Design of a Robotic Coach for Motor, Social and Cognitive Skills Training Toward Applications With ASD Children

Laura Santos, *Graduate Student Member, IEEE*, Alice Geminiani, *Member, IEEE*, Paul Schydlo, Ivana Olivieri, José Santos-Victor, *Senior Member, IEEE*, and Alessandra Pedrocchi, *Senior Member, IEEE*

***Abstract*—Socially assistive robots may help the treatment of autism spectrum disorder(ASD), through games using dyadic interactions to train social skills. Existing systems are mainly based on simplified protocols which qualitatively evaluate subject performance. We propose a robotic coaching platform for training social, motor and cognitive capabilities, with two main contributions: (i) using triadic interactions(adult, robot and child), with robotic mirroring, and (ii) providing quantitative performance indicators. The key system features were accurately designed, including type of protocols, feedback systems and evaluation metrics, contemplating the requirements for applications with ASD children. We implemented two protocols, *Robot-Master* and *Adult-Master*, where children performed different gestures guided by the robot or the adult respectively, eventually receiving feedback about movement execution. In both, the robot mirrors the subject during the movement. To assess system functionalities, with a homogeneous group of subjects, tests were carried out with 28 healthy subjects; one preliminary acquisition was done with an ASD child. Data analysis was customized to design protocol-specific parameters for movement characterization. Our tests show that robotic mirroring execution depends on the complexity and standardization of movements, as well as on the robot technical features. The feedback system evaluated movement phases and successfully estimated the completion of the exercises. Future work includes improving platform flexibility and adaptability, and clinical trials with ASD children to test the impact of the robotic coach on reducing symptoms. We trust that the proposed quantitative performance indicators extend the current state-of-the-art towards clinical usage of robotic-based coaching systems.**

***Index Terms*—Robotic coaching, embodied mirroring, robotic therapy, human-robot interaction, movement metrics.**

Manuscript received November 30, 2020; revised April 10, 2021 and May 18, 2021; accepted June 13, 2021. Date of publication June 21, 2021; date of current version June 30, 2021. This work was supported in part by the Portuguese Foundation of Science and Technology under Project SFRH/BD/145040/2019 and SFRH/BD/139092/2018, in part by the CerebNest (partnering project of the Human Brain Project) through the U.S. National Institutes of Health under Grant 1R01HD081346-01A, in part by the Subaward University of Southern California-Politecnico di Milano (USC-POLIMI) under Grant 61430868, and in part by the Fondazione Zorzi through the Project Interactive mirrOring Games wIth sOCial rObot (IOGIOCO). *(Corresponding authors: Laura Santos; Alice Geminiani.)*

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Politecnico di Milano.

Laura Santos is with the NEARLab, Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milan, Italy, and also with the Institute for Systems and Robotics, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal (e-mail: laurajoana.espinosa@polimi.it).

Alice Geminiani is with the NEARLab, Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milan, Italy, and also with the Department of Brain and Behavioral Sciences, University of Pavia, 27100 Pavia, Italy (e-mail: alice.geminiani@polimi.it).

Paul Schydlo and José Santos-Victor are with the Institute for Systems and Robotics, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal.

Ivana Olivieri is with IRCCS, Fondazione Don Carlo Gnocchi, 20148 Milan, Italy.

Alessandra Pedrocchi is with the NEARLab, Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milan, Italy.

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2021.3091320, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2021.3091320

I. Introduction

SOCIALLY assistive robots are designed to establish close and effective interactions between robots and humans with the aim of providing assistance, enhancing therapy and achieving measurable progress in convalescence, rehabilitation, learning and well-being [1]. With respect to virtual avatars, embodied robotic systems tend to elicit more spontaneous reactions and imitation behaviours from subjects [2]. This is particularly important in rehabilitation applications where movements play a role, like in protocols where the robot proposes exercises to be mirrored by subjects.

