Evaluating the Engagement with Social Robots

Salvatore Anzalone, S Boucenna, Serena Ivaldi, Mohamed Chetouani

Evaluating the Engagement with Social Robots

Salvatore M. Anzalone · Sofiane Boucenna · Serena Ivaldi · Mohamed Chetouani

Abstract To interact and cooperate with humans in their daily-life activities, robots should exhibit human-like "intelligence". This skill will substantially emerge from the interconnection of all the algorithms used to ensure cognitive and interaction capabilities. While new robotics technologies allow us to extend such abilities, their evaluation for social interaction is still challenging. The quality of a human-robot interaction can not be reduced to the evaluation of the employed algorithms: we should integrate the engagement information that naturally arises during interaction in response to the robot's behaviors. In this paper we want to show a practical approach to evaluate the engagement aroused during interactions between humans and social robots. We will introduce a set of metrics useful in direct, face to face scenarios, based on the behaviors analysis of the human partners. We will show how such metrics are useful to assess how the robot is perceived by humans and how this perception changes according to the behaviors shown by the social robot. We discuss experimental results obtained in two human-interaction studies, with the robots Nao and iCub respectively.

S. M. Anzalone · M. Chetouani
Sorbonne Universités, UPMC Université Paris 06, ISIR UMR7222, Paris, France
CNRS, Institut des Systèmes Intelligents et de Robotique UMR7222, Paris, France
Tel. : +33.1.44.27.51.41
Fax : +33.1.44.27.51.45
E-mail: anzalone@isir.upmc.fr

S. Boucenna
ETIS Laboratory, UMR CNRS 8051, the University of Cergy-Pontoise, ENSEA, Cergy-Pontoise, France
E-mail: boucenna@ensea.fr

S. Ivaldi
Inria, Villers-lès-Nancy, F-54600, France
Intelligent Autonomous Systems Lab, TU Darmstadt, Germany
Université de Lorraine, Loria, UMR n 7503, Vanduvre-lès-Nancy, F-54500, France
CNRS, Loria, UMR n 7503, Vandoeuvre-lès-Nancy, F-54500, France
E-mail: serena.ivaldi@inria.fr

1 Introduction

The development of social robots focuses on the design of living machines that humans should perceive as realistic, effective partners, able to communicate and cooperate with them as naturally as possible [13]. To this purpose, robots should be able to express, through their shapes and their behaviors, a certain degree of "intelligence" [31]. This skill entails the whole set of social and cognitive abilities of the robot, which makes the interaction "possible" in a human-like manner, through exchange of verbal and nonverbal communication, learning how to predict and adapt to the partner's response, ensuring engagement during interactions, and so on. The development of such abilities, for a robotics researcher, translates into the implementation of several complex algorithms to endow the robot with different cognitive and social capabilities: multimodal people tracking [5], face recognition [79], gesture recognition [17], speech recognition [18, 21], object learning [40], motor skills learning [7], action synchronization [2, 54], just to name a few. Each algorithm or module is evaluated in the metric space of its specific problem. If we limit ourselves at evaluating their performance or their coordination, we make the mistake of evaluating their efficiency as algorithms [59] rather than their capability to obtain a desired effect once they are used in a
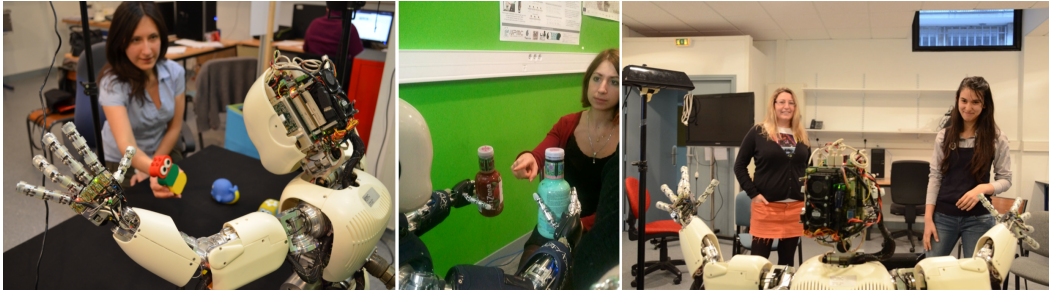
Fig. 1: Typical human-robot interactions scenarios with the iCub robot [52].

human-robot interaction context. If all those modules worked perfectly, would the robot be perceived as intelligent? The answer is not granted, for example recent studies showed that humans prefer to interact with a "non-perfect" robot that makes mistakes and exhibit some uncertainty or delays [63,1].

Evaluating the quality of the experiences with a social robot is critical if we want robots to interact socially with humans to provide assistance and enter their private and personal dimension [67]. But how can we evaluate whether a robot is capable of engaging a human in social tasks? Do we have metrics to determine whether different robot behaviors can improve the quality of such human-robot interaction? Most importantly, can we find good metrics that can be retrieved by cheap sensors (e.g., a Kinect) and in "natural" interaction scenarios, without recurring to the use of invasive measuring devices (e.g., eye-trackers or motion capture systems)?

Measuring the quality of the experiences [57] involving social robots can be a quite challenging task, involving the assessment of several aspects of the interaction, such as the user expectations, his feelings, his perceptions and his satisfaction [23,46]. A characterizing feature of the user experience is given by the ability of robot to engage users in the social task. As stated by [55]: "Engagement is a category of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control". The paper will focus on engagement as characterizing feature of the quality of the experiences with social robots, defining it as "the process by which individuals involved in an interaction start, maintain and end their perceived connection to one another" [65]. In direct, face-to-face scenarios, measurable changes of the human partners behaviors will reflect this engagement through both physiological variations [51], as heart rate [77] or skin conductivity changes [45], and movements [60], as synchronous and asynchronous motions such as nodding [47], gazing [66,

58] or mimicry [11]. Such movements correspond to the non-verbal body language that humans use to communicate each other [31,14]. In the light of this observation, it seems possible to infer the engagement of a person involved in a social interaction with the robot through an analysis of his non-verbal cues. With social interactions we do not refer exclusively to cooperative scenarios, in which, for instance, nodding or joint attention can be seen as feedback given to the partners. To some extent, the same holds to competitive and deceptive interactions [72], where the dynamics of non-verbal behaviors are still used as feedback to humans, for instance, to communicate boring, misunderstandings, rejections or surprise. In any case, variations of non-verbal cues between study groups can inform about the engagement of the partners involved in the social task. This assumption is then very general and able to include a large variety of social interactions and it becomes a powerful instrument to evaluate and, in some cases, to manipulate the synergy between the peers.

Several social signals have been proposed in literature to study the engagement. Hall et al. [34] manipulated nonverbal gestures, like nodding, blinking and gaze aversion, to study the perceived engagement of the human participants, retrieved by a post-experimental questionnaire. Significant works focusing on engagement during verbal interactions were proposed by Rich and Sidner. In particular, in [58] manually annotated engagement was analyzed through mutual and directed gaze, and correlated with spoken utterances. In [66] and [65], via manual labeling, gaze signals have been used by the authors distinguishing between head nods and quick looks; in [38], where the authors combined different gaze behaviors, captured using eye tracker, for conversational agents. Ivaldi et al. [39] also used post-experimental questionnaires to evaluate engagement, but obtained indirect measurements of engagement through the rhythm of interaction, the directional gaze and the timing of the response to robot's stimuli, captured by the use of RGB-D data. In [60], engagement is automatically evaluated from videos of in-

teractions with robots, using visual features related to the body posture, precisely to the inclination of the trunk/back. Similar measures have been used to evaluate behaviors in medical contexts [12] using audio features and video analysis [10] [62] [61].

In this paper we propose a methodology to evaluate the engagement aroused during interactions between social robots and human partners, based on metrics that can be easily retrieved from off-the-shelves sensors. Such metrics are principally extracted by static and dynamic behavioral analysis of posture and gaze, and have been supporting our research activities in human-robot interaction.

We remark that our study is focused on methodologies that do not require intrusive devices that could make the human-robot interaction unnatural, such as eye-trackers or wearable sensors. We choose to work with cheap sensors like Kinects and microphones that can be easily placed in indoor environments and are easy to accept for ordinary people. These features are important since we target real applications with users that are not familiar with robotics. Users' perception and need is an element that must be taken into account by the experimental and robotics setting [71].

