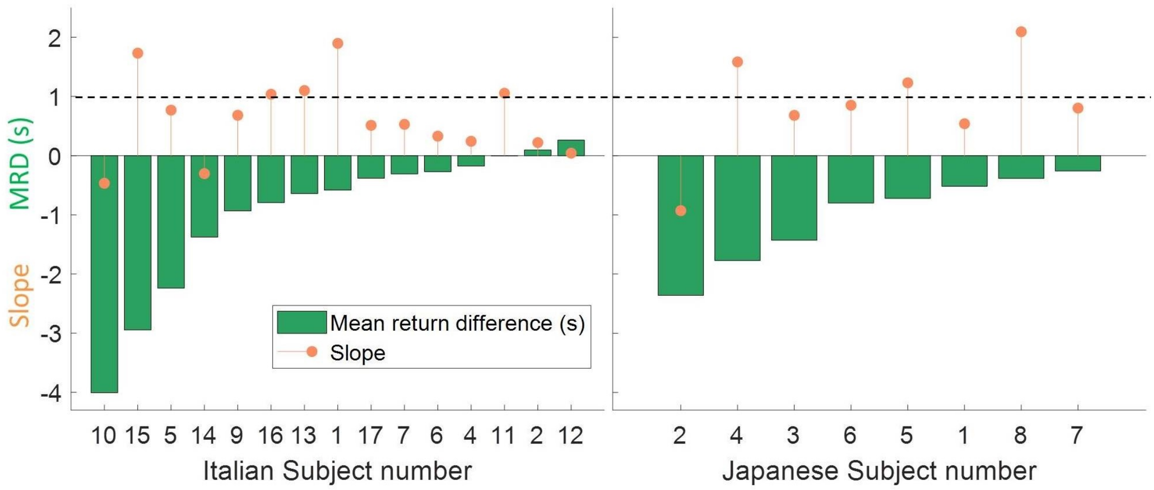

Figure 2.12 *Stem plots represent the slope corresponding to the amount of adaptation of participants. These slopes have been computed for the one-box condition, but those computed for the two-box condition are very similar. Bars represent the "Mean return difference" (MRD), that is the mean difference in timing between the instants in which the robot and the participant returned their hand to the starting position to get a new block. A negative value means that the participant arrived first.*

2.4 Limitations and future work

A possible confounding variable in the experiment was that by design we established the restriction to start each repetition together with the robot. Participants might, therefore, have been induced to slow down when the robot slowed down, in order to synchronize their arrival to the start zone with that of iCub. In other words, they might have planned their actions with the covert aim of coordinating with the starting action of the robot. If this assumption is correct, we might expect that people who showed a stronger adaptation to the robot speed (slope closer to one) were also arriving at the starting zone in synchrony with the robot. We therefore computed the “mean return difference” as the difference in timing between the instants in which the robot and the participant returned their hand to the starting position to get a new block. A correlation between high slopes (close to one) and small mean return difference would mean that they slowed down intentionally. The computation of slope and mean return difference showed no connection between the two measures, as can be seen in Figure 2.12. The distribution of the stems, representing the adaptation (slope), is not clearly related to the bars showing the difference in return timing. It seems therefore unlikely that the common start of repetition timing during the trials influenced the adaptation to the robot. The presence of a third agent with the active role of passing blocks could have influenced the interaction, even though the experimenters were trained to release the objects in the hand of the participant and robot in a stereotyped manner which was constant across all conditions. Moreover participants witnessed only the release of blocks into their hands since the experimenter picked up the blocks while participants were completing the previous trial looking at the robot or at the box. For this reason we believe it’s unlikely that the speed of the experimenter had influenced the participants. Future experiments will avoid the presence of another person that could bias the results. Another problem to discuss is the number of participants. Availability in Japan was limited thus we could not have the same amount of participants in the two countries. Overall, testing a higher amount of people in the future would lead to more reliable results and allow a more detailed data analysis, possibly strengthening our findings and improving the comparison between the two countries. It is important to note that the selected task might not be considered a proper joint action, such that the two agents overtly collaborate to achieve a shared goal. However the two have to perform a simultaneous action in a quasi-shared space and this entails a certain degree of coordination. It could be relevant in the future to assess whether a more explicitly collaborative task could be more affected by cultural difference. Last, in this experiment we only measured the implicit adaptation and we did not enquire about the subjective perception of the participants. Such information could help interpreting more in depth the behavioral results.

Chapter 3

Vitality

3.1 Inspiration

In the previous section we presented a study investigating movement-based implicit communication with a focus on coordination timing, but during human-human interactions we exchange many different signals including indications on our current affective status and attitude towards other agents. Actions can be performed in different ways, for example, a hand shake can be kind or vigorous, a caress can be delicate or rushed communicating the positive or negative attitude of the agent. Similarly, words can be pronounced with a pleasant or unpleasant tone also conveying the agent's attitude. These different forms of communication have been named "vitality forms" by Daniel Stern (2010). Vitality forms have a strong influence on interaction between humans. The same gesture has different consequences when it is performed aggressively or kindly, and humans are very sensitive to these subtle differences in others' behaviors. The aim of this work was twofold: 1) endow a humanoid robot with vitality forms allowing it to generate gentle and rude actions; 2) investigate whether and how the observation of these actions influence the perception and motor behaviour of the human partner. To address these issues, using fMRI and kinematic techniques, we carried out two different studies. Drawing inspiration from modulations of human voice and motion associated to different vitality forms, we modified a passing action and a passing voice command performed by the robot to convey an aggressive or kind attitude. In a series of experiments we demonstrated that the humanoid was consistently perceived as aggressive or kind. The neural responses to the observation of the robot actions matched those recorded for the observation of human actors exhibiting similar vitality forms. Importantly, this happened most likely because the appropriate features of human behavior were ported on the robotic platform. Also human behavior changed in response to the different robot attitudes with human partners showing more aggressive kinematics when interacting with an aggressive robot.

The opportunity of humanoid behavior to express vitality enriches the array of nonverbal communication that can be exploited by robots to foster seamless interaction. Such behavior might be crucial in emergency and in authoritative situations in which the robot should instinctively be perceived as assertive and in charge, as in case of police robots or teachers.

3.2 fMRI studies

A fascinating possibility is that the new generations of robots will be endowed with the capacity to express vitality forms appearing more authentic and comfortable to interact with. In this view, an interesting question is to investigate whether the observation of actions generated by a humanoid robot (iCub) with a gentle or rude vitality form may elicit the activation of the dorso-central insula found activated in previous studies during the observation of human actions (27). For this purpose, two fMRI studies were carried out. The aim of the first study was to assess whether the observation of the iCub robot mimicking human vitality forms produced the activation of the insula. In this study, sixteen participants were required to pay attention either on video-clips showing a human actor offering an object in a gentle and rude way or on video-clips showing very similar actions generated by the iCub robot with a low and high velocity. Results of this first study showed that only the observation of human actions produced the activation of the dorso-central insula. In the second study, we tried to improve the stimuli to have a better comparison between the robot and the human. We chose an actress whose anthropometric measures were more similar to those of the robotic platform iCub to ensure a higher compatibility between the robot and the human stimuli. We asked her to perform an offering gesture gently or rudely and its kinematic parameters (i.e. peak velocity, length trajectory) were recorded by the Optotrack motion capture system and remapped into the kinematic model of the iCub robot. This allowed us to replicate her actions with the same vitality forms. Then, we assessed whether the observation of these robotic actions generated with human kinematics produced the activation of the dorso-central insula. The results of this second study showed that, the observation of robotic actions endowed with human vitality forms produced a BOLD signal increase in the dorso-central insula. This insular activity was very similar to that obtained during the observation of human actions.

3.3 First fMRI study

In contrast with previous fMRI studies which investigated the goal understanding of robotic actions (24; 61; 62) (what), the present study is focused on a new fundamental action property: the form (how). For this purpose one fMRI experiment was carried out to assess whether the observation of the iCub robot mimicking human vitality forms produced the activation of the insula. In this study, sixteen participants were required to pay attention either on video-clips showing a human actor offering an object in a gentle and rude way or on video-clips showing very similar actions generated by the iCub robot with a low and high velocity. Results of this first study showed that only the observation of human actions produced the activation of the dorso-central insula.

3.3.1 Methods

Participants & Data collection

Sixteen healthy right-handed participants (Mean age = 26.5 years ± 3.8 SD, 6 females, 10 males) took part in this experiment. All participants had normal or corrected-to-normal vision. Informed consent was obtained from all participants and the experiment was approved by the ethics committee of the University of Parma (UNIPRMR750v1) in accordance with the Declaration of Helsinki.

Imaging data were collected on a 3 Tesla Discovery MR750 GE scanner equipped with an eight-channel receiver head coil. Functional images were acquired using a gradient EPI sequence with a TR of 2500 ms, TE of 30 ms, flip angle of 90°, parallel imaging acceleration factor of 2, 205 x 205 mm^2 field of view, voxel size of 2.5 x 2.5 x 3 mm^3. The scanning sequence comprised 229 ascending sequential volumes composed by 37 axially slices. Additionally, a high resolution T1-weighted structural image (1 x 1 x 1 mm^3) was acquired with a TR of 8100 ms, TE of 3.2 ms, flip angle of 12° for each participant.

Generation of the stimuli

VICON Motion Capture System (Vicon OMG, UK MX2 model, sampling frequency: 100 Hz.) was used to record the kinematic features of actions performed by both the male actor and the robot. In particular, six infrared cameras recorded the 3D position at regular time intervals of a 5 markers placed in different positions of the right arm: one on the thumb, one on the index finger, one on the wrist, one on the elbow and one on the shoulder. In all videos, the actor and the robot performed the actions starting from the same initial position and

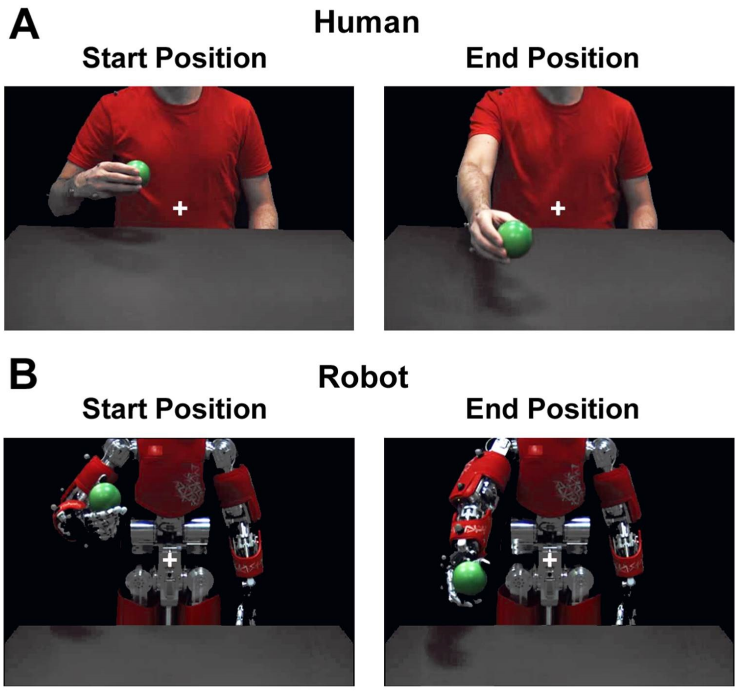

Figure 3.1 *Example of the video-clips observed by participants during the experiment. The images show the human actor and the iCub robot in the start (AB, left panels) and in the end positions (AB, right panels).*

reaching the same final position (see Figure 3.1). During the action execution, the natural and ecological expression of human actor was preserved as much as possible avoiding excessive artificial manipulation of kinematic variables. All actions performed were also filmed by a high definition camera (Panasonic HCX 900) in order to obtain video stimuli to present in the fMRI study. The movements of the iCub were generated with the module described in section 2.1, trying to match the speed and trajectory of the human actor for the rude and gentle vitality forms. The use of this module granted that the movements generated complied with the properties of biological motion and in particular to the Two-Thirds power law. However, with this approach, the distinction between the two vitality forms was limited to a difference in peak velocity, while the acceleration profile remained unchanged. The module also had some limitations that didn't allow to reach the same speed of the human actor.