Using robotic coaches to train motor skills is becoming increasingly important for older people and children [3], [4]. For children, social robots are being used within interactive games, aimed at improving social skills, as in the case of autism spectrum disorder(ASD) [5]. Motor rehabilitation only recently started to be considered in standard therapies of autism, since social skill deficits appear to be connected to motor [6], [7] and mirroring skill impairment [8], [9]. Therefore, socially assistive robots acting as imitation coaches have been developed, training both social and motor capabilities through embodied mirroring [10], [11]. In available prototypes, the interactions are usually dyadic, between robot and subject [11]–[13]. This is a limitation for effective training of engagement and eventual translation to daily life, since the generalization from a human-robot to a human-human interaction is difficult [14]. That is why triadic interactions should be considered, in which another person is present [10], [14].

In this case, the protocol could include a first step where the robot learns exercises demonstrated by an expert, e.g.

a therapist [12]. Exercises usually involve simple arm gestures [8]; however, complexity of movements should be tailored to subject/therapy requirements, to make the game challenging and effective, avoiding frustration (too difficult tasks) or boredom (too easy tasks). In a second step, the robot shows the movement to the end-user, and invites him/her to imitate, while providing feedback on the execution [8]. Two forms of imitation can be implemented: anatomical or mirror. In the first, the subject estimates and reproduces robot actions using the anatomically congruent limb. In the latter, the subject imitates the robot as if looking at a mirror [15], [16].

One of the key requirements for these applications is the capability of the robotic system to perceive a person's movements. Systems like the Microsoft Kinect (now on referred just as Kinect) have become popular, being the 3D coordinates of the tracked subject joints directly extracted from images recorded with the camera [17]. Despite its lower accuracy, its minimum intrusiveness to the user and minimal calibration procedure make it the most used in robotic mirroring coaching applications, especially for ASD children where non-intrusiveness is mandatory [11], [18].

Evaluating the performance of subjects in executing movements is another key feature for these systems. This can be measured by assessing the quality of imitation done or the number of gestures correctly performed. In [10], a Dynamic Time Warping algorithm is used to calculate the similarity between the subject imitation and robot movement, while in [18], [19], a simple posture matching between Kinect acquired data and the robot position evaluates the imitation quality. On the other hand, the automatic identification of correct gestures can be based on the error between the measured and the expected joint angles of the subject determined by the task [12], [20], or on rules taking into account angles of the several degrees of freedom [11], [21]. The robot, then, needs to interact with the subject providing feedback on movement execution. Multiple feedback solutions can be adopted: sounds, lights or movements [18]. Vocal feedback can be general on movement correctness [13], or include specific indications on how to improve the movement [12]. Different feedback modalities can be combined to increase engagement.

Available studies in robotic mirroring do not provide systematic quantitative performance measures of the exercises, and rely on the subjective evaluation by professionals. This prevents the comparison of different strategies and the objective evaluation of participants performance. Moreover, only few studies in the robotic mirroring field evaluate the complexity of protocols. Since they are mostly pilot studies, the protocols are constructed for specific scenarios, directly with the patients, and without analysing the possible design issues emerging from the interaction with the robot (delay on the execution of the gestures by the robot, understanding of the instructions given, etc). In addition, the vast majority of the proposed protocols involve just two main actors (the robot and the subject), neglecting important cooperation aspects that are fundamental in ASD therapies.

In this work, we present the design of a platform for robotic coaching through triadic mirroring training, to be applied during therapeutic sessions with ASD children. The treatment definition joined clinical experience and needs with the literature on robotic training [11], [22]. In designing the system, we considered the following main research questions:

(i) how to design the best protocols for robotic mirroring involving triadic interactions and training motor, social and cognitive skills? (which protocols are better for mirroring? which exercises are more complex/difficult?)

(ii) how can the robot/system evaluate the movements executed by a person and provide the appropriate feedback?

(iii) which metrics are better for inter-subject overall comparison along time? (which movements are more standard among subjects and allow the comparison between them? which performance metrics can be used for a continuous evaluation of the subject?)