2 Material and Methods

To evaluate the engagement, here we address direct, face-to-face, interaction scenarios, where a robot is used to elicit behaviors on humans. This is the case, for example, of a robot playing interactive games with a human partner, or the case of a human tutoring the robot in a cooperative or learning task. The choice of such kind of scenarios, as in Figure 1, does not represent a huge limitation on the validity and on the use of the proposed methodology: many social interactions between humans usually occur in similar conditions.

In this scenario, we assume the human standing in front of the robot. An RGB-D sensor is placed close to the robot, in the most convenient position to capture (as much as possible) the environment and the interacting partner. The information about the robot status and behavior is continuously logged; at the same time, the RGB-D sensor captures, synchronizes and stores the data perceived about the environment and the human [3]. The human posture and his head movements are tracked, according to the processing pipeline shown in Figure 2. Such information is then statically and dynamically analyzed to retrieve: body posture variation, head movements, synchronous events, imitation cues and joint attention.
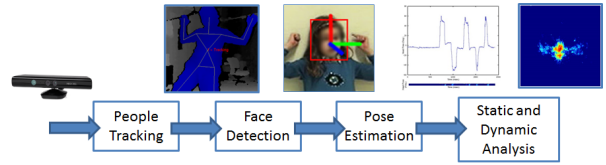

Fig. 2: The algorithm's pipeline employed for the extraction of social measures.

2.1 People Tracking

The presence, position and activity of people interacting with the robot is processed by a posture tracking system based on the data captured by the RGB-D sensor. The humans location and their body parts are tracked in the visible range of the camera to extract gestures, body movements and posture.

Precisely, the depth data perceived by the sensor is processed through a background subtraction algorithm [64]; from each point of the foreground, local features are calculated and classified to assess to which body part (among 31 possible patches) they belong. Finally, through a randomized decision forest, a dense probabilistic skeleton is obtained. Joints of the body are calculated according to the position and the density of each patch.

The tracking algorithm provides a complete map of the body, characterized by the position and the orientation of 15 anatomical parts, including the two arms, the two legs, the torso and the head. Concerning the latter, the algorithm is not able to retrieve an accurate orientation of the head: to accomplish this task, we need a different, dedicated approach that we describe in the following.

2.2 Head Movements

Once the presence of the human partner is found and his body is tracked, the information about the head pose can be extracted.

From the 3D information about the body of the person interacting with the robot, the estimated position of the head is back projected over the RGB image captured by the sensor, to obtain the coordinates in the image space in which the face should appear. A rectangular area of the camera image, centred on such coordinates, is then cropped to retrieve the face, as shown in Figure 3. A face tracking algorithm is then applied to retrieve the head pose: our face tracking implementation is based on a Constrained Local Model approach [24]. This class of trackers is based on a statistical model of the faces based on a set of constrained landmarks,
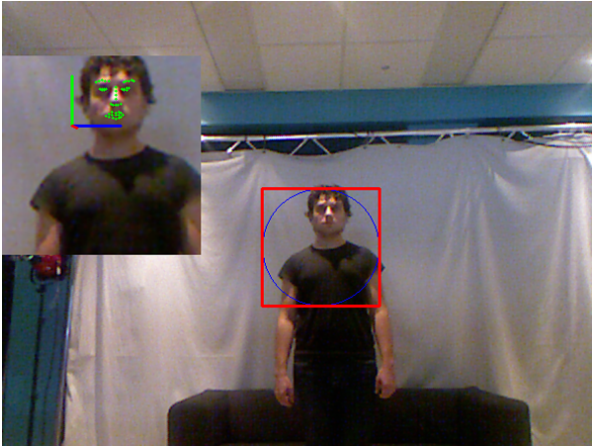
Fig. 3: The face detection and the head pose extraction.

Fig. 4: The histogram heat map during the jumping jack exercise. Red spots highlight the start/stop position of each body articulation.

such as the face shape, its texture, its appearance. Such landmarks are used to approximate the contour of the face, the lips, the eyes, the nose. The algorithm tries to adapt iteratively the shape defined by the landmarks to find the best fit. The result of the algorithm is the best fitting set of facial landmarks approximating the actual face. From the facial landmarks, the orientation of the whole head is calculated and integrated in the full body model.

The head pose provides only an approximation of the gazing direction, since it cannot capture the eyes movements or their direction. However, it continues to be an informative estimator in the case in which potential targets of the interactions are displaced respect to the person's field of view. In such scenarios, the objects are located in a way that they are not visible unless the participants turn their heads toward them, then they are forced to turn the head toward the targets to gaze them. In absence of high resolution cameras that could provide more accurate images of the eyes, the head orientation provides a fair estimate of the human gaze direction. Most importantly, it does not require invasive devices as wearable eye-trackers.

2.3 Static Analysis

We hereby extract and analyze the information related to the posture and gaze of the human interacting with the robot. Histograms are used to study the distribution of the measured data.

Figure 4 shows the 2D histogram of the position of each joint of a person while performing the "Jumping jack" exercise. The time distribution of the joints/bodies positions is conveniently represented as a heat map overlapped with a snapshot of the person performing such movement. In this exercise, the person jumps from
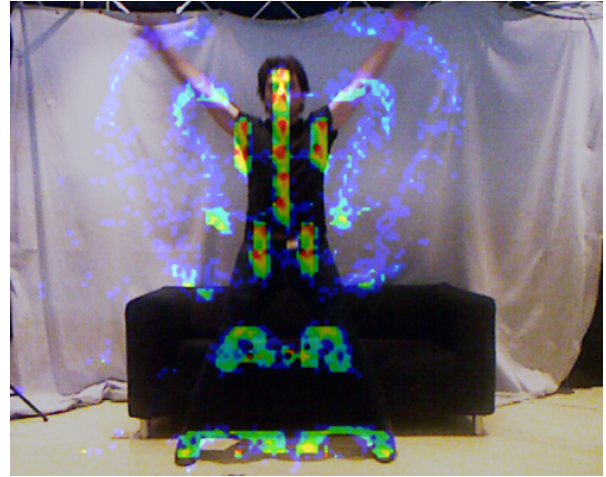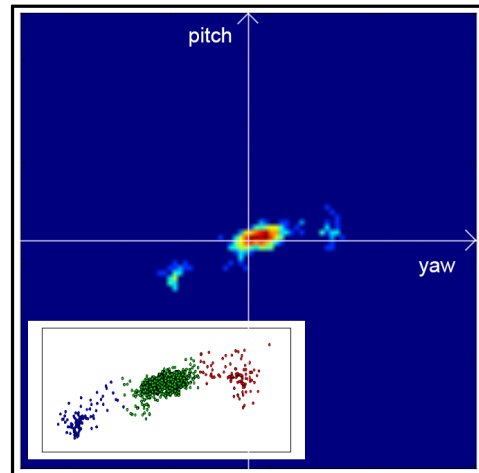

Fig. 5: The histogram heat map of a person's head movements. Peaks correspond to three different focuses of attention, on the center, on the left and on the right. Overlapped, the k-means classification of such data.

a standing position with the feet together and arms at the sides of the body, to a position with legs spread wide and hands over the head. The heat map shows hot spots over the positions in which the body joints spend more time. In particular, red spots depict the start/stop position of each joint during the jumping jack movement. The heat map allows to capture with a simple visualization the posture information in time, such as the movement of the trunk and its stability. Also, it is able to show the variability of the trajectories of arms and legs.

A similar analysis can be done for the gaze. Figure 5 shows the 2D histogram of a person gazing to a tutor,
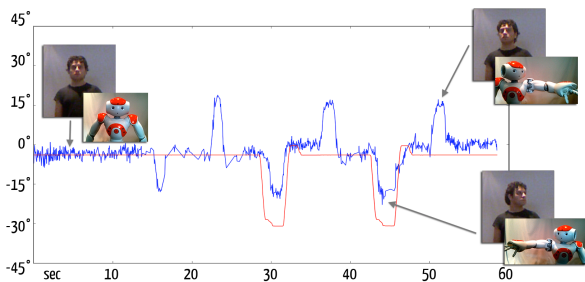
Fig. 6: The evolution over time of the robot's arm joint, in red, overlapped to the head yaw movements of the human, in blue. This highlights two synchrony events between the pointing gesture of the robot and the gazing behavior of the human.

Fig. 7: The evolution over time of the robot's arm joint, in red, overlapped to the arm movements of the human, in blue. This highlights a synchrony event in terms of imitation, between the pointing gesture of the robot and the pointing behavior of the human.