After video recording, using the Tracker software, the 2D kinematic profiles of all actions presented in the video-clips were analyzed. For this purpose, a specific point of the agent's hand (human or robot) corresponding to the index finger, was marked for all video clips. In particular, for each video frame, the position of the index finger was tracked in the space from the beginning to the end of action. For all actions, using both X and Y values of each tracked point, the module of the velocity of gentle and rude actions was calculated and averaged. Figure 3.2 shows the kinematic profiles of human and robotic actions presented to the participants in this study (perceived velocity). Due to the limitations of the module used to generate robotic actions, we had to speed up the frame-rate of the iCub videos to reach a perceived velocity that was closer to the one of the human actor.

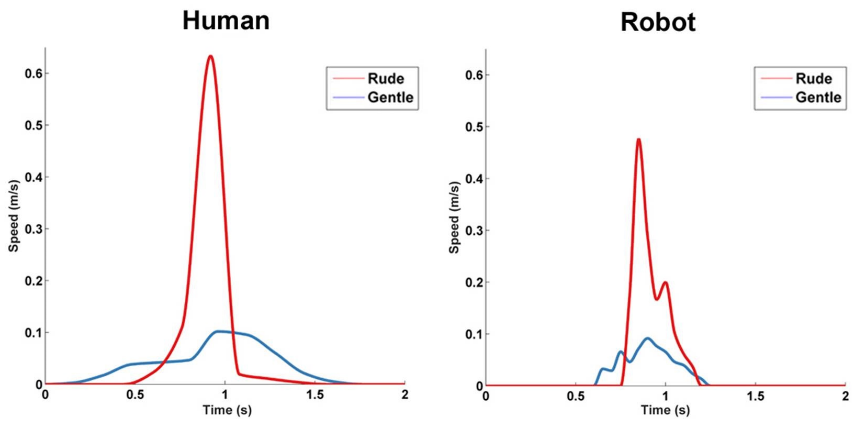

Figure 3.2 *Physical properties of the video stimuli. Graphs depict the profiles of the mean 2-D velocity perceived by participants during the observation of gentle and rude actions performed by human actors and iCub robot.*

Experimental design

Participants were presented with video-clips showing a male actor offering different objects (apple, orange, packet of crackers, bottle, ball) with a gentle and rude vitality forms or without any vitality form (jerky actions; control condition). More specifically, the control stimuli were obtained by presenting one static frame of the action every 400 ms (5 frames in total from the beginning to the end of the action). The aim of the control stimuli was to allow participants to understand the action goal without conveying any vitality form information. Additional video-clips showed a humanoid robot (iCub) generating the same actions with two velocities (high and low) mimicking the gentle and rude vitality forms expressed by the human actor. As a control, participants were presented with video-clips showing the iCub robot generating the same actions with a jerky movement. Subsequently, in order to obtain a very similar peak velocity between the kinematic curves of actions performed by the human actor and those generated by the iCub robot, the frame rate of video-clips was adjusted accordingly (Figure 3.2). Note that, for all video-clips, the face area of both the human actor and the iCub robot was not shown to avoid possible confound effects due to attention to the facial expression.

Participants laid in the scanner in a dimly lit environment. The stimuli were presented via digital video system (VisuaSTIM) with 30 dB noise-attenuating headset with 40Hz to 40 kHz frequency response and with a 500.000 px x 0.25 square inch resolution with horizontal eye field of 30°. The digital transmission of the signal to the scanner was via optic fiber. The software E-Prime 2 Professional (Psychology Software Tools, Inc., Pittsburgh, USA, http://www.pstnet.com) was used both for stimulus presentation and the recording of participants' answers. Video stimuli were presented in blocks of five consecutive stimuli of the same condition each lasting 2.5 s (conditions: human gentle actions, human rude actions; human control actions; robot gentle actions; robot rude actions; robot control actions). An inter block period of 12.5 s without video stimuli was present between two consecutive blocks. In the 10% of cases, in the inter block period catch trials were randomly presented and participants were required to indicate the correct vitality form observed by pressing a button on a response box placed inside the scanner. The study 1 was composed of 3 functional runs with a total of 9 blocks (45 single trials) for each condition, presented in a randomized order. Each functional run lasted about 10 min. During the presentation of the stimuli, participants were asked to fixate a white cross presented in the center of the screen and to focus on the way in which actions were performed.

3.3.2 Results

The observation of actions performed by human gently and rudely vs. baseline produced the activation of the occipital lobe, bilateral inferior parietal lobule extending to the supramarginal gyrus and bilateral premotor cortex extending to the inferior frontal gyrus. In addition, there was a bilateral activation of the insular cortex. A very similar activation pattern was observed for the robot condition with a large extension of the frontal areas and the lack of the activity of the insular cortex. In contrast, the observation of video-clips showing jerky actions performed by both human and iCub robot (control stimuli) produced a weaker activation of the same pattern (see Fig. 3.3).

The direct contrast between the observation of human actions performed with vitality forms and the control actions (jerky actions) revealed the activation of the left dorso-central insula (human rude actions vs. human control actions; human gentle actions vs. human control actions). On the contrary, the contrast between the observation of the robotic actions mimicking the human vitality forms (actions generated with low and high velocities) and the

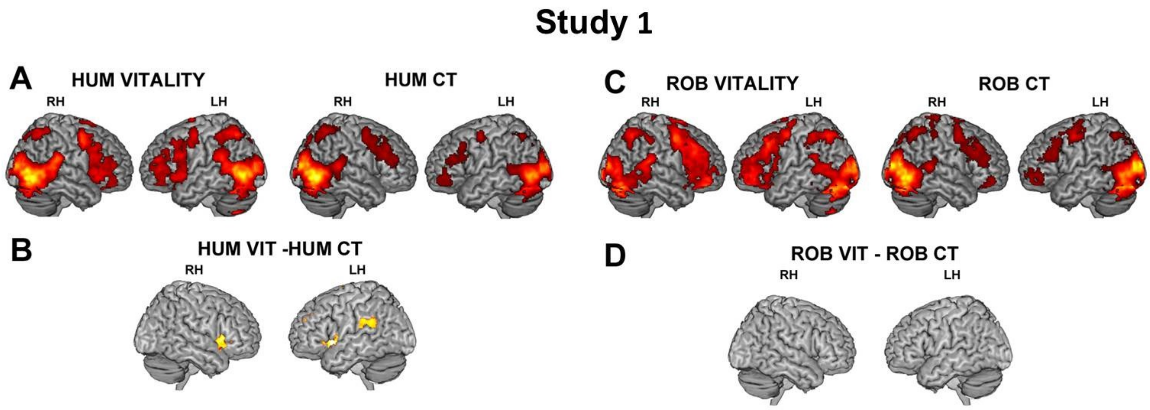

Figure 3.3 *Brain activation during the processing of actions performed by human actors and iCub robot. LH, left hemisphere; RH, right hemisphere.*

robotic control actions did not reveal any activation (robotic rude actions vs. robotic control actions; robotic gentle actions vs. robotic control actions).

The direct contrast between the observation of human actions and robotic actions [human (rude actions + gentle actions) – robot (rude actions + gentle actions)] revealed the activation of the left dorso-central insula. On the contrary, the opposite contrast did not produce any activation. It is important to note that the activation of the dorso-central insula was also found in the direct contrast between human gentle actions versus robot gentle actions. Considering the BOLD signal extracted, the t-test revealed significant difference between human and iCub robot during the observation of rude and gentle actions (Figure 3.4 $p < 0.05$).

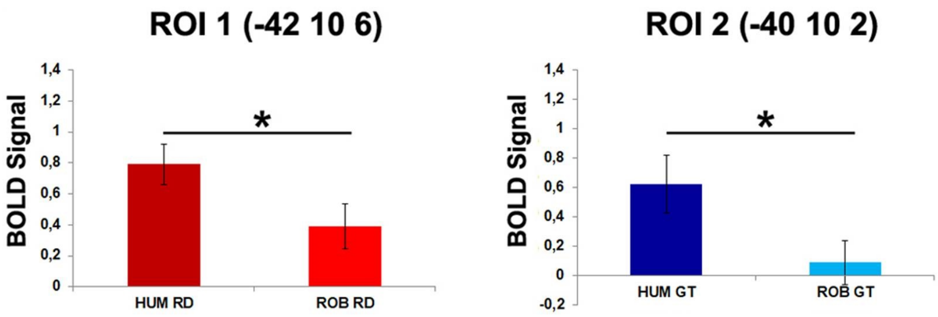

Figure 3.4 *BOLD signal indicating the insula activity during the processing of actions performed by human actors and iCub robot. The bar graphs indicate the comparisons between human and robot in gentle and rude conditions. Asterisk indicates p<0.05. (RD = rude, GT = gentle).*

3.4 Second fMRI study and subjective evaluation

In the first experiment described above, we showed that only the observation of human actions endowed with vitality forms produced the activation of the dorso-central insula. It was crucial then to understand whether the lack of activation in response to robotic stimuli was due to the nature of the agent (i.e. a robot) or to the difference of the subtle properties of the robotic movements with respect to those of the human. To address this question we tried to reduce the differences between the kinematics of human and robotic actions, and we carried out a second fMRI experiment.

The main improvements included:

- Putting a red t-shirt to the robot. The movement of the mechanical joints of the iCub shoulder are quite different from the motion of a human being. This could weaken the impression of biological plausibility of the robot movement. The red t-shirt masked the joint.
- Changing the human actor. We choose an actress whose size was more similar to the robot, especially for the length of the arm and consequently the amplitude of the movements. This facilitated the mapping of human motion to the robot.
- Avoiding manipulation of the videos. In the previous study, in order to generate the aggressive stimuli, we had to speed up the frame-rate of the videos since we could not reach the same speed of the actor only through the use of the module (see section 2.1. This could have interfered with the naturalness of the robot motion.