To address these questions, two different protocols have been designed. They have been tested on a group of healthy adults and children with more homogeneous behaviour and reactions, when compared to an alternative choice of ASD children, where the diversity of each child's specific symptoms would lead to different reactions to the system. Furthermore, we assessed our system in a clinical setting, testing it with one ASD child and one therapist. We advanced the state of the art, by developing a triadic system, where both the adult and child can participate actively, within a semantic framework, allowing the training of social and cognitive capabilities beyond just mirroring and motor skills. Moreover, we propose new quantitative measures to evaluate the level of difficulty, repeatability and mirroring of the movements chosen for each protocol. These measures could be exploited also to simultaneously guide the construction of new protocols in robotic mirroring coaching and serve as possible metrics for the continuous evaluation of subject performance.

II. Methods

A. Mirroring system

The mirroring coaching platform consists of the NAO robot (Aldebaran Robotics), and the Kinect as a motion tracking system [23]. NAO is a small and portable humanoid robot, widely used in robotic coaching systems [12], [13], [18]. In the current application, the controlled joints corresponded to the three degrees of freedom of the robot upper limb: Elbow Roll, Shoulder Roll and Shoulder Pitch. The Kinect tracked two people simultaneously at a rate of 30 Hz. The 3D positions of each joint in the upper limb were monitored for each person. A moving median filter (5 samples) was applied for impulsive noise reduction. The keypoints positions extracted in the Kinect reference frame were mapped to the NAO reference frame. The angles between one limb segment and the reference frame axes, or between two limb segments were calculated by customized algorithms, and used as robot control angles (details in [23], [24]).

B. Acquisition Protocols and Participants

The robotic mirroring system was used to implement different exercises in a triadic interaction, between an adult, a child and the robot (Figure 1(a)). The three actors were positioned in a triangle, with the Kinect camera positioned above the NAO