Fig. 8: The evolution over time of the robot's head yaw, in red, overlapped to the head yaw movements of the human, in blue. This highlights several synchrony events in terms of joint attention, between the head movement of the robot and the gazing behavior of the human.

standing in front of him, and over objects on his two sides. The resulting heat map is again a very convenient visualization tool: it shows the focus of attention of the person, highlighting the correspondent hot spots of the head gazing towards the tutor and towards the two objects. It must be noted that the gaze direction is projected on the pitch-yaw plane of the head, since the gaze is approximated by the head orientation as described in Section 2.2.

One possible way to study the head movements is by applying data mining algorithms. In the bottom-left corner of Figure 5, we can see the three areas found applying a clustering algorithm – precisely k-means (k=3, as the 3 hot spots). The information related to the clusters, such as their average, barycenter, density and variance, can be used to extract useful descriptive features of the gaze behaviors of the humans interacting with the robot. The analysis of such signals can also provide information about the head stabilization during fixation. As we will show in the next Section, the clusters information can be used for example to compare the outcome of different experimental conditions.

2.4 Dynamic Analysis

The histogram analysis of the head movements and of the body posture gives only a partial description of the human behaviors, because it does not capture the movement dynamic.

The time analysis can be useful to individuate synchronous events and phenomena occurring when the interacting agents synchronize their actions. In particular it can capture causality, synchronous and delayed events [27]. Figure 6 reports the elicitation of a gaze in response to a robot attention cue: precisely, the robot points its arm twice towards some objects. The figure
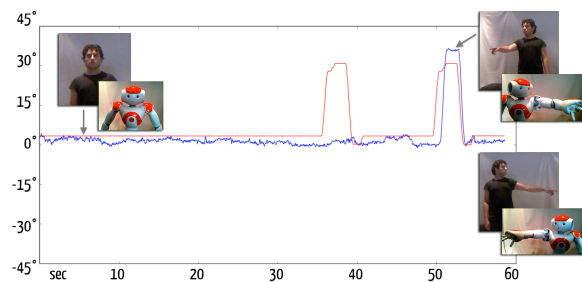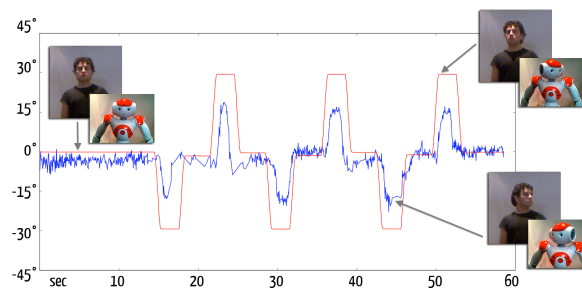
plots in blue the yaw movement of the human head, and in red the movement of the shoulder of a robot: we can see that the head movement is induced by the robot goal-directed action. Figure 7 shows in blue the behavior of the shoulder of a person, and in red the same data from the robot. The plot highlights how the robot fails the first elicitation, while the human imitates it in the second elicitation.

The time between the beginning of the robot's arm movement and the beginning of the human gaze can be interpreted as a measure of the effectiveness of the nonverbal communication abilities of the robot [9]. Humans could respond as fast as if they were communicating with another human, if the robot was readable and "interactable" [39]. If the robot lacks in communication abilities, humans could struggle on understanding the communication cue, thus responding slower than in the ideal case; this delay, if not controlled, can make the response non-contingent. Lastly, humans may not re-

Analysis	Information
Static	**Focus of Attention**
	Clustering of gaze
	Head stability
	Head pose, variance
	Body posture stability
	Barycenter, bodies/joints pose, variance
Dynamic	**Joint attention**
	Clustering of gaze, response times
	Synchrony
	Response times, rhythm of interaction
	Imitation
	Response times, rhythm of interaction, variance

Table 1: Static and dynamic metrics for evaluating the quality of dyadic and triadic face-to-face interactions.

spond at all, because of a complete inability of the robot to communicate with nonverbal-behaviors or gestures.

The synchrony of human-robot movements and the contingent response to reciprocal cues are critical features for evaluating the quality of imitation and joint attention [48,30]. Figure 8 highlights the joint attention elicited by the robot towards a human: in blue the yaw head movement of the human, in red the robot's. Here, the robot is always able to elicit joint attention, as there is a contingent response to each attention cue – a gaze towards an object on the left or right.

Among the temporal features of an interaction, we can account the rhythm of interaction [39] or the pace of interaction [58]: this measure relates to the time between consecutive interactions. The more the pace of human-robot interaction tends to the one of human-human interaction, the more this interaction is perceived as natural [35].

An important matter is the automatic discovery of events, such as beginning and end of interactions. This can be relatively easy from the robot's point of view, since its actions are typically determined by a state machine or some parametrized policy: it is trivial to get the time of the perception events triggering a behavior. On the contrary, it becomes trickier to retrieve the events describing human behaviors from the flow of RGB-D data. One possible way to discriminate easily between activity and inactivity period is to analyze the time spectrum of the joint trajectories, and threshold the energy of such signals computed across a sliding window.

3 Case studies

The presented methods have been successfully employed in two different human-robot interaction experiments. Both experiments focused on a triadic interaction where
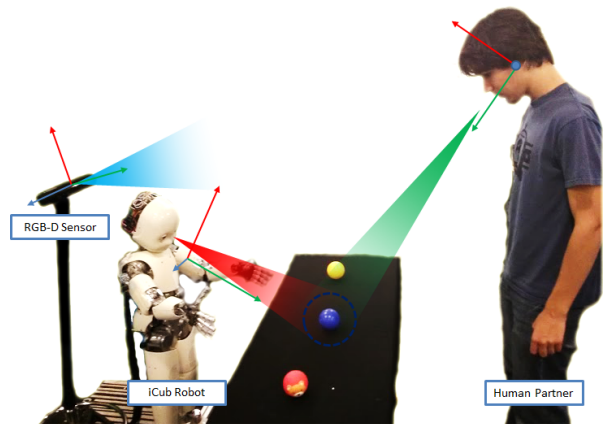

Fig. 9: The iCub robot learns about objects colors from a human partner in a tutoring scenario.

the robot tried to engage the human partner during a specific task. The first case study is a social teaching experiment, where a human teaches the color of some objects to the iCub humanoid robot [39]. In the second case study, the Nao robot tries to elicit behaviors on children affected by autism spectrum disorder and on children in typical development. This section presents the two studies and report on the results that were obtained applying our evaluation methods to discriminate between behaviors from different conditions in the task and different groups.

3.1 Interactive Behaviors Assessment

In this scenario, the robot interacts with a human partner to improve its knowledge about the environment. The two peers stand in front of each other, as shown in Figure 9. The robot can interrogate the human about the objects on the table, to discover their color properties. A simple speech recognition system, based on a fixed dictionary [59], is used to retrieve the verbal information from the human. The match between color information and object is possible thanks to the shared attention system: the robot is capable to select, among the different objects, the one currently observed by the human. The ability of the robot to retrieve the focus of attention of the human is based on the estimation of the partner's head orientation provided by the head tracking system. Remarkably, the head orientation is not only used for the post-experiment evaluation of the interaction, but it is used in runtime to provide to the robot control system the information about the gaze of the human partner. This way, the robot can gaze at the same direction.

Teaching phase

RI

Green.

I don't know this object.   What is its color?

HI

Green.

Please, choose an object.   What is its color?

Verification phase

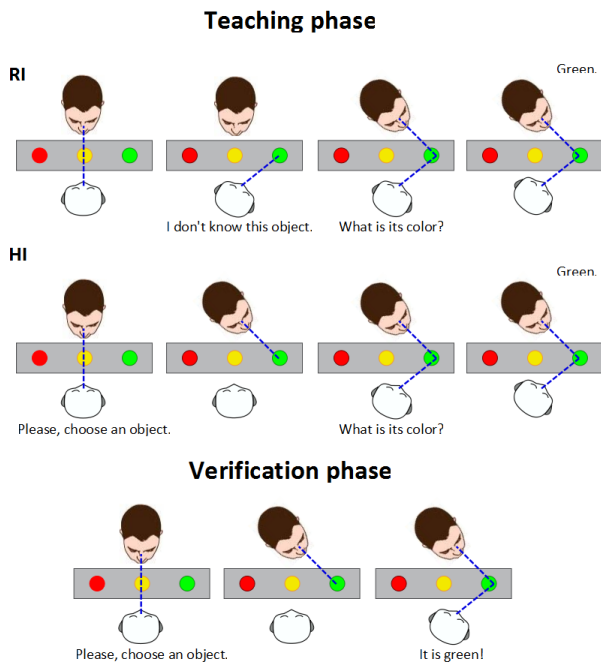Please, choose an object.   It is green!