The experimental paradigm was kept equal to the previous study. In addition, we decided to also ask for some subjective evaluations of the video stimuli, to verify if the different vitality forms were perceived correctly by participants.

3.4.1 Methods

Participants & Data collection

Sixteen healthy right-handed participants, different from the ones of the first study (Mean age = 25.7 years ± 3.4 SD, 5 females, 11 males) took part in this experiment. For the subjective evaluation we had 20 participants (Mean age: 24, SD: 2 years, 5 males, 15 females). All participants had normal or corrected-to-normal vision. Informed consent was obtained from all participants and the experiment was approved by the ethics committee of the University of

Parma (UNIPRMR750v1) in accordance with the Declaration of Helsinki. For specifications of imaging data, see the previous section.

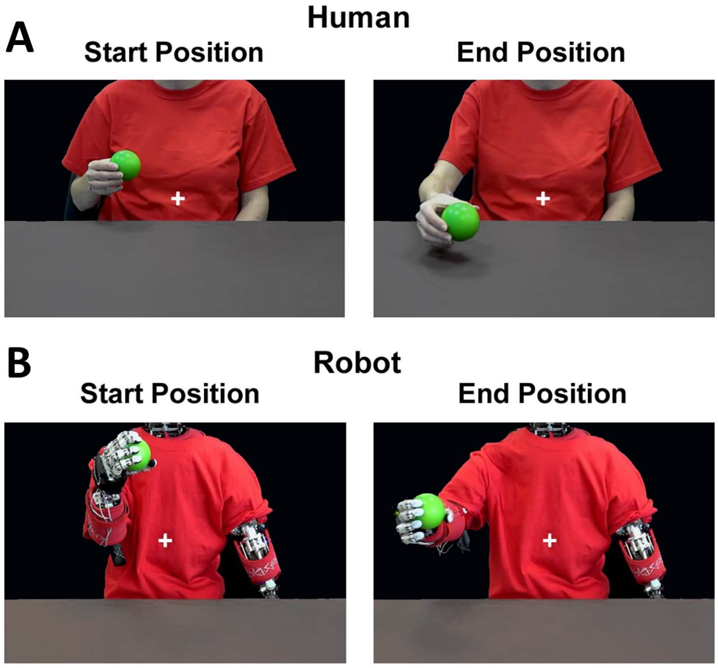

Figure 3.5 *Example of the video-clips observed by participants during the experiment. The images show the human actor and the iCub robot in the start (AB, left panels) and in the end positions (AB, right panels).*

Generation of the stimuli

In this study, we choose a female actress whose arm size was more similar to the one of the iCub. The kinematic features of the actions performed by both the actress and the robot were recorded from the same marker positions as described above, but using a different system: the NDI OptoTrack (sampling frequency: 200 Hz). The choice of this new recording system was supported by the presence of active markers. These specific markers helped to avoid possible confounding effects caused by the reflective metallic parts of the robot. This time, the kinematic features recorded from the actress were used to generate the movements of the robot. We took advantage of the captured data and translated it to the motion space of the robot using the HandProfiler module, generating movements with velocity and acceleration profiles that were closer to the ones of the actress, as can bee seen from figure 3.6. Thanks to this technique we did not have to modify the video stimuli since the aggressive vitality form was directly embedded in the movements of the iCub, making the motion appear more natural and biologically plausible.

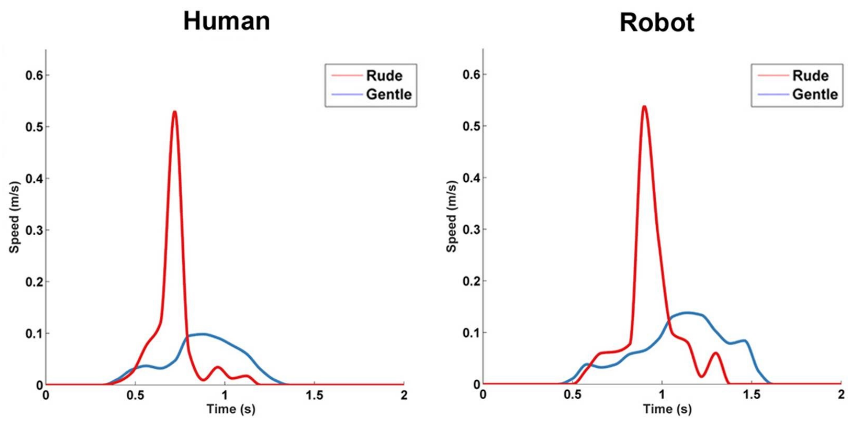

Figure 3.6 *Physical properties of the video stimuli. Graphs depict the profiles of the mean 2-D velocity perceived by participants during the observation of gentle and rude actions performed by human actors and iCub robot.*

Experimental design and subjective evaluation

The design of this experiment has already been described in the previous section. In addition before the fMRI, we decided to gather subjective evaluations of the video stimuli, to verify whether the different styles of the actions were perceived in the same manner for the human and the robot. Participants were presented with video-clips showing human or robot expressing gestures or voice with vitality forms. Both visual and auditory stimuli were presented in two forms: gentle or rude. After visual or auditory stimulus perception, participants were immediately requested to indicate on a Likert scale reported in figure 3.7) the vitality forms perceived.

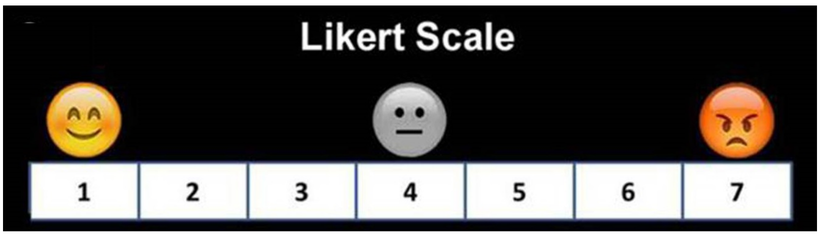

Figure 3.7 *Here is displayed the rating scale presented to participants for the subjective evaluation experiment*

3.4.2 Results

Regarding the subjective evaluation, the plot of figure 3.8, indicate that the vitality of all the stimuli was correctly recognized for both the human and the robot stimuli. This is confirmed by one Two Way Repeated Measures ANOVA on the ratings for the different movies, with Agent (human/robot) and Style (gentle/aggressive) as factors. The difference between Style is highly significant (action: $F(1,19)=298.7$, $p<0.001$), whereas the difference between Agents is not significant (action: $F(1,19)=0.84$, $p=0.37$). From the analysis it emerges also a significant interaction between Agent and Style (action: $F(1,19)=14.1$, $p=0.001$). In particular the human aggressive style shows significantly higher ratings than the robot (Bonferroni post-hoc, $p = 0.002$). However, the distinction between aggressive and gentle behavior is highly evident for both agents.

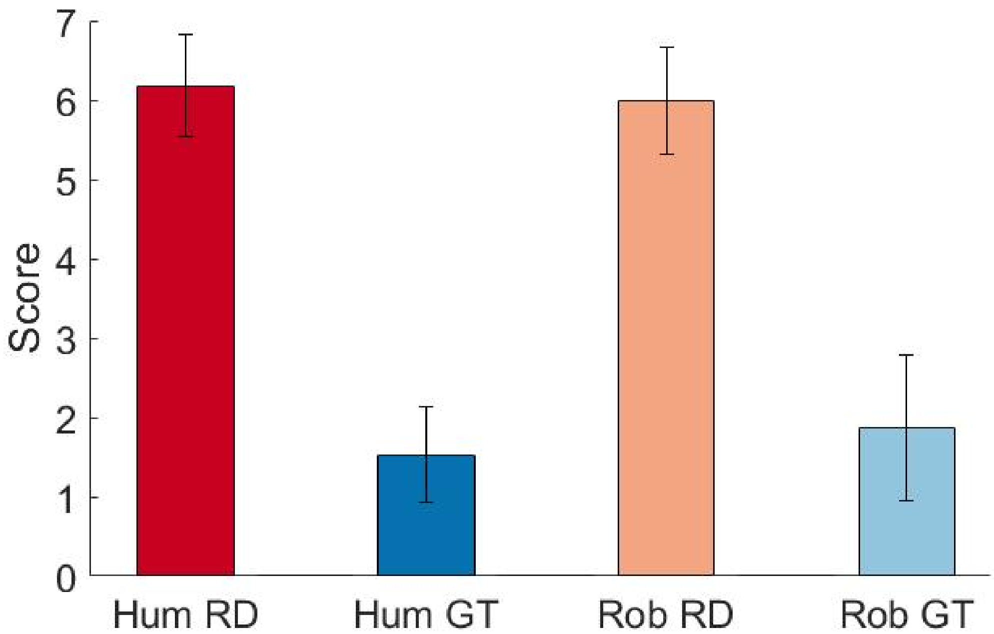

Figure 3.8 *Subjective evaluation of robot behaviors with different vitality, conveyed with different features. Gentle and rude between the robot and actor seem to be perceived almost in the same manner by participants.*

Results indicated that the observation of actions performed by human gently and rudely vs. baseline produced the common activation of the occipital lobe, bilateral inferior parietal lobule, bilateral premotor cortex and inferior frontal gyrus. A similar activation pattern was observed for the robot condition with a large extension of the frontal area in the right hemisphere. Furthermore, the observation of video-clips showing control stimuli (jerky actions) performed by both human and robot produced similar activation patterns more extensive in the right hemisphere for the robot condition (Fig. 3.9).

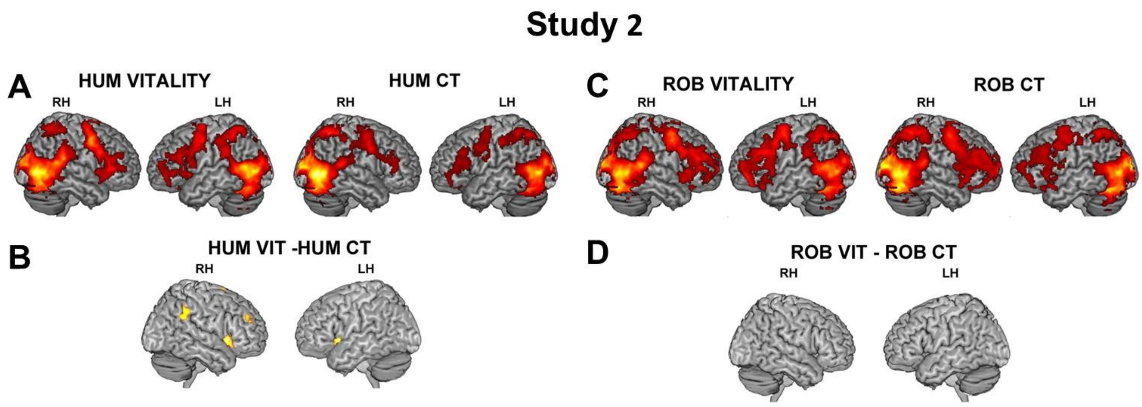

Figure 3.9 *Brain activation during the processing of actions performed by human actors and iCub robot. LH, left hemisphere; RH, right hemisphere.*

The direct contrast between human actions performed with vitality forms (gentle, rude) and the control condition (jerky actions) revealed the activation of the left dorso-central insula (human rude actions vs. human control actions; human gentle actions vs. human control actions). On the contrary, the contrast between the observation of robotic action and the control condition did not produce any activation (robot rude actions vs. robot control actions; robot gentle actions vs. robot control actions).