Fig. 10: A schematic representation of the experimental protocol to study the effect of the robot initiative in a tutoring scenario. The teaching phase changes according to the partner that begins the interaction: robot initiative (RI) or human initiative (HI). In the verification phase the robot always asks the human to chose an object.

In this tutoring scenario, described in detail in [39], the authors investigated whether the initiative of the robot could produce an effect in the engagement of the human partner. The experiments consisted in a teaching phase, where the robot had to learn the colors of all the objects, and a verification phase where it had to tell to the human the colors of all the objects. The authors manipulated the robot initiative in the teaching phase, as shown in Figure 10.

Two conditions were tested. In a first condition (RI) the robot initiates the interaction by selecting an object, gazing at it, and interrogating the human about its properties. In the second condition (HI) the human decides which object to teach, by gazing at it once the robot is ready. The experiments were performed by 13 participants without previous interactions with the robot: 7 people (26 ± 3 years old) in the RI case, 6 people (22 ± 1 years old) in the HI case.

Head movements have been analyzed with the methods discussed in the previous Section. Figure 11 shows samples of the estimated gaze of some participants. Both static and dynamic features related to the validation stage have been retrieved. The static analysis of the gaze shows four hot spots in both conditions. These
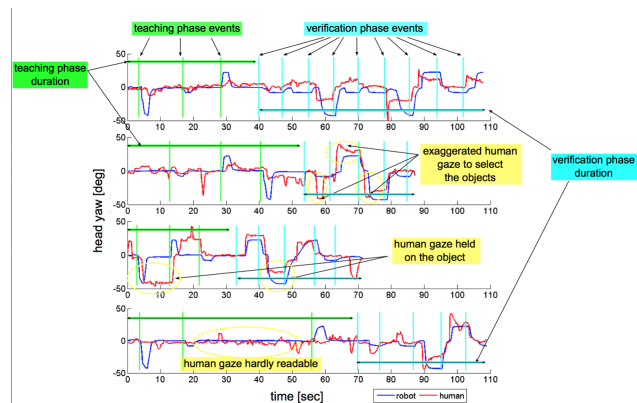

Fig. 11: Examples of gaze behaviors during the experiments. The superposition of human and robot gaze is used to study the reaction time to the robot's attention stimuli. Each vertical bar marks the beginning of a new interaction.

hot spots correspond to the head gazing over the robot and over the three objects placed on the table. The differences between the two conditions are highlighted by the clustering of the data using the k-means algorithm, as shown in Figure 12.

The dynamic analysis of the head movements show statistical relevant differences between the two groups: the reaction time in response to the robot attention stimuli over an object in the verification stage is faster if the robot initiates the interaction ($p<0.005$), rather than if the human initiates. This result is confirmed ($p<1.5e-5$) by the analysis of the pace of the interaction, the time interval between consecutive robots attention stimuli during the verification stage. The pace is faster if robot manifests proactive behaviors, initiating the interaction.

3.2 Autism Assessment

The proposed evaluation methodology has been used in an interactive scenario to match differences between children affected by Autism Spectrum Disorder (ASD) and children in Typical Development (TD). In this assessment scenario, described in detail in [6], a robot is placed in front of the child and used as an instrument to elicit joint attention. As shown in Figure 13, two images of a cat and of a dog conveniently placed on the environment are used as targets for the attention for the two peers. The RGB-D sensor provides to the robot the capability to look at the child and, at the same time; it stores all the information related to the behavior of the children, paired and synchronized with the movements of the robot.

(a) Human leader condition
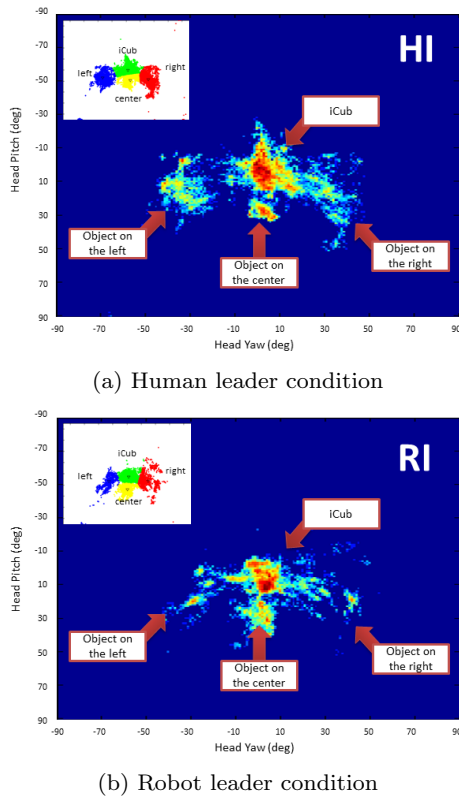

(b) Robot leader condition

Fig. 12: The heat maps of the human gaze (head yaw on X-axis, pitch on Y-axis) in the two conditions (HI and RI) highlights differences in the human gazing behavior. We can observe four different areas of focus of attention: the robot (in front of the human) and the three objects. Their location was chosen to conveniently highlight the three areas of gaze.
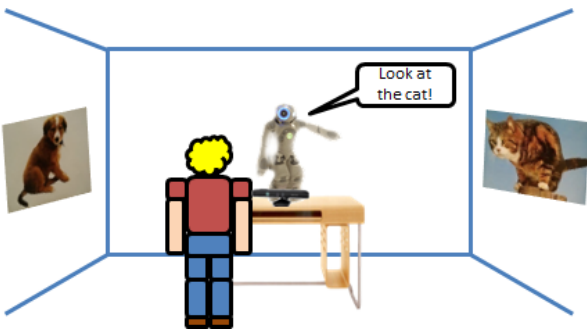

Fig. 13: A Nao robot tries to elicit joint attention over two focus of attention in an interactive scenario.

The experiment is composed by three stages in which the robot tries to induce joint attention increasing the informative content it provides to the human. In the first stage the robot gazes over the two focuses of attention; then it gazes and points over them; finally it
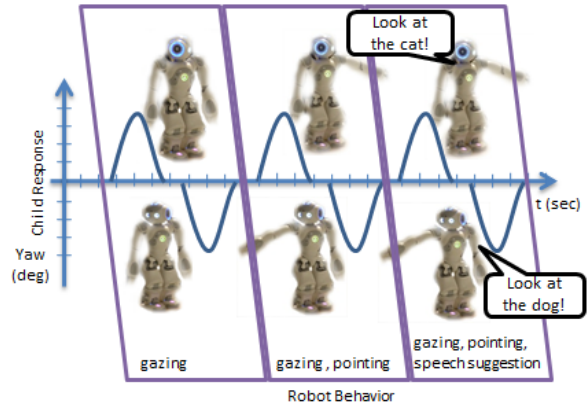

Fig. 14: In the experimental protocol, the robot tries to elicit joint attention in children in different conditions that mix multimodal social cues: gazing, pointing and vocalization.

gazes, points and vocalizes "look at the cat", "look at the dog", as shown in Figure 14.

Thirty-two children have been chosen for the experiments:

- Group ASD: 16 people (13 males, 5 females), 9.25 ± 1.87 years old.
- Group TD: 16 people (9 males, 6 females), 8.06 ± 2.49 years old.

In this case, head movements and posture have been analyzed and compared between the two groups. Using generalized linear mixed models, we found a significant higher variance of the yaw movements in TD children rather than in the children with ASD (b = 1.66, p = 0.002). The analysis showed also a significant effect on the yaw movements in accord to the induction modalities used to stimulate Joint Attention: higher variance has been found in vocalizing + pointing compared to pointing (b = 1.52, p <0.001) and compared to gazing only (b = 1.55, p <0.001). At the same time, pitch movements analysis revealed a lower variance in TD children (b = -0.84, p = 0.019) rather than children affected by ASD.