Differently with results of the previous study, the direct contrast between the observation of human actions and robotic actions did not revealed any activation. Furthermore, no significant activation was found in the contrast human rude actions vs. robot rude actions and in the contrast human gentle actions vs. robot gentle actions (figure 3.10).

3.5 Kinematic study

Considering the work described in the previous sections, we assume that the generated movements for the iCub effectively convey rude and gentle vitality forms to the human observer. The next step would be to test these motion styles in a real interaction with the robot. The aim is to assess the behavioral responses of human participants to the robot actions showing different vitality forms in an interactive setting.

Di Cesare et al. (27) carried out a kinematic study to assess whether visual and auditory properties of vitality forms expressed by others influenced the motor response of participants. They presented video-clips displaying a male and female actor performing a "giving request"

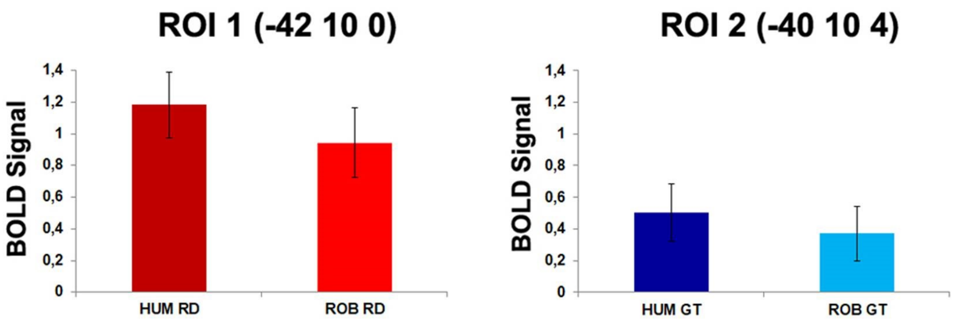

Figure 3.10 *BOLD signal indicating the insula activity during the processing of actions performed by human actors and iCub robot. The bar graphs indicate the comparisons between human and robot in gentle and rude conditions. (RD = rude, GT = gentle).*

(give me) or a "taking request" (take it) in visual, auditory and mixed modalities. These requests could express rude or gentle vitality forms. After watching the video, participants had to perform an action. Their results showed that participants' kinematic features were influenced by vitality forms for both the visual and auditory stimuli. In particular, parameters such as the peak speed and acceleration of their subsequent movements were coherent with the vitality forms that they experienced immediately before: after a rude request, they were faster and displayed a higher acceleration, with respect to the gentle request. Our goal was the reproduction of this effect during a human-robot interaction.

First, we designed and carried out a pilot study to analyze the effect of vitality forms on the kinematic response of participants during a real human-robot interaction. Then, after addressing the possible issues of this pilot, we performed a final experiment involving both video stimuli and live interaction with the robot.

3.5.1 Methods

Participants and data

In the pilot, we tested a total of 10 participants (all right-handed, 4 females, mean age: 27, SD: 3 years). In the final experiment we tested a total of 10 participants (all right-handed, 5 females, mean age: 26.4, SD: 1.8 years). In both the studies we gathered kinematics data of the hand and arm with 5 markers (see fig. 3.11 for positioning) thanks to a NDI OptoTrack system, with 200 Hz sampling frequency. We also performed one questionnaire after the interaction, giving participants the possibility to give open comments. Only in the final experiment we had two questionnaires before the experiment. These were used to better interpret possible particular kinematic behavior of participants: the first regarding aggression and the second about the attitude toward robots. All the ten participants of the final experiment, started with the video phase, then performed the live interaction after some months to avoid a possible bias given by the memory of a live interaction with the robot, that could be stronger than just observing in video.

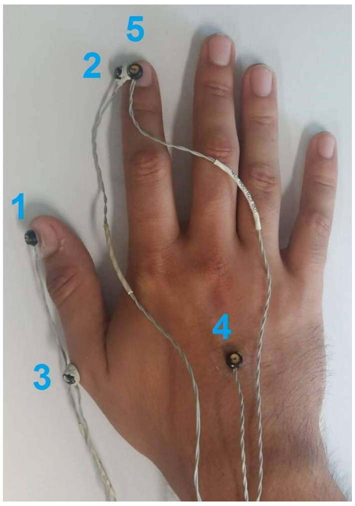

Figure 3.11 *Markers positioning for both studies.*

Experimental design

The experimental setting of the pilot study is depicted in Fig 3.12. After the iCub action execution (passing the object) or word pronunciation ("take it") participants had to take a ball held by the robot at approximately 30 cm from their right hand. The participant comfortably sat in front of the robot with small headphones to hear the robotic voice indications, covered by hearing protectors to avoid experimental biases due to the noise of the moving motors. Between the participant and the robot, we placed a small table with marks indicating the starting position of the right hand and two different targets (yellow and orange) on which the ball had to be placed by the participant.

The robot performed one action (pass the object) or pronounced one action verb in Italian language ("prendi" that is "take it"), with two different vitality forms. The two actions and voices were designed to show rude and gentle robot behavior towards the participant, inspired by the protocol by Di Cesare et al. (27). The kinematics of the actions were the ones described in section 3.4.1, which had proven to be effective in conveying the vitality as shown by observers' neural response. To generate a robotic voice we recorded the voice of a human actor that pronounces the following motor command: "take it" in rude and gentle

Figure 3.12 *Experimental Setup included the working table with two target areas (yellow and orange) and the humanoid robot iCub standing in front of the participant. The face of the robot was covered to avoid bias regarding eyes or facial information.*

forms. We then manipulated some physical properties e.g. the pitch and duration (Cool Edit Pro Software). Finally, the intensity of action verbs was equated for loudness in order to match the corresponding gentle and rude vitality forms of the human voice.

The face of the robot was covered with a black piece of cloth held by two poles at the proper height, since the salience of the information had to be conveyed only either by the action or by the speech. Participants were instructed to replace the ball in the hand of the robot with their left hand. Before the beginning of the experiment, each participant performed a first training phase of ten repetitions, in which they had to take the ball and place it on the target indicated by the robot. During this phase, the robot posture was fixed in the final passing position. After the training, we presented the stimuli in three blocks of 16 repetitions. Each one of the 4 conditions (gentle action, gentle voice, rude action, rude voice) was performed 4 times per block in a randomized order. The sequence of the conditions in each block was identical for every participant. Each repetition had the structure depicted in Fig. 3.13. At first a neutral robotic voice indicates the target (yellow or orange), then one of the four possible stimuli is presented. Note that for the voice stimuli, the hand of the robot was always in the final passing position, equal to the final position of the passing action. When the stimulus is over, the participant takes the ball and puts it on the previously indicated target area. Eventually, a neutral robotic voice asks the participant to put back the ball in place.

The voice messages that indicated the target or the re-positioning of the ball, were designed to be neutral and to avoid any influence on the participant. Also the return movement was programmed to respect biological motion but peak velocity speed was reduced to show neutral vitality. To assess the potential effect of different vitality on participants' behavior during the interaction, we analyzed a few kinematic features of their movement drawing inspiration from a previous study on vitality forms in human-human interaction (27): maximum hand speed during reaching and return – i.e. positioning of object in the target position; maximum hand acceleration during reaching and return. At the end of the experiment, we asked the participants how they would describe the audio and movements of the robot and we left the possibility to give open comments about the experiment.

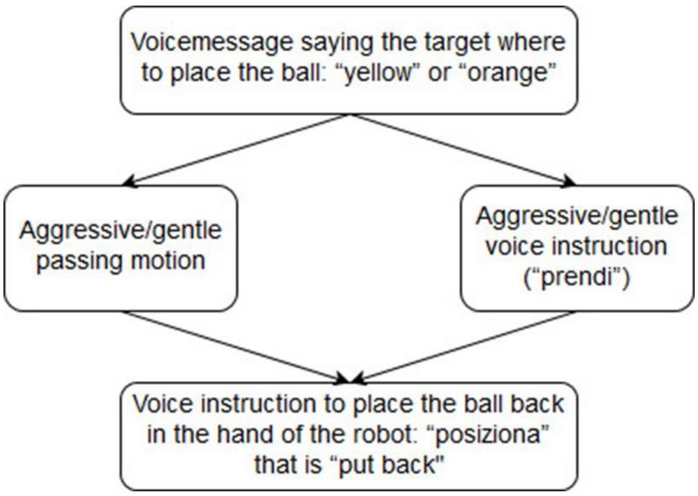

Figure 3.13 *Flowchart showing the sequence of one trial with the possible variations.*

3.5.2 Results of the pilot experiment

In the pilot experiment, we assessed whether participants' action kinematics was affected by the vitality expressed by the robot during a passing interaction. Fig. 3.14 shows the maximum speed averaged across all participants for the reach-to-grasp and the return phase of the movement in response to the robot gentle action or rude action, and the same measures for the voice instructions.

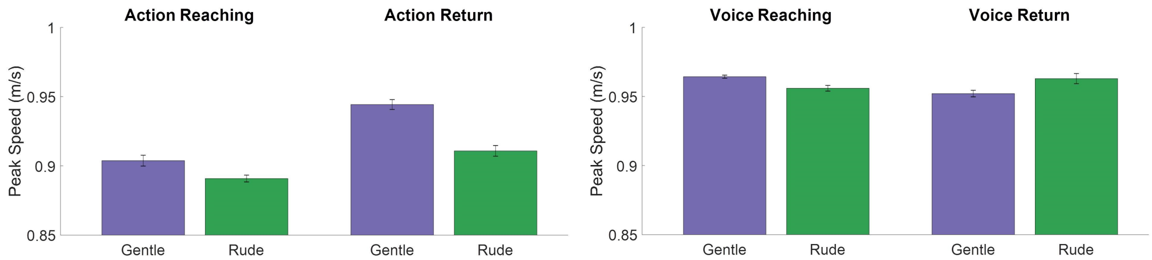

Figure 3.14 *Mean among all participants of the maximum hand speed for the reaching and return phase of the movement in response to the robot's action and voice. The error bars represent standard error of the mean.*

The motion seems slightly faster when the stimulus was gentle, but the main variation can be found between voice and action. In fact, through an two-Way Repeated Measures ANOVA, we did not find a significant difference between the two vitality forms, but only between action and voice ($F(1,9) = 16.13$, $p<0.01$) and no significant interaction. This observation is confirmed also by analyzing the behavior of individual participants as displayed in Fig. 3.16 The properties of the second part of the action, from the grasping to the placement on the table, are approximately equivalent to the ones about the "reaching" part of the movement, just described: a Two-Way Repeated Measures ANOVA confirms again significant difference only between action and voice ($F(1,9) = 16.83$, $p<0.01$).