As highlighted in Figure 15, both the heat maps of the head pitch and yaw movements show a central hot spot: this area represents the gaze of the child towards the robot. The two lobes corresponding to the two focuses of attention on the sides of the room are less highlighted in ASD children rather than in TD children. An analysis of the clusters obtained using k-means on the TD children data shows that both left and right directions gathered 30.2% of all the occurrences. Applying the same k-means model to ASD children data shows that left and right represented just 8.72% of all the occurrences (Fisher's exact test, p = 2.2×10^{-16}): during

(a) Head movements in TD condition
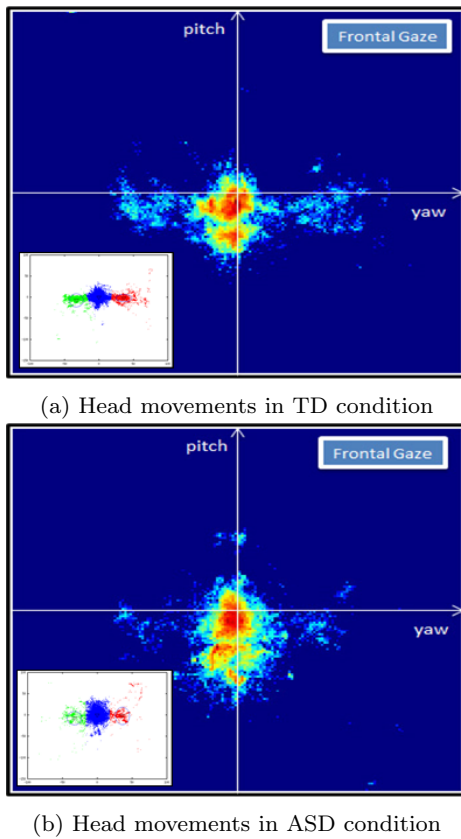

(b) Head movements in ASD condition

Fig. 15: The heat maps of the children head yaw (on X-axis) and pitch (on Y-axis) in the two conditions highlights differences on their behavior: ASD children present a lower response to the elicitation and less stability of their gazing towards the robot.

the Joint Attention task, TD children gazed over the focus of attention placed on the room 4.6 times more frequently during than the children with ASD (95% Confidence; Interval: 4.4-4.6). Those results highlight an higher response to the robot's elicitation by children in typical development, while less stability on the gazing is found in ASD children.

A similar analysis has been performed using the body pose, Figure 16, and body posture data, Figure 17. In particular, the displacements of each child from the zero position shown a higher stability in TD children: using multivariate regression, the pose variance was significantly lower than in ASD children, within all axes (x, estimates = 28.1, p = 0.001; y, estimates = 7, p = 0.006; z, estimates = 12, p = 0.009). A similar behavior has been found from the analysis of body posture data, considering the pitch and the yaw of the trunk. Also, in this case, ASD children data results less stable than TD children data: posture variance was significantly lower in the TD children than in ASD children, within all axes (x, estimates = 13.9, p = 0.0016; y, estimates =

(a) Trunk pose in TD condition
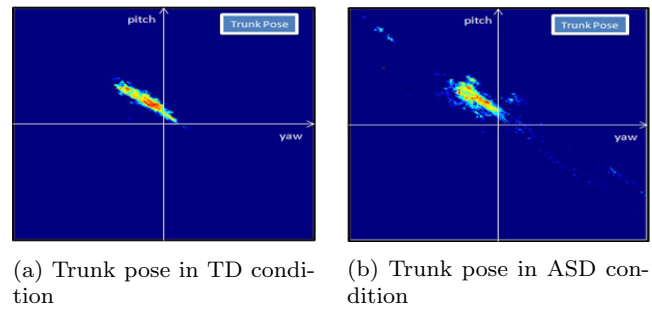
(b) Trunk pose in ASD condition

Fig. 17: Heat maps of the body pose of the children in the two conditions highlight differences among the two groups: ASD children posture is less stable than in TD children.

9.2, p = 0.016; z, estimates = 1.6, p = 0.003). Such results highlight lower stability of the body posture in ASD children rather than in TD children.

4 Discussion

We proposed in this paper a methodology for analyzing the engagement between humans and social robots in direct, face-to-face, interactive scenarios. The methodology is based on an evaluation of the engagement aroused by a robot during the interaction, focusing on the nonverbal behaviors expressed by the human partner. Both static and dynamic interaction cues have been considered, as they can be used to extract different meaningful measures. The ones described in Section 2 were able to characterize different aspects of the social interaction between humans and robots in two different use cases.

In both scenarios, the human and the robot established a mutual communication. In such contexts, a correct comprehension and proper use of nonverbal behaviors are essential tools to achieve an "optimal" interaction: to provide readable behaviors, and to arouse on human partners the illusion of a social intelligence. The importance of nonverbal behaviors has been highlighted by developmental sciences [50]. Toddlers learn about the world in a social way. They develop communication skills through nonverbal cues, and such skills gradually evolve together with verbal language [70]. Imitation, joint-attention, gesticulation, synchrony are all learned in the very first stages of childhood development, and seem to be pivotal traits of the developmental process [69,43]. In adulthood, those become semi-automatic, almost involuntary behaviors, influenced by the culture, used in daily communications, eventually in combination with spoken language to reinforce it or to completely alter its meaning.

(a) Displacement in TD condition (front)

(b) Displacement in ASD condition (front)

(c) Displacement in TD condition (top)
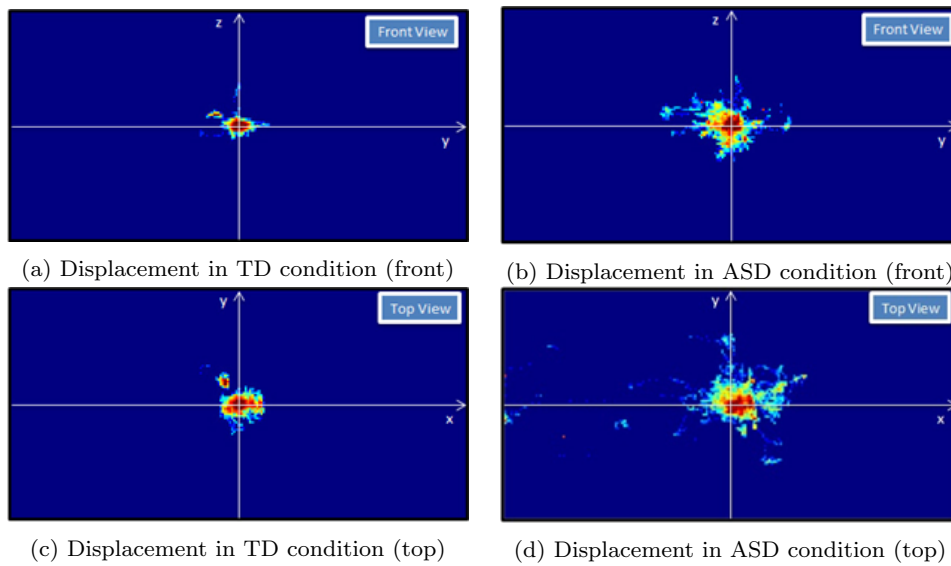
(d) Displacement in ASD condition (top)

Fig. 16: Heat maps of the trunk displacement from the zero position of the children in the two conditions highlight differences among the two groups: ASD children's position in the space is less stable than in TD children.

The measurement of nonverbal signals during interactions with robots can provide information about the engagement between pairs [27]. The static analysis of the movements of the body articulations can reveal how humans respond to the robot stimuli, if they respond as engaged partners or not. The analysis of the gaze behavior can be used to model the attention system of the human partner and improve joint attention. The dynamic analysis can be used to study the motor resonance, synchrony of movements, and can improve imitations and gestures. A robot capable to capture the attention of the human partner should leverage all those nonverbal cues to increase the engagement.

4.1 A practical set of measures

Similar measures can be retrieved using motion capture systems. However, usually such systems use marker based technologies: they require passive or active beacons that should be worn by the user. This do not only increase the complexity of the system, but they critically reduces the naturalness of the interaction. The proposed system, instead, is based on a simple RGB-D camera, a marker-less technology that can be still used to track human movement [33, 4]. Despite its lower resolution, this system allows researchers to explore the engagement in very natural scenarios, without the restrictions and the complexity imposed by wearable devices and marker holders.

While such measures have been developed to enable studies in naturalistic settings, those can be aggregated

with the features obtained from physiological responses of the participants in specially designed experiments during which participants would forget the existence of worn sensors, to establishing natural interactions as much as possible. In such case, it would be possible to capture a larger dynamic of possible interactions and, at the same time, to study the neurophysiological bases of the engagement [28, 76].

Several researches in social robotics make use of post-experiment questionnaires to gather information about the engagement after the experiments [42, 41]. Unfortunately, while quick and easy to analyze, questionnaires can be strongly affected by several kind of biases [22]. Without being exhaustive, it is possible to find at least three important sources of errors in questionnaires: their design, the experimental subjects, and the experimenter. The design of the questionnaire can introduce artifacts due to complexity, ambiguity or specificity of the questions, or due to the number, too few or too many, of the answers' options (question wording [56]). The subjects can also introduce errors, because of their unconscious will to be positive experimental subjects and to provide socially desirable responses (response bias [32]). Lastly, the researchers can also be a source of error with their tendency to interpret the answers as a confirmation of their hypothesis (confirmation bias [53]). The measures presented in this paper can be used as a practical and objective tool to explore the interaction with robots; they can also serve as a complement to verify and eventually to reinforce the results obtained by questionnaires and surveys.