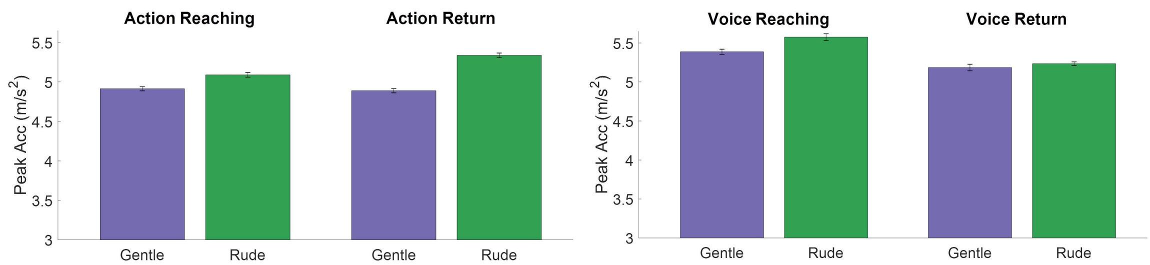

Figure 3.15 *Mean among all participants of the maximum hand acceleration for the reaching and return phase of the movement in response to the robot's action and voice. The error bars represent standard error of the mean.*

Results regarding peak acceleration are displayed in the charts of Fig. 3.15. In this case, contrary to what happens for speed, acceleration is higher when the robot is aggressive, a difference which reaches significance in the return phase of the movement (Fig 8B, $F(1,19) = 5.95$, $p<0.05$). In the reaching phase instead the main difference is between voice and action and is confirmed by a Two-Way Repeated Measures ANOVA ($F(1,19) = 13.33$, $p<0.01$).

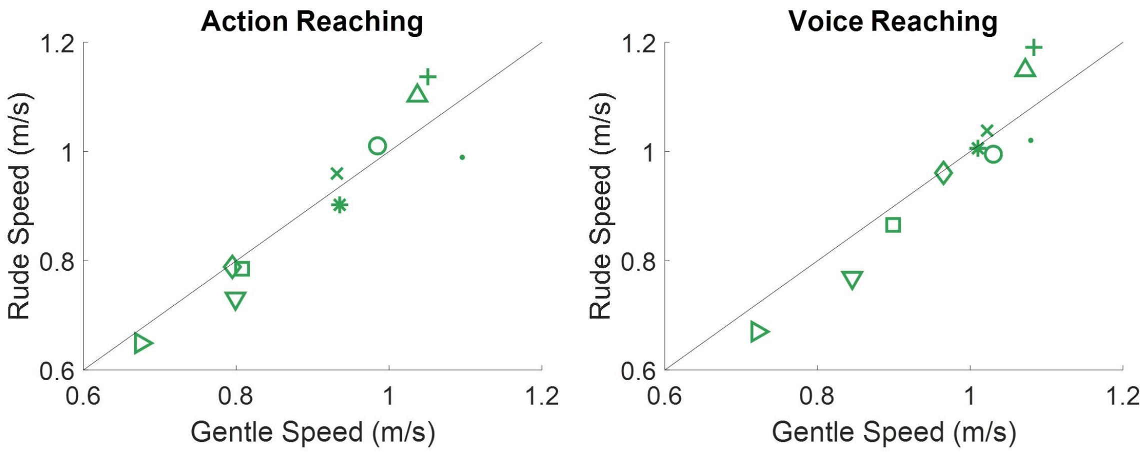

Figure 3.16 *Individual speed of participants for the gentle and rude conditions in response to the action and voice of the robot. If a participant maintained the exact same speed, the corresponding marker would be on the identity line.*

The last results regard subjective answers to the question “how would you describe the voices and movement of the robot?” asked after the experiment. All participants described the rude stimuli using at least one of these words: “aggressive”, “commanding”, “angry”, “rude”, while the gentle stimuli were defined “kind”, “calm”, “relaxed”. Moreover, 8 out of 10 declared that, in their opinion, the difference between the rude and gentle audio was approximately the same if compared to the rude and gentle motion. These comments extend the findings of the first subjective evaluation experiment [sec] that the robot was able to convey vitality forms. In particular, the vitality exhibited by the robot communicated the correct attitudes also when participants did not have only to provide a binary evaluation but were free to choose an arbitrary description.

Eventually, we can say that the subjective recognition of the “style” of the robot is similar for both modalities of communication: action and voice. The two vitality forms are also clearly perceived during an actual interaction with the robot, conveying the perception of the robot being “calm” or “aggressive/commanding”. The behavioral reactions, however, show only minor differences between the two types of robot actions, with just a tendency to show an increase in hand acceleration and speed of grasp aperture in response to an aggressive rather than a gentle robot behavior. Moreover, participants’ motor response was significantly faster and more accelerated in response to robot’s voice rather than robot’s action. These findings do not entirely replicate what has been shown in human-human interaction, where the style and emotion conveyed by voice and movement of the agent, influenced similarly the motor response of the receiver (27). In particular, participants exhibited significantly faster speed and higher acceleration when they watched a video showing rude behavior, on the contrary, they were slower and more relaxed when the video displayed gentle behavior.

Updates to the final version of the experiment

After the pilot experiment, we decided to adjust some details in order to avoid biases and obtain more reliable results. The main changes include:

- We added a video part to the experiment, equal to the interaction with the robot, to have a direct comparison within the same participants in the two different conditions. This would help to relate our work with the human-human interaction study (27) that only included video stimuli, and to investigate whether there are variation between the video and live interaction. The videos were arranged to show the same size of the real iCub during the live interaction.
- The robot did not hold the ball in its hand anymore, but performed a pointing instead. This was necessary to have consistency between the two phases of the experiment, since in the video part we could not move the object like it happened in the passing action during the live interaction.
- The robot wore a red t-shirt. As mentioned in section 3.4 it was useful to hide the shoulder of the robot which is characterized by non-biological motion.
- We also decided to add two questionnaires before the experiment in order to have more information to interpret the kinematic data. The questionnaires were: the Italian validated version of the Aggression Questionnaire by Buss and Perry (63; 64) and the Italian version of the Negative Attitude towards Robots scale (65). With these two we could have a better understanding of the participants' movement and reactions to the iCub.
- We added a set of open questions to be answered after each of the two phases of the experiment. This way we could also verify whether the actions of the robot were perceived in the same way, regardless of the two modalities (video or live interaction).

We did not change the design and flow of the repetitions and blocks, since it was satisfactory in the pilot for both the amount and quality of the data gathered during the sessions.

Figure 3.17 *Experimental Setup improved after the pilot experiment. This is an example of the video phase of the experiment, with the ball held by a tripod in front of the participant.*

3.5.3 Results of the final experiment

Considering the results of the pilot, we tried to improve the final experiment with the modifications explained in the previous section. The main difference consists in the addition of a video phase, identical to the live interaction to allow a comparison of the participant's behavior in these two conditions. To be consistent across the two contexts of video and live interaction, the robot could not hold the ball in its hand anymore, therefore we put it on a tripod in front of the iCub. The passing action was then transformed into an open hand pointing action. We also covered the robot with a red t-shirt, to hide the mechanics of the shoulder joint that are very different to the human's articulation and could have worsen the interaction. Eventually we introduced two questionnaires to possibly give a more precise interpretation to behavioral results.

We performed a GLM statistical analysis on the reaching and return parts of the movement for both speed and acceleration, with post-hoc Bonferroni corrected comparisons. From now on the conditions will be called with the following names:

- CONTEXT to indicate the *live* or *video* condition.

- MODALITY to indicate the *action* or *voice* condition.
- VITALITY to indicate the *gentle* or *rude* stimuli.

In general, for speed we found significant differences for the reaching phase in MODALITY (F(1,9)=23.43, p<0.001) and VITALITY (F(1,9)=14.00, p<0.01). There is also a significant interaction of MODALITY-VITALITY (F(1,9)=15.32, p <0.01) and CONTEXT-MODALITY-VITALITY (F(1,9)=50.11, p<0.001). For the return phase, we found significance only in VITALITY (F(1,9)=7.64, p<0.05) and an interaction of MODALITY-VITALITY (F(1,9)=34.99, p <0.001). More in detail, Fig. 3.18 shows the peak speed averaged across all participants for the reach-to-grasp and the return phase of the movement in response to the robot gentle action or rude action, for both the live and video phases of the experiment. It is clear that vitality of the action modulates the speed of the participant, especially in the reaching phase of the movement where the speed difference between gentle and rude is extremely significant for both the video and live parts. Also the video rude vs live rude and video gentle vs live gentle are significantly different, leading us to think that the physical presence of the robot may have an influence on participant's behavior.

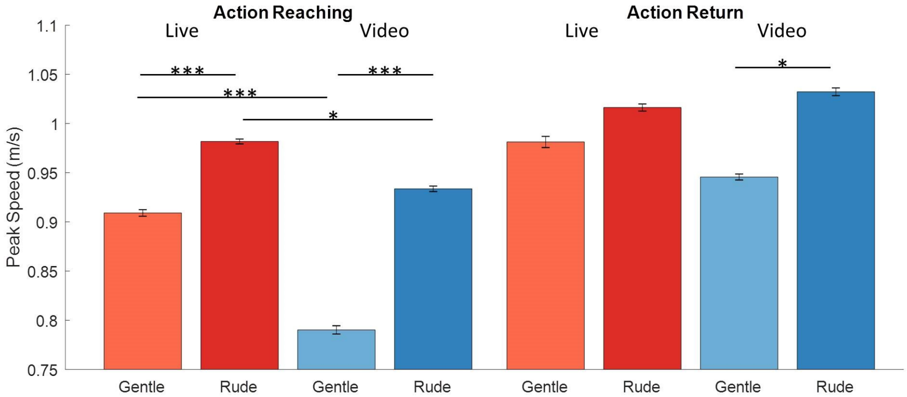

Figure 3.18 *Mean among all participants of the maximum hand speed for the reaching and return phase of the movement in response to the robot's action. The error bars represent standard error of the mean. The symbols indicate the level of significance of post-hoc analysis with Bonferroni correction: * p<0.05, ** p<0.01, *** p<0.001.*

Approximately the same results can be seen for the voice condition in Fig. 3.19, with one important exception: there is no significant difference between the gentle and rude vitality in the video condition. This result suggests that the presence of the robot in this particular interaction configuration could have a role in the modulation of this effect. In both MODALITY conditions, the return phase of the movement does not present almost any strong effect, with the exception of gentle and rude vitality in the action video condition.

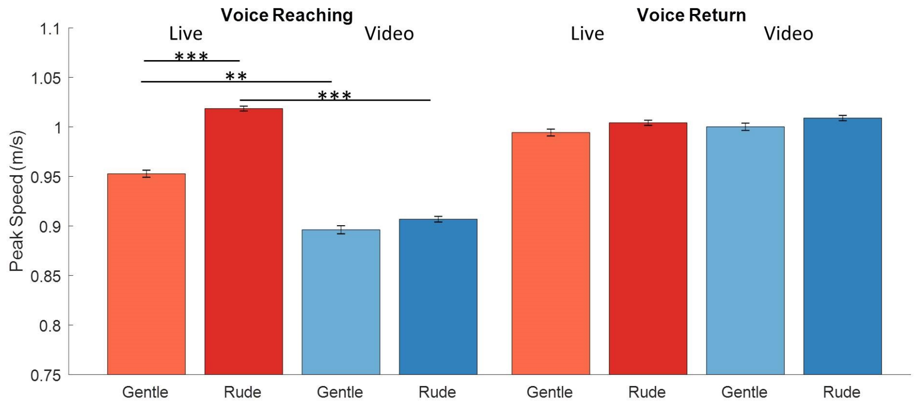

Figure 3.19 *Mean among all participants of the maximum hand speed for the reaching and return phase of the movement in response to the robot's voice. The error bars represent standard error of the mean. The symbols indicate the level of significance of post-hoc analysis with Bonferroni correction: * p<0.05, ** p<0.01, *** p<0.001.*

We went a bit further in the analysis of the speed and inspected the individual behavior of participants in the reaching phase. In Fig. 3.20 we plotted the speed of every single participant in the rude (Y-axis) and gentle (X-axis) conditions. Both graphs show that most of the markers are located above the identity line, meaning that the speed of participants was faster in response to the rude behavior of the robot and slower when the iCub was gentle. This result confirms the more general ones showed above. Another thing that can be noted from the image is that there were no big differences between action and voice reaching for individual participants.

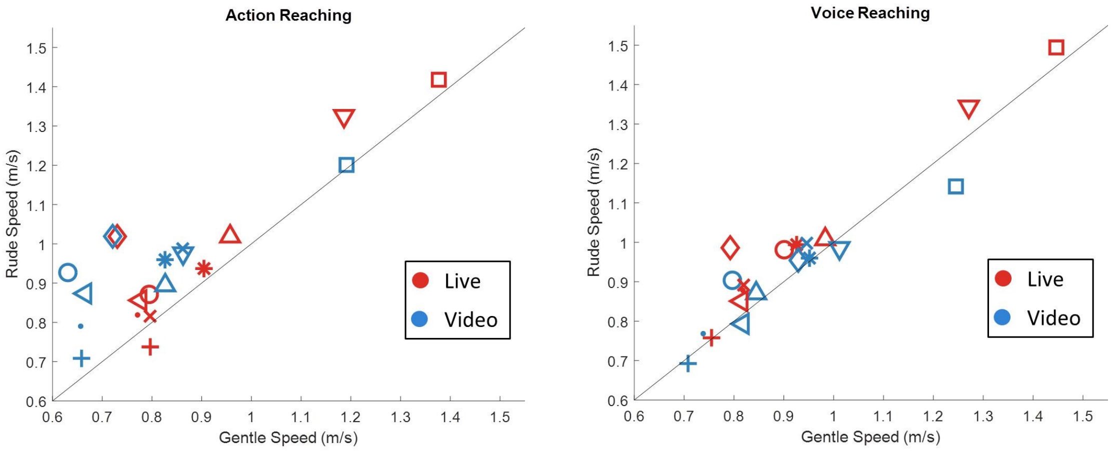

Figure 3.20 *Individual speed of participants for the gentle and rude conditions in response to the action and voice of the robot. If a participant maintained the exact same speed in the different vitality conditions, the corresponding marker would be on the identity line.*

Results for acceleration peaks are less strong but similar to the ones regarding speed. In general we found significant differences in MODALITY (F(1,9)=16.67, p<0.01) and VITALITY (F(1,9)=14.74, p<0.01). There is also a significant interaction of MODALITY-VITALITY (F(1,9)=10.12, p <0.05) and CONTEXT-MODALITY-VITALITY (F(1,9)=22.20, p<0.01). For the return phase, we found significance only in VITALITY (F(1,9)=15.11, p<0.01). In particular, in Fig. 3.21 are displayed the acceleration peaks for the reaching and return phase of the movement. There are significant differences between rude and gentle vitality for both the video and live interaction, and in the rude vitality forms between video and live. These effects are almost equal to the ones regarding speed, corroborating our ideas about automatic imitation of vitality forms. The same occurs less strongly for voice stimuli as shown in Fig. 3.22. Significant difference here is only between gentle and rude vitality in the live condition, and between the gentle stimuli in live and video modalities.

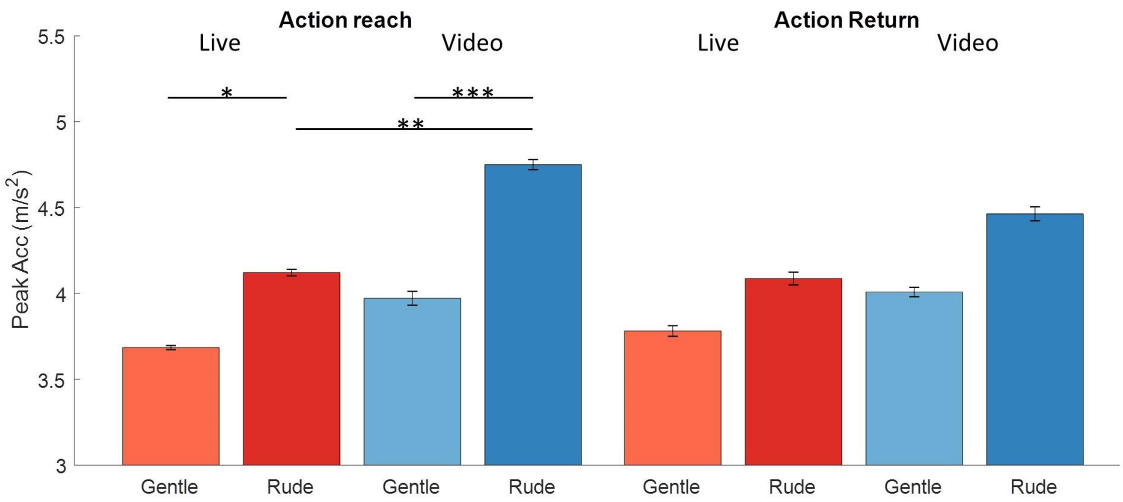

Figure 3.21 *Mean among all participants of the maximum hand acceleration for the reaching and return phase of the movement in response to the robot's action. The error bars represent standard error of the mean. The symbols indicate the level of significance of post-hoc analysis with Bonferroni correction: * p<0.05, ** p<0.01, *** p<0.001.*

We did not find any significant correlation between the Aggression Questionnaire and NARS scores and the kinematic behavior. On the contrary, the open comments questionnaires performed after the two phases confirm a good recognition of the different voices and actions. In particular, the rude vitality form was described as "aggressive", "fast", "unpleasant", "commanding" and also "rude", while for the gentle one, participants used words such as "slow", "calm", "kind", "helpful" and "gentle".

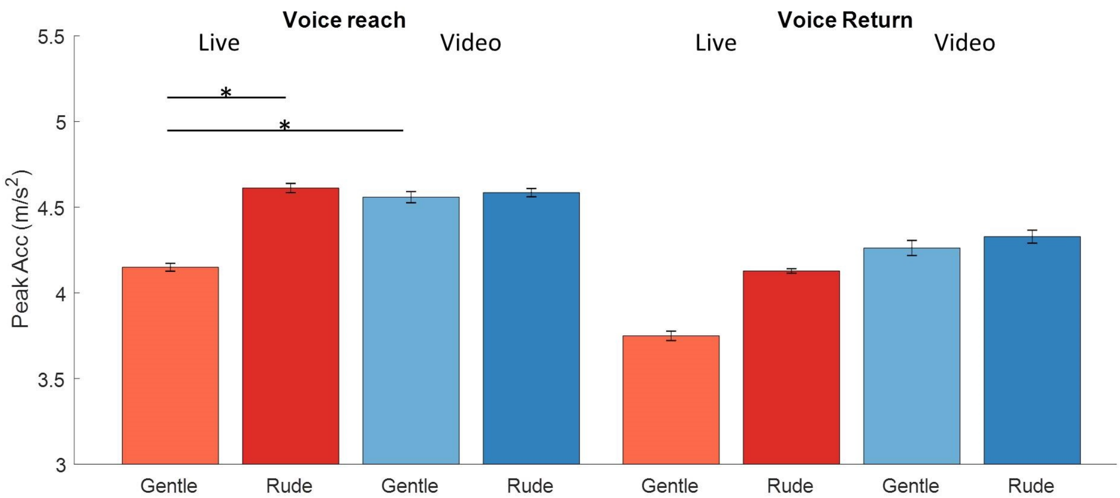

Figure 3.22 *Mean among all participants of the maximum hand acceleration for the reaching and return phase of the movement in response to the robot's voice. The error bars represent standard error of the mean. The symbols indicate the level of significance of post-hoc analysis with Bonferroni correction: * p<0.05, ** p<0.01, *** p<0.001.*

3.5.4 Limitations and future work

In this study, our main goals included: giving the ability to express different vitality forms with actions and voice to the iCub humanoid robot and eventually verify the differences between presenting stimuli in with a video or during a live interaction.

Considering these results we can say that they are coherent with the previous fMRI studies. In the context of our experiment, the different vitality forms are explicitly recognized for all conditions. Implicit communication instead, influences the subsequent action performed by participants in almost all cases, but not always. In fact, while action stimuli provokes an imitation of the style of the robot in every condition, this is not always true for the voice commands. It is important to note that even though the audio stimuli were provided exactly in the same modality for the live and video phases, in the latter case the imitation effect was not triggered. On the contrary, the voice in presence of the robot and the actions in both live and video conditions, influenced the subsequent movements of participants very clearly in speed and acceleration, driving them to match the style of the stimulus observed right before.

Even if there was no contact between the participant and the robot in both the video and live condition, the contagion was strong and the physical presence of the robot could have had an effect on the perception of the voice. We are not used to deal with vitality forms of robotic voices. We believe that the audio could have been more easily associated to the robot when it was present in front of the participant, while this association failed for the video phase regardless of whether it was presented with identical modality. The results in this study confirm that the physical presence of the robot has an impact on certain aspects of the interaction, whereas other properties are preserved even in video-based scenarios, in line with previous literature (20; 66; 67).