4.2 Readability of robot's nonverbal cues

As social animals, humans are extraordinarily able to infer information about their partners, and to build models of the other and of their society. Nonverbal behaviors play a central role for making this inference [29].

In the first scenario, the presented metrics have been used to show that humans react faster to the attention cues of a proactive robot. It is possible to speculate about the manipulation of the proactive behavior of the robot to strengthen the engagement, regulate the rhythm of interaction, and arouse in people the perception of social intelligence. The engagement, here, comes essentially from the readability provided by the nonverbal cues. This result is confirmed also in the experiment with children affected by ASD and TD children: a significant difference in the two groups has been found according to the amount of information expressed by the nonverbal language of the robot. The more modalities the robot uses to communicate its focus of attention (from gazing, to gazing and pointing, to vocalizing), the more its behavior becomes readable by the children.

The results obtained in the two case studies confirm that the proposed measures are effective to study the engagement. These metrics can be used by the robot as a continuous, online feedback signal to evaluate (and eventually manipulate) the engagement with the human partner [5].

Future studies, however, will focus on the use of the presented metrics in long term scenarios, in which the novelty effect of the robot became less relevant with time. In such settings people will interact day-by-day with the robot becoming accustomed to its behaviors; at the same time, human subjects could adapt their own behaviors to the robot.

4.3 The "bias" of the anthropomorphic design

People have the natural tendency of projecting human-like features on animals and inanimate objects. This is the so called "anthropomorphism" [49]: as "social machines", we seek in the unknown the same intelligence patterns we are used to recognize in our peers, projecting our social intelligence. The robot is not perceived as a machine: people have frequently the illusion that the robot understands them, needs their help and wants to communicate. During an interaction, the human can naturally develop a feeling a "partnership" with the robot [15]. The anthropomorphic design of the robots can help the readability of their behaviors, facilitating the interaction with human partners [16].

The robots used in our experiments, iCub and Nao, have a baby-like, humanoid shape, which makes them particularly suited for interaction but also introduces an anthropomorphism bias in their human partners. These robots communicate implicitly, just with their design, very human-like traits such as a personality, emotions and intention, and arouse a sense of co-presence [78]. The presented metrics can be used to study the perceived engagement with other types of robots. They should be as well able to highlight differences due to different types of robot, even if it is difficult to make predictions about the human reaction to non-humanoid robots or "headless" robots. It would be very interesting to see if our results with humanoids hold in the case of androids and very human-like robots. A relational analysis with respect to the uncanny valley could not be so quite obvious [36, 37]. Our intuition is that the presented metrics should be able to highlight the different reactions and behaviors of the human partners in all cases, but it is difficult to imagine how much the results will diverge in similar experiments involving androids (maybe revealing aversion effects).

We plan future experiments where we will use the proposed metrics to assess the engagement between humans and different types of robots. Since the robot design practically impacts the span of their behaviors, we will carefully ponder such a study, considering the limits and capabilities of each robot and evaluating their "social intelligence" on comparable tasks and desired behaviors.

4.4 Are we measuring social intelligence?

Explaining the concept of "intelligence" is a non-trivial problem [44]. Intelligence could be intuitively associated to the ability of humans to understand the world. However, this definition still lacks of generality, due to the observation of certain kinds of intelligence in the living. The idea of intelligence in humans is context-dependent. The psychometric approach to human intelligence provides a definition according to three points of view [68]: the abstract intelligence, as the ability of understanding and managing ideas and symbols; the mechanical intelligence, as the capability of working with concrete objects; the social intelligence [20], as "ability to get along with others" [73].

These definitions can be also employed in robotics, with an interesting parallelism [25, 26]. The abstract intelligence can be identified with the capability of the robots of learning, reasoning using symbols, exploring the knowledge and deducing new facts. This roughly corresponds to the area of "Artificial Intelligence". We can refer mechanical intelligence to the perceptuo-motor intelligence or body intelligence, the ability to interact

with the physical world, to perceive it and to coordinate proper actions on it. This kind of intelligence comes from the robot embodiment. The robot should be able to capture the relevant information from the world and to link them to high level, abstract, symbols. Reasoning on these symbols should take into account the physical possibilities and capabilities provided by the robot's embodiment in its physical world.

Finally, social intelligence would refer to the ability of robots to act socially with humans, to communicate and interact with them in a human-like way, following social behaviors and rules attached to their role, to learn and adapt their own behaviors throughout their lifetime, incorporating shared experiences with other individuals into their understanding of self, of others, and of the relationships they share [16]. The domain of "Social Signal Processing" [74] aims to provide computers and robots with social intelligence, addressing to a correct perception, accurate interpretation and appropriate display of social signals [75].

Expressing social intelligence is a key feature to achieve an optimal interaction with humans. The perception of "social intelligence" can be aroused if the robot is capable to exhibit social cues: accomplishing coherent behaviors, following social rules and communicating with the humans in natural way.

The proposed methodology focuses on the analysis of non-verbal human's behaviors during interaction with robots. We can speculate about the use of the presented metrics as a feedback of the social intelligence perceived by the humans. From this point of view, responsive behaviors produced by a robot will induce in the human partners a perception of intelligence that can be quantitatively captured by the proposed measures by observing changes to the human's reactions to the robot social cues. This interpretation comes out from the experiments we discussed.

In the first experiment, it is possible to speculate about the social intelligence expressed by the robot and perceived by the human partner, according to slight changes on the plot of the interaction. Questionnaires given to the participants of the experiments, reported as more intelligent the robot that initiates the interaction. In our view this can be attributed to the increased readability of the proactive case, which makes the human "aware" of the robot status and creates the illusion of a greater intelligence than in the other case. This illusion could be one of the reasons for the human interact notably faster.

The second experiment is remarkable since Autism is characterized by a lack of social intelligence [8,19]. Here, the behaviors shown by the robot and, at the same time, the plot of the interaction do not vary be-

tween the two conditions, so the differences are due to the different ability of the children to recognize social cues. The proposed metrics to evaluate the engagement highlight the lacking of social intelligence in the ASD children, showing behavioral differences between the two groups.

4.5 Conclusions

In this paper a set of metrics has been proposed to evaluate the engagement between humans and robots in direct, face to face scenarios. Those metrics have been applied to study the interaction in two different use cases, characterized by natural settings and different objectives, and to assess effectively different human responses to robot behaviors. In both the scenarios, the metrics confirmed the importance of the study of non-verbal cue to improve the interactions between humans and robots. Nevertheless, thanks to their easiness of use in real world scenarios, due to employment of non-intrusive sensors, such metrics present a strong potential for scalability and a further generalization to different applications and contexts. Limitations of the metrics would be studied in future works, in particular in long-term scenarios, in which human subjects will be accustomed to the behaviors of the robot, and according to the use of different robotic designs, anthropomorphic and not.

References

1. Admoni, H., Dragan, A., Srinivasa, S.S., Scassellati, B.: Deliberate delays during robot-to-human handovers improve compliance with gaze communication. In: Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction, HRI '14, pp. 49–56 (2014)
2. Andry, P., Blanchard, A., Gaussier, P.: Using the rhythm of nonverbal human-robot interaction as a signal for learning. IEEE Transactions on Autonomous Mental Development **3**(1), 30–42 (2011)
3. Anzalone, S.M., Chetouani, M.: Tracking posture and head movements of impaired people during interactions with robots. In: New Trends in Image Analysis and Processing–ICIAP 2013, pp. 41–49. Springer (2013)
4. Anzalone, S.M., Ghidoni, S., Menegatti, E., Pagello, E.: A multimodal distributed intelligent environment for a safer home. In: Intelligent Autonomous Systems 12, pp. 775–785. Springer (2013)

5. Anzalone, S.M., Ivaldi, S., Sigaud, O., Chetouani, M.: Multimodal people engagement with icub. In: Biologically Inspired Cognitive Architectures 2012, pp. 59–64. Springer (2013)

6. Anzalone, S.M., Tilmont, E., Boucenna, S., Xavier, J., Jouen, A.L., Bodeau, N., Maharatna, K., Chetouani, M., Cohen, D.: How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3d+ time) environment during a joint attention induction task with a robot. Research in Autism Spectrum Disorders **8**(7), 814–826 (2014)