Indeed the experiment presented also some limitations that could have influenced the results. First, we did not analyze the features of the voice vitality in depth as the ones of the movements, therefore the audio stimuli can probably be improved a lot in future studies. Other problems include the low number of participants and the order of conditions. All the ten participants in fact, performed the video phase first and the live interaction after some months. We chose this procedure to avoid a possible bias given by the memory of a live interaction with the robot, that could be stronger than just observing in video. We plan to extend the sample size and reverse the order of conditions with ten more participants, to address these limitations.

Part III

Section Three

Chapter 4

Discussion

4.1 Major goals & results

The main goal of my three year PhD has been investigating motion in human-robot interaction. In particular we focused on biological motion, and the various implicit signals that it naturally conveys. Human motion is in fact unconsciously recognized by the brain since early stages of development (26), facilitating the cooperation between individuals: thanks to such signals our interactions are more efficient and pleasant (17). Giving the iCub the ability to move following the rules of human motion, would present some advantages in various situations, from basic communication to more complex backgrounds such as elderly home care and law enforcement. Two aspects of biological motion were addressed: how it triggers adaptation in a cooperative framework and how it can be used to communicate an affective status. First of all, we programmed the robot to move its arms following the Two-thirds power law, a well known regularity of human actions. Then, taking inspiration from human-human interaction, we ported these skills to the iCub and performed two main studies, one to induce participants to adapt to the robot during a collaboration, and another to transmit vitality forms through the robot's movements. In the first study we designed an experiment in which a person and the robot had the same task of filling a box with Lego blocks. We changed the speed of the iCub across the trials, to investigate whether participants would adjust their own speed accordingly. Results show that on average our participants were slower when the robot was slow and faster when the robot was fast, adapting their velocity to their automated humanoid partner. We performed this experiment in two very different countries, Italy and Japan, to verify if cultural background could have an influence on this behavior. Both Italian and Japanese participants showed approximately the same level of adaptation, leading us to think that these low level mechanisms, as motor resonance in the form of automatic imitation, are not influenced by cultural differences. In the second study we started with the generation of different passing actions to display rude and gentle behavior, the so called vitality forms (34). After verifying that these movements were correctly designed to explicitly communicate rudeness and kindness, we used them in different experiments. The first one focused on fMRI analyses to verify whether a robot exhibiting vitality forms could trigger the same neural responses of a human actor moving with the same styles. We managed to obtain a very similar activation of the dorso-central part of the insula in the brain of participants only with a proper design, porting the movements of a human actor to the iCub. A second series of experiments tested the kinematic response to the vitality forms of the robot with videos and in a real interaction. The exercise included a video phase and a live interaction with the iCub. In both circumstances, participants were influenced by the behavior of the robot, replicating the style of the motion displayed by the iCub. Their movement features resembled the ones of an aggressive action after the observation of an aggressive motion of the robot, while on the contrary they performed more gentle action when the robot was gentle itself. These results replicate the findings of previous studies on human-human interaction we took inspiration from (27), showing that humanoid robots can express vitality forms and consequently affect the behavior of the human counterpart during a simple interaction.

4.2 Limitations & future work

As mentioned in the Introduction and in the chapters describing each study, there are some limitations on these works. In general it is difficult to bring the robot in the wild and perform more ecological experiments to investigate the transmission of implicit signals in more realistic environment. Considering our results and the fact that such signals are processed unconsciously, we trust that this problem is not crucial to appropriately explore the context of implicit communication during an interaction. How does this scale to richer interactions involving multiple people and more complex signals needs further study. Again, considering the results obtained in both studies, we believe we have properly implemented a form of human-like motion, making the actions of the robot biologically plausible. If not so, our participants would have probably experienced some sort of uncanny effect derived from the difference between humanoid appearance and non-human behavior, undergoing through an unpleasant interaction. This does not seem the case, since none of the participants involved manifested bother or frustration during the experiments, nor in the informal questions we asked afterwards.

As mentioned in the relative sections, both studies had some limitations. In particular the collaboration involved the presence of a third person that could have affected the interaction, but it is important to note that this did not prevent participant to adjust their motion to the speed of the robot and display a sort of adaptation. In order to enhance our findings, we plan to design a different experiment with a new task to avoid the presence of a third agent and strengthen the feeling of collaboration between the participant and the robot. Also a higher number of participants will be essential. Regarding the vitality study, we took care of generating only one passing/pointing action and two basic styles, rude and gentle. Our strong aspiration is to understand more in depth the features that differ among various vitality forms. Learning how to manipulate such parameters would allow the generation of many diverse styles, starting from the same neutral action. Such generalization could be applied to many fields, from other humanoids to virtual avatars, passing through different kind of robots. Eventually, a challenging and ambitious goal will be the combination of perception and action in robotics. The ability to recognize biological motion and act consequently is crucial to achieve a natural, safe and efficient human-robot interaction in real environments, drawing robots closer to our daily lives.

4.3 Impact

Much research has already been trying to humanize machines making them more aware and respectful of human necessities (25). Robots should adjust their behavior on human needs: this will allow a proper communication through anticipation and understanding of our actions. New technologies can lead to more powerful and efficient robots and these new capabilities should be put at the service of new key concepts such as predictability and understanding skills to greatly improve interaction with humans. It is important to note that humanization is not strictly bound to anthropomorphic appearance, in fact we could say it is more related to a collection of skills that would allow robots to understand and predict human intentions and limitations. It suggests that the robot has a model of the human behavior and uses this representation to communicate in a way that is intuitive for the human partner. This model can be applied to machines with very different embodiments, from humanoids to smart cars.

During the three years of my PhD, I followed an approach based on human-human interaction models, trying to replicate them with an iCub humanoid robot. When we cooperate, there are many implicit signals in the way we perform an action that can suggest our goals and intention, but also the properties of an object or our emotional status. A robot should be able to perceive and send such signals during an interaction, to both understand and be understood by the human counterpart. To reach the same level of safety and efficiency, human-robot interaction must be mutual.

In social science it's been known for a long time that highly ambiguous and unpredictable events and tasks are often evaluated as unpleasant and can be cause of fear and anxiety. These negative feelings come from the difficulty to understand and adapt to such situations. Also the uncanny valley effect in robotics has been linked to similar perception of ambiguity and unpredictability, caused by a mismatch between expectations and reality. To address this problem, the focus of my work has been the design of predictable motion for collaborative actions and the communication of an affective state. This would allow the robot's actions to be more readable and to communicate also different emotions that can be useful in various situations. For example, a robot designed to be a teacher would benefit from moving in a gentle fashion to avoid uneasiness, while in some cases it could act in a more authoritative way to convince people to quickly follow its instructions, depending on the circumstances. This way human-robot interaction can become more natural and safe at all levels and be perceived as such. These applications could be useful in other context such as for virtual agents, already widely spread online as automatic support tools for websites and companies, and more in general in many fields of human-machine interaction.

Figure 4.1 *RBCS and CONTACT groups of IIT. Luca Garello, Gaia Risso, Francesco Rea, Dario Pasquali, Giulia Albanese, Valeria Falzarano, Ana Tanevska, Alexander Aroyo, Giulia Petrella, Laura Bandini, Sara Nataletti, Alessia Vignolo, Alessandra Sciutti, Omar Eldardeer, Linda Lastrico, Giulia Belgiovine, Elena Lechuga, Radoslaw Niewiadomski, Romain Toebosch, Anne Bloem, Inge Hootsmans, Lena Opheij, Giuseppe Di Cesare, Motonobu Aoki, Anna Folso, Jacopo Zenzeri, Carlo Mazzola, Lukas Grasse, Joshua Zonca, Jonas Gonzales, Giulio Sandini*

References

[1] G. Knoblich, S. Butterfill, and N. Sebanz, "Psychological research on joint action: theory and data.," *Psychol. Learn. Motiv. Adv. Res. Theory*, vol. 54, pp. 59–101, 2011.

[2] R. Johansson, G. Westling, A. Bäckström, and J. Flanagan, "Eye-hand coordination in object manipulation.," *J. Neurosci*, vol. 21, p. 6917–6932, 2001.

[3] J. Flanagan and R. Johansson, "Action plans used in action observation.," *Nature*, vol. 424, p. 769–771, 2003.

[4] K. Scherer, "Expression of emotion in voice and music.," *J. Voice*, vol. 9, p. 235–248, 1995.

[5] M. Karg, A. Samadani, R. Gorbet, K. Kuhnlenz, J. Hoey, and D. Kulic, "Body movements for affective expression: a survey of automatic recognition and generation.," *IEEE Trans. Affect.Comput.*, vol. 4, pp. 341–359, 2013.

[6] K. Lohan, H. Lehmann, C. Dondrup, F. Broz, and H. Kose, "Enriching the human-robot interaction loop with natural, and semantic and symbolic gestures.," *Human-Humanoid Interaction, and Humanoid Robotics: A Reference*, 2017.

[7] E. Bicho, L. Louro, and W. Erlhagen, "Integrating verbal and nonverbal communication in adynamic neural field architecture for human-robot interaction.," *Human-Humanoid Interaction, and Humanoid Robotics: A Reference*, vol. 4, p. 5, 2010.

[8] S. Khan, S. Bendoukha, and M. Mahyuddin, "Dynamic control for human-humanoid interaction.," *Human-Humanoid Interaction, and Humanoid Robotics: A Reference*, 2017.

[9] M. Mahyuddin and G. Herrmann, "Cooperative robot manipulator control with human "pinning"for robot assistive task execution.," *International Conference on Social Robotics*, pp. 521–530, 2013.

[10] J. Fink, "Anthropomorphism and human likeness in the design of robots and human-robot interaction.," *Lect. Notes Comput. Sci.*, vol. 7621, p. 199–208, 2012.

[11] C. Breazeal, "Toward sociable robots.," *Rob. Auton. Syst.*, vol. 42, p. 167–175, 2003.

[12] L. Bonini, P. Ferrari, and L. Fogassi, "Neurophysiological bases underlying the organization of intentional actions and the understanding of others' intention.," *Conscious. Cogn.*, vol. 22, p. 1095–1104, 2013.

[13] J. Złotowski, D. Proudfoot, K. Yogeeswaran, and C. Bartneck, “Anthropomorphism: opportunitiesand challenges in human–robot interaction.,” *Int. J. Soc. Robot.*, vol. 7, p. 347–360, 2015.

[14] T. Chaminade and G. Cheng, “Social cognitive neuroscience and humanoid robotics.,” *J. Physiol.*, vol. 103, p. 286–295, 2009.

[15] J.-D. Boucher, U. Pattacini, A. Lelong, G. Bailly, F. Elisei, S. Fagel, P. Dominey, and J. Ventre-Dominey, “I reach faster when i see you look: gaze effects in human-human and human-robot face-to-face cooperation.,” *Front. Neurorobot.*, vol. 6, p. 3, 2012.

[16] H. Admoni and B. Scassellati, “Social eye gaze in human-robot interaction: a review,” *Journal of Human-Robot Interaction*, vol. 6, no. 1, pp. 25–63, 2017.