7. Argall, B.D., Browning, B., Veloso, M.: Teacher feedback to scaffold and refine demonstrated motion primitives on a mobile robot. Robotics and Autonomous Systems **59**(3-4), 243–255 (2011)

8. Baron-Cohen, S.: Mindblindness: An essay on autism and theory of mind. MIT press (1997)

9. Bertenthal, B.I., Boyer, T.W., Han, J.M.: Social attention is not restricted to the eyes: Pointing also automatically orients direction of attention (2012)

10. Boucenna, S., Anzalone, S., Tilmont, E., Cohen, D., Chetouani, M.: Learning of social signatures through imitation game between a robot and a human partner. Autonomous Mental Development, IEEE Transactions on (2014)

11. Boucenna, S., Gaussier, P., Andry, P., Hafemeister, L.: A robot learns the facial expressions recognition and face/non-face discrimination through an imitation game. International Journal of Social Robotics **6**(4), 633–652 (2014)

12. Boucenna, S., Narzisi, A., Tilmont, E., Muratori, F., Pioggia, G., Cohen, D., Chetouani, M.: Interactive technologies for autistic children: A review. Cognitive Computation pp. 1–19 (2014)

13. Breazeal, C.: Toward social robots. Robotics and Autonomous Systems **42**, 167–175 (2003)

14. Breazeal, C., Kidd, C.D., Thomaz, A.L., Hoffman, G., Berlin, M.: Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 383–388 (2005)

15. Breazeal, C.L.: Sociable machines: Expressive social exchange between humans and robots. Ph.D. thesis, Massachusetts Institute of Technology (2000)

16. Breazeal, C.L.: Designing sociable robots. MIT press (2004)

17. Brethes, L., Menezes, P., Lerasle, F., Hayet, J.: Face tracking and hand gesture recognition for human-robot interaction. In: IEEE International Conference on Robotics and Automation, vol. 2, pp. 1901–1906 (2004)

18. Brick, T., Scheutz, M.: Incremental natural language processing for hri. In: ACM/IEEE International Conference on Human-robot Interaction, HRI '07, pp. 263–270. ACM, New York, NY, USA (2007)

19. Bruner, J., Feldman, C.: Theories of mind and the problems of autism (1993)

20. Cantor, N., Kihlstrom, J.F.: Personality and social intelligence. Prentice-Hall Englewood Cliffs, NJ (1987)

21. Cantrell, R., Scheutz, M., Schermerhorn, P., Wu, X.: Robust spoken instruction understanding for hri. In: 5th ACM/IEEE International Conference on Human-Robot Interaction, pp. 275–282 (2010)

22. Choi, B.C., Pak, A.W.: A catalog of biases in questionnaires. Prev Chronic Dis **2**(1), A13 (2005)

23. Crespi, N., Molina, B., Palau, C., et al.: Qoe aware service delivery in distributed environment. In: Advanced Information Networking and Applications (WAINA), 2011

24. Cristinacce, D., Cootes, T.: Feature detection and tracking with constrained local models. In: Proc. British Machine Vision Conference, vol. 3, pp. 929–938 (2006)

25. Dautenhahn, K.: Getting to know each otherartificial social intelligence for autonomous robots. Robotics and autonomous systems **16**(2), 333–356 (1995)

26. Dautenhahn, K.: Socially intelligent robots: dimensions of human–robot interaction. Philosophical Transactions of the Royal Society B: Biological Sciences **362**(1480), 679–704 (2007)

27. Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., Cohen, D.: Interpersonal synchrony : A survey of evaluation methods across disciplines. IEEE Transactions on Affective Computing **3**(3), 349– 365 (2012)

28. Delaherche, E., Dumas, G., Nadel, J., Chetouani, M.: Automatic measure of imitation during social interaction: A behavioral and hyperscanning-eeg benchmark. Pattern Recognition Letters (2014)

29. Ekman, P., Friesen, W.V.: The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Nonverbal communication, interaction, and gesture pp. 57–106 (1981)

30. Fischer, K., Lohan, K., Saunders, J., Nehaniv, C., Wrede, B., Rohlfing, K.: The impact of the contingency of robot feedback on hri. In: Collaboration Technologies and Systems (CTS), 2013 International Conference on, pp. 210–217. IEEE (2013)

31. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Robotics and Autonomous Systems **42**, 143 – 166 (2003)

32. Furnham, A.: Response bias, social desirability and dissimulation. Personality and individual differences **7**(3), 385–400 (1986)

33. Ghidoni, S., Anzalone, S.M., Munaro, M., Michieletto, S., Menegatti, E.: A distributed perception infrastructure for robot assisted living. Robotics and Autonomous Systems (2014)

34. Hall, J., Tritton, T., Rowe, A., Pipe, A., Melhuish, C., Leonards, U.: Perception of own and robot engagement in human-robot interactions and their dependence on robotics knowledge. Robotics and Autonomous Systems **62**(3), 392 – 399 (2014)

35. Harris, T.K., Banerjee, S., Rudnicky, A.I.: Heterogeneous multi-robot dialogues for search tasks (2005)

36. Ishiguro, H.: Interactive humanoids and androids as ideal interfaces for humans. In: Proceedings of the 11th international conference on Intelligent user interfaces, pp. 2–9. ACM (2006)

37. Ishiguro, H.: Android science. In: Robotics Research, pp. 118–127. Springer (2007)

38. Ishii, R., Shinohara, Y., Nakano, T., Nishida, T.: Combining multiple types of eye-gaze information to predict users conversational engagement. 2nd workshop on eye gaze on intelligent human machine interaction (2011)

39. Ivaldi, S., Anzalone, S.M., Rousseau, W., Sigaud, O., Chetouani, M.: Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement. Frontiers in Neurorobotics **8**(5) (2014)

40. Ivaldi, S., Nguyen, S., Lyubova, N., Droniou, A., Padois, V., Filliat, D., Oudeyer, P.Y., Sigaud, O.: Object learning through active exploration. IEEE Transactions on Autonomous Mental Development **6**(1), 56 – 72 (2014)

41. Kamide, H., Mae, Y., Kawabe, K., Shigemi, S., Hirose, M., Arai, T.: New measurement of psychological safety

for humanoid. In: Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, pp. 49–56. ACM (2012)

42. Kamide, H., Mae, Y., Takubo, T., Ohara, K., Arai, T.: Development of a scale of perception to humanoid robots: Pernod. In: Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on, pp. 5830–5835. IEEE (2010)

43. Kaplan, F., Hafner, V.: The challenges of joint attention (2004)

44. Kihlstrom, J.F., Cantor, N.: Social intelligence. Handbook of intelligence **2**, 359–379 (2000)

45. Kulic, D., Croft, E.A.: Affective state estimation for human–robot interaction. Robotics, IEEE Transactions on **23**(5), 991–1000 (2007)

46. Laghari, K.U.R., Connelly, K.: Toward total quality of experience: A qoe model in a communication ecosystem. Communications Magazine, IEEE **50**(4), 58–65 (2012)

47. Lee, C., Lesh, N., Sidner, C.L., Morency, L.P., Kapoor, A., Darrell, T.: Nodding in conversations with a robot. In: CHI'04 Extended Abstracts on Human Factors in Computing Systems, pp. 785–786. ACM (2004)

48. Lee, J., Chao, C., Bobick, A.F., Thomaz, A.L.: Multi-cue contingency detection. International Journal of Social Robotics **4**(2), 147–161 (2012)

49. Lemaignan, S., Fink, J., Dillenbourg, P.: The dynamics of anthropomorphism in robotics. In: Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, pp. 226–227. ACM (2014)

50. Miller, P.H.: Theories of developmental psychology. Macmillan (2010)

51. Mower, E., Feil-Seifer, D.J., Mataric, M.J., Narayanan, S.: Investigating implicit cues for user state estimation in human-robot interaction using physiological measurements. In: Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on, pp. 1125–1130. IEEE (2007)

52. Natale, L., Nori, F., Metta, G., Fumagalli, M., Ivaldi, S., Pattacini, U., Randazzo, M., Schmitz, A., G. Sandini, G.: Intrinsically motivated learning in natural and artificial systems, chap. The iCub platform: a tool for studying intrinsically motivated learning. Springer-Verlag (2012)

53. Nickerson, R.S.: Confirmation bias: A ubiquitous phenomenon in many guises. Review of general psychology **2**(2), 175 (1998)

54. Obhi, S.S., Sebanz, N.: Moving together: toward understanding the mechanisms of joint action. Experimental Brain Research **211**(3), 329–336 (2011)

55. O'Brien, H.L., Toms, E.G.: What is user engagement? a conceptual framework for defining user engagement with technology. Journal of the American Society for Information Science and Technology **59**(6), 938–955 (2008)

56. Payne, S.L.: The art of asking questions. (1951)

57. Raake, A., Egger, S.: Quality and quality of experience. In: Quality of Experience, pp. 11–33. Springer (2014)

58. Rich, C., Ponsler, B., Holroyd, A., Sidner, C.L.: Recognizing engagement in human-robot interaction. In: Proc. ACM/IEEE Int. Conf. Human-Robot Interaction (HRI), pp. 375–382 (2010)

59. Rousseau, W., Anzalone, S.M., Chetouani, M., Sigaud, O., Ivaldi, S.: Learning object names through shared attention. In: Int. Conf. on Intelligents Robots and Systems - Workshop on Developmental Social Robotics (2013)

60. Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P.W., Paiva, A.: Automatic analysis of affective postures and body motion to detect engagement with a game companion. In: 6th ACM/IEEE International Conference on Human-Robot Interaction, pp. 305–311 (2011)

61. Scassellati, B.: Quantitative metrics of social response for autism diagnosis. In: Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on, pp. 585–590. IEEE (2005)

62. Scassellati, B.: How social robots will help us to diagnose, treat, and understand autism. In: Robotics research, pp. 552–563. Springer (2007)

63. Short, E., Hart, J., Vu, M., Scassellati, B.: No fair!! an interaction with a cheating robot. In: 5th ACM/IEEE International Conference on Human-Robot Interaction, pp. 219–226 (2010)

64. Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., Moore, R.: Real-time human pose recognition in parts from single depth images. Communications of the ACM **56**(1), 116–124 (2013)

65. Sidner, C., Lee, C., Kidds, C., Lesh, N., Rich, C.: Explorations in engagement for humans and robots. Artificial Intelligence (2005)

66. Sidner, C.L., Kidd, C.D., Lee, C., Lesh, N.: Where to look: A study of human-robot engagement. In: Proceedings of the 9th International Conference on Intelligent User Interfaces, pp. 78–84 (2004)

67. Tapus, A., Mataric, M., Scasselati, B.: Socially assistive robotics [grand challenges of robotics]. Robotics Automation Magazine, IEEE **14**(1), 35–42 (2007)

68. Thorndike, E.L.: Intelligence and its uses. Harper's magazine (1920)

69. Tomasello, M.: Joint attention as social cognition. Joint attention: Its origins and role in development pp. 103–130 (1995)

70. Tomasello, M., Farrar, M.J.: Joint attention and early language. Child development pp. 1454–1463 (1986)

71. Vaussard, F., Fink, J., Bauwens, V., Retornaz, P., Hamel, D., Dillenbourg, P., Mondada, F.: Lessons learned from robotic vacuum cleaners entering the home ecosystem. Robotics and Autonomous Systems **62**(3), 376 – 391 (2014)

72. Vázquez, M., May, A., Steinfeld, A., Chen, W.H.: A deceptive robot referee in a multiplayer gaming environment. In: Collaboration Technologies and Systems (CTS), 2011 International Conference on, pp. 204–211. IEEE (2011)

73. Vernon, P.E.: Some characteristics of the good judge of personality. The Journal of Social Psychology **4**(1), 42–57 (1933)

74. Vinciarelli, A., Pantic, M., Bourlard, H.: Social signal processing: Survey of an emerging domain. Image and Vision Computing **27**(12), 1743–1759 (2009)

75. Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., Schröder, M.: Bridging the gap between social animal and unsocial machine: A survey of social signal processing. Affective Computing, IEEE Transactions on **3**(1), 69–87 (2012)

76. Weisman, O., Delaherche, E., Rondeau, M., Chetouani, M., Cohen, D., Feldman, R.: Oxytocin shapes parental motion during father–infant interaction. Biology letters **9**(6), 20130,828 (2013)

77. Yannakakis, G.N., Hallam, J., Lund, H.H.: Entertainment capture through heart rate activity in physical interactive playgrounds. User Modeling and User-Adapted Interaction **18**(1-2), 207–243 (2008)

78. Zhao, S.: Toward a taxonomy of copresence. Presence: Teleoperators and Virtual Environments **12**(5), 445–455 (2003)

79. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Comput. Surv. **35**(4), 399–458 (2003)

5 Biographies

5.1 Salvatore Maria Anzalone

Salvatore M. Anzalone received his B.S. in Computer Engineering (2004) and M.S.Degree in Intelligent Systems (2006) from the University of Palermo. From the same University he received his Ph.D. degree in Computer Engineering (2010), guided by Prof. A. Chella and Prof. R. Sorbello. He cooperated with Prof. H. Ishiguro at the Osaka University during one year of his Ph.D. (2009) and during his post-doc year at University of Padova (2010), guided by Prof. E. Pagello and Prof. E. Menegatti. He received a grant from the Japan Society for the Promotion of Science and from April 2011 to April 2012 was a research fellow at the Intelligent Robotics Laboratory directed by Prof. H. Ishiguro at Osaka University (2011-2012). From May 2012 he is a Post-Doc researcher at University Pierre et Marie Curie in Paris, France, working with Prof. M. Chetouani at the Institute for Intelligent Systems and Robotics (ISIR-UPMC). In November 2013 he visited again the laboratory of Prof. H. Ishiguro at the Osaka University thanks to a grant received by the Japan's National Institute of Information and Communications Technology. His research focuses on Personal Robots, in particular on the modelization of human's identities and behaviours, to give to the robots the ability of interacting with humans in a more strict, customized and reliable way. Last experiences at ISIR-UPMC in Paris involve the use of personal robots for autistic children in the FP7 European Project Michelangelo.

5.2 Sofiane Boucenna

Sofiane Boucenna is Associate Professor at Cergy-Pontoise University in ETIS laboratory. He obtained its PhD at the Cergy Pontoise University in France in 2011, where he worked with the Neurocybernetic team of the Image and Signal processing Lab (ETIS). Then, (2012-2014) he worked as post-doc at Pierre et Marie Curie University in the Institut des Systmes Intelligents et de Robotique lab (ISIR). His research interests are focused on the modelling of cognitive mechanisms and the development of interaction capabilities such as imitation, emo-

tion and social referencing. Currently, he attempts to assess the effect of the type of partners (adults, typically developing children and children with autism spectrum disorder) on robot learning.

5.3 Serena Ivaldi
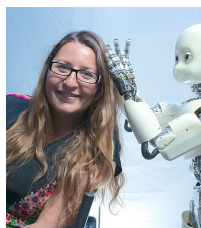
Serena Ivaldi received her PhD in Humanoid Technologies in 2011, jointly at the University of Genoa and Italian Institute of Technology. There she also held a research fellowship in the Robotics, Brain and Cognitive Sciences Department, under the supervision of Giorgio Metta. From 2011 to 2014 she was a postdoctoral researcher in ISIR, University Pierre et Marie Curie, then in the Intelligent Autonomous Systems Lab in TU Darmstadt, as a postdoctoral researcher, under the supervision of Jan Peters. Since November 2014 she is a researcher of INRIA in Nancy. Her research focuses on robots interacting physically and socially with humans and environment. She is also interested in developmental robotics, blending learning, perception and control.

5.4 Mohamed Chetouani

Mohamed Chetouani is the head of the Artificial Perception research group. He received the M.S. degree in Robotics and Intelligent Systems from the University Pierre and Marie Curie (UPMC), Paris, 2001. He received the PhD degree in Speech Signal Processing from the same university in 2004. In 2005, he was an invited Visiting Research Fellow at the Department of Computer Science and Mathematics of the University of Stirling (UK). Dr. Chetouani was also an invited researcher at the Signal Processing Group of Escola Universitaria Politecnica de Mataro, Barcelona (Spain). He is currently an Associate Professor in Signal Processing and Pattern Recognition at the University Pierre et Marie Curie. His research activities, carried out at the Institute of Intelligent Systems and Robotics, cover the areas of non-linear signal processing, feature extraction, pattern classification and fusion for human centered interaction analysis: verbal and non-verbal communication, physiological signals. He is an Associate Editor of the Cognitive Computation Journal (Springer) and

a Guest Editor for a Special Issue in Speech Communication on Non-Linear and Non-Conventional Speech Processing. He served as a chairman of several international workshops on non-linear speech processing, human-robot interaction and human centered multimodal signal processing.