[17] A. Sciutti, A. Bisio, F. Nori, G. Metta, L. Fadiga, and G. Sandini, “Robots can be perceived as goal-oriented agents.,” *Interact. Stud.*, vol. 14, p. 329–350, 2014.

[18] A. Sciutti, L. Patanè, F. Nori, and G. Sandini, “Understanding object weight from human and humanoid lifting actions.,” *IEEE Trans. Auton. Ment. Dev.*, vol. 6, p. 80–92, 2014.

[19] R. Liepelt, W. Prinz, and M. Brass, “When do we simulate non-human agents? dissociating communicative and non-communicative actions.,” *Cognition*, vol. 103, p. 286–295, 2009.

[20] D. Eizicovits, Y. Edan, I. Tabak, and S. Levy-Tzedek, “Robotic gaming prototype for upper limb exercise: Effects of age and embodiment on user preferences and movement,” *Restorative neurology and neuroscience*, vol. 36, no. 2, pp. 261–274, 2018.

[21] S. Kashi and S. Levy-Tzedek, “Smooth leader or sharp follower? playing the mirror game with a robot,” *Restorative neurology and neuroscience*, vol. 36, no. 2, pp. 147–159, 2018.

[22] E. Oztop, D. Franklin, and T. Chaminade, “Human – humanoid interaction: is a humanoid robot perceived as a human.,” *Int. J. Humanoid. Robot.*, vol. 2, p. 537–559, 2005.

[23] A. Bisio, A. Sciutti, F. Nori, G. Metta, L. Fadiga, G. Sandini, and T. Pozzo, “Motor contagion during human-human and human-robot interaction.,” *PLoS One*, vol. 9, 2014.

[24] V. Gazzola, G. Rizzolatti, B. Wicker, and C. Keysers, “The anthropomorphic brain: the mirrorneuron system responds to human and robotic actions.,” *Neuroimage*, vol. 35, p. 1674–1684, 2007.

[25] Schulz, Trenton, Torresen, Jim, Herstad, and Jo, “Animation techniques in human-robot interaction user studies: A systematic literature review,” *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 8, no. 2, p. 12, 2019.

[26] F. Simion, L. Regolin, and H. Bulf, “A predisposition for biological motion in the newborn baby.,” *Proc.Natl.Acad.Sci.*, vol. 105, p. 809–813, 2008.

[27] G. D. Cesare, E. D. Stefani, M. Gentilucci, and D. D. Marco, “Vitality forms expressed by others modulate our own motor response: A kinematic study.,” *Front. Hum. Neurosci*, 2017.

[28] A. Sciutti, L. Schillingmann, O. Palinko, Y. Nagai, and G. Sandini, “A gaze-contingent dictating robot to study turn-taking.,” *IEEE International Conference on Human-Robot Interaction*, p. 137–138, 2015.

[29] C. Chao and A. Thomaz, “Timing in multimodal turn-taking interactions: Control and analysis using timed petri nets.,” *Human-Robot Interact.*, vol. 1, p. 4–25, 2012.

[30] G. Hoffman and C. Breazeal, “Effects of anticipatory perceptual simulation on practiced human-robot tasks.,” *Auton. Robots*, vol. 28, p. 403–4023, 2010.

[31] D. Eizicovits, Y. Edan, I. Tabak, and S. Levy-Tzedek, “Robotic gaming prototype for upper limb exercise: Effects of age and embodiment on user preferences and movement.,” *Restorative Neurology and Neuroscience*, vol. 36, pp. 261–274, 2018.

[32] F. Kaplan, “Who is afraid of the humanoid? investigating cultural differences in the acceptance of robots.,” *International Journal of Humanoid Robotics*, vol. 1, p. 465–480, 2004.

[33] T. L. Chartrand and J. L. Lakin, “The antecedents and consequences of human behavioral mimicry.,” *Annual Review of Psychology*, vol. 64, p. 285–308, 2013.

[34] D. Stern, *Forms of vitality: Exploring dynamic experience in psychology, and the arts, and psychotherapy, and development.* Oxford University Press, 2010.

[35] L. Takayama, D. Dooley, and W. Ju, “Expressing thought: improving robot readability with animation principles.,” *6th ACM/IEEE Int. Conf. Human-Robot Interact.*, 2011.

[36] M. J. Gielniak, C. K. Liu, and A. L. Thomaz, “Generating human-like motion for robots.,” *Int. J. Rob. Res.*, 2013.

[37] A. Sciutti, M. Mara, V. Tagliasco, and G. Sandini, “Humanizing human-robot interaction: On the importance of mutual understanding.,” *IEEE Technol. Soc. Mag.*, vol. 37, p. 22–29, 2018.

[38] T. Hashimoto, S. Hitramatsu, T. Tsuji, and H. Kobayashi, “Development of the face robot saya for rich facial expressions.,” *SICE-ICASE International Joint Conference*, 2006.

[39] T. Fukuda, J. Taguri, F. Arai, M. Nakashima, D. Tachibana, and Y. Hasegawa, “Facial expression of robot face for human-robot mutual communication.,” *IEEE Int. Conf. Robot. Autom.*, 2002.

[40] G. Sandini, A. Sciutti, and F. Rea, “Movement-based communication for humanoid-human interaction.,” *Humanoid Robotics: A Reference*, p. 1–29, 2017.

[41] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin, “Effects of nonverbal communication on efficiency and robustness in human-robot teamwork.,” *IEEE/RSJ International Conference on Intelligent Robots and Systems, and IROS*, 2005.

[42] O. Palinko, F. Rea, G. Sandini, and A. Sciutti, "A robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration.," *IEEE International Conference on Intelligent Robots and Systems, and IROS*, 2016.

[43] A. Sciutti and G. Sandini, "Interacting with robots to investigate the bases of social interaction.," *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2017.

[44] S. Planalp, "Varieties of cues to emotion in naturally occurring situations.," *Cogn. Emot.*, 1996.

[45] F. E. Pollick, H. M. Paterson, A. Bruderlin, and A. J. Sanford, "Perceiving affect from arm movement.," *Cognition*, 2001.

[46] J. Montepare, S. B. Goldstein, and A. Clausen, "The identification of emotions from gait information.," *J. Nonverbal Behav.*, vol. 11, pp. 33–42, 1987.

[47] P. Viviani and T. Flash, "Minimum-jerk, and two-thirds power law, and isochrony: converging approaches tomovement planning.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 21, p. 32–53, 1995.

[48] G. Catavitello, Y. P. Ivanenko, F. Lacquaniti, and P. Viviani, "Drawing ellipses in water: evidence for dynamic constraints in the relation between velocity and path curvature.," *Experimental Brain Research*, vol. 234, p. 1649–1657, 2016.

[49] N. Noceti, F. Rea, A. Sciutti, F. Odone, and G. Sandini, "View-invariant robot adaptation to human action timing.," *SAI Intelligent Systems Conference*, p. 804–821, 2018.

[50] U. Pattacini, F. Nori, L. Natale, G. Metta, and G. Sandini, "An experimental evaluation of a novel minimum-jerk cartesian controller for humanoid robots.," *EEE/RSJ International Conference on Intelligent Robots and Systems, and IROS*, p. 1668–1674, 2010.

[51] T. Chaminade, D. Franklin, E. Oztop, and G. Cheng, "Motor interference between humans and humanoid robots: Effect of biological and artificial motion.," *4th IEEE International Conference on Development and Learning*, 2005.

[52] A. Mörtl, T. Lorenz, and S. Hirche, "Rhythm patterns interaction - synchronization behavior for human-robot joint action.," *PLoS One*, vol. 9, 2014.

[53] G.Metta, L. Natale, F. Nori, and G. Sandini, "The icub project: An open source platform for research in embodied cognition.," *IEEE Workshop on Advanced Robotics and its Social Impacts*, pp. 24–26, 2011.

[54] L. Amoruso and C. Urgesi, "Contextual modulation of motor resonance during the observation of everyday actions.," *Neuroimage*, vol. 134, pp. 74–84, 2016.

[55] L. Amoruso, A. Finisguerra, and C. Urgesi, "Tracking the time course of top-down contextual effects on motor responses during action comprehension.," *Journal of Neuroscience*, vol. 36, p. 11590–11600, 2016.

[56] B. Rauchbauer, J. Majdandžić, A. Hummer, C. Windischberger, and C. Lamm, "Distinct neural processes are engaged in the modulation of mimicry by social group-membership and emotional expressions.," *Cortex*, vol. 70, pp. 49–67, 2015.

[57] P. Molenberghs, V. Halász, J. B. Mattingley, E. J. Vanman, and R. Cunnington, "Seeing is believing: Neural mechanisms of actionperception are biased by team membership.," *Human BrainMapping*, vol. 34, p. 2055–2068, 2013.

[58] M. Baldassarre and S. Feller, "Cultural variations in personal space: theory, and methods, and evidence.," *Ethos*, vol. 3, p. 481–503, 1975.

[59] G. Metta, P. Fitzpatrick, and L. Natale, "Yarp – yet another robot platform, and version 2.3.20.," *International Journal of Advanced Robotic Systems*, vol. 3, 2006.

[60] "Icubforwardkinematics.," *http://wiki.icub.org/wiki/ICubForwardKinematics*, 2014.

[61] L. Oberman, J. McCleery, V. Ramachandran, and J. Pineda, "Eeg evidence for mirror neuron activity during the observation of human and robot actions: Toward an analysis of the human qualities of interactive robots.," *Neurocomputing*, vol. 70, p. 2194–2203, 2007.

[62] B. Urgen, S. Pehlivan, and A. Saygin, "Distinct representations in occipito-temporal, and parietal, and premotor cortex during action perception revealed by fmri and computational modeling.," *Neuropsychologia*, vol. 127, pp. 35–47, 2019.

[63] A. Fossati, C. Maffei, E. Acquarini, and A. D. Ceglie, "Multigroup confirmatory component and factor analyses of the italian version of the aggression questionnaire.," *European Journal of Psychological Assessment*, vol. 19, no. 1, p. 54, 2003.

[64] A. Buss and M. Perry, "The aggression questionnaire.," *Journal of personality and social psychology*, vol. 63, no. 3, p. 452, 1992.

[65] D. Syrdal, K. Dautenhahn, K. Koay, and M. Walters, "The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study," *Adaptive and emergent behaviour and complex systems*, 2009.

[66] J. Li, "The benefit of being physically present: A survey of experimental works comparing copresent robots, and telepresent robots and virtual agents," *International Journal of Human-Computer Studies*, vol. 77, pp. 23–37, 2015.

[67] S. Levy-Tzedek, S. Berman, Y. Stiefel, E. Sharlin, J. Young, and D. Rea, "Robotic mirror game for movement rehabilitation," in *2017 International Conference on Virtual Rehabilitation (ICVR)*, pp. 1–2, IEEE, 2017.