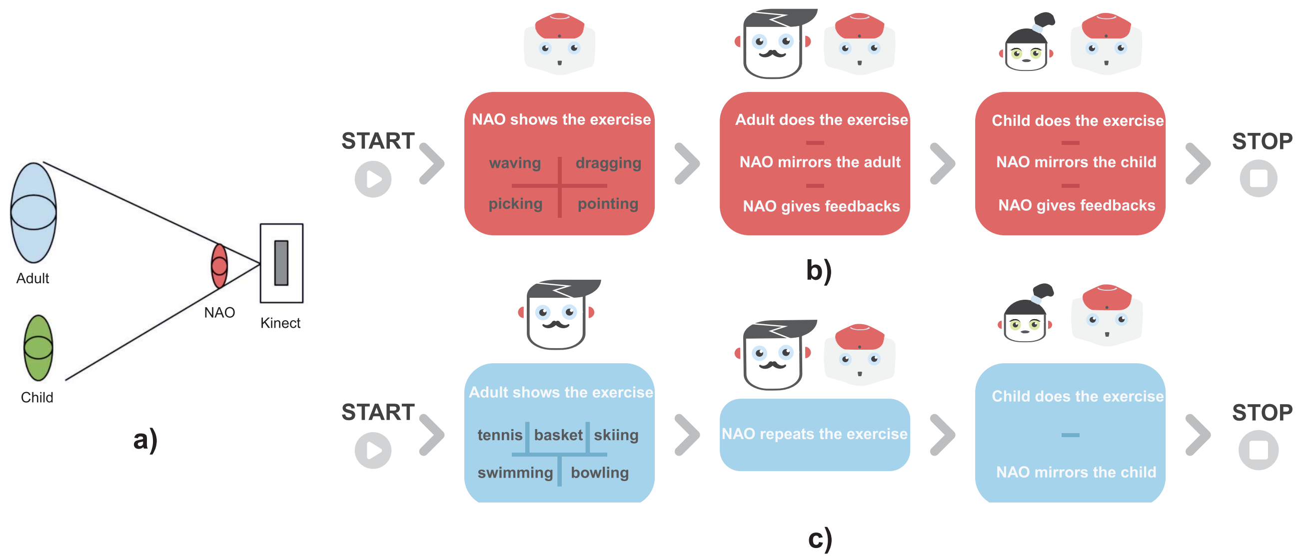

Fig. 1. Diagram of the triadic interaction (adult-child-robot), illustrating the geometry between the participants and the perception system (Kinect) used during a protocol session (a) and block diagrams of the two game protocols. (b) *Robot-Master* protocol: first the NAO shows the exercise, then the adult repeats it while the robot is mirroring, and finally the child does the exercise, mirrored by the robot. (c) *Adult-Master* protocol: the adult leads the game, showing the exercise, which is then repeated by the robot and finally by the child while the robot is mirroring. Adapted from [23].

robot and opposite to the two people, so that the Kinect could track both subjects without occlusions. It was assumed that the left-most person with respect to the Kinect was always the adult. The protocols were designed in collaboration with the rehabilitation center IRCCS Fondazione Don Carlo Gnocchi in Milan (Italy), with the aim to build an interactive game with meaningful gestures, training both social and motor skills for a prospective application in ASD children therapy.

Two games were conceived, the *Robot-Master* and the *Adult-Master* protocols, as shown in Figure 1. The games were contextualized in a semantic framework in order to address also cognitive skill training: the NAO played the role of an individual coming to Earth from outer space.

These two game modalities were designed in view of future clinical applications to allow to analyze any possible difference in the child's behaviour and engagement, depending on whether the action is lead by the robot or the adult. Moreover, the *Robot-Master* protocol involves simple movements, which allows to directly analyse the participants' joint trajectories to evaluate performance. Instead, in the *Adult-Master* protocol, the complexity of the movements required the definition of a meta-parameters set to capture the overall movement characteristics. This analysis was carried out in the space of features describing the joint trajectories.

Within the same session, each protocol was executed once by a child and an adult (referred to as pair/couple), with the *Adult-Master* protocol following the *Robot-Master* protocol. The number of acquisitions was 15, resulting in a dataset of 30 individuals in total, including 14 healthy children (between 5 and 10 years old, 7 boys and 7 girls) and 14 adults close to each child (usually one of the parents), naive to the platform. They were recruited from schools and selected based on their age. The remaining two individuals were a 5-year-old girl, diagnosed with ASD according to the Diagnostic and Statistical Manual of Mental Disorders V [25] and to the Autism Diagnostic Observation Schedule [26] (comparison score = 9), and her therapist experienced with technology-mediated treatments. This last acquisition was performed in a clinical setting in Fondazione Don Carlo Gnocchi in Milan. Participants were asked to follow the robot instructions and mirror its movements. The study was approved by the Ethical Committee of Politecnico di Milano (reference number: 18/2019; date: 19/09/2019); the adults and child's legal guardians signed an informed written consent prior to the acquisitions.

C. Robot-Master Protocol

In the *Robot-Master* protocol, NAO leads the game (Figure 1(b)): it shows a movement, then gives a "go" signal (robot pointing towards the adult) asking the adult to repeat it. After the "go" signal, the robot has some time to start processing the recorded data, and then it mirrors the adult. After, the robot gives the "go" signal to the child, who executes the gesture while NAO is mirroring. If the movement is performed correctly, NAO gives a positive vocal feedback ("Bravo" or "Grande", i.e. "Great") and its LEDs turn green. Otherwise, if the adult or the child do not finish the exercise within 20 seconds, the LEDs on NAO eyes become red. This time limit was determined from clinical experience.

1) Movements: With the aim to train gestures in a semantic grounding, the movements chosen for this protocol were related to the theme of "Space". Specifically, starting from the standing position with the arms along the body, the subjects had to do: (i) "waving" (greeting movement done with the left arm), (ii) "dragging" (reaching the top left position and back with two arms with the objective of moving clouds from the sky), (iii) "picking" (reaching the top position and back with each arm sequentially with the objective of "taking out stars") and finally (iv) "pointing" (pointing to a top position with one arm; participants were supposed to point to a planet).

2) Feedback System: For each movement, positive feedback was given whenever the current pose joint angles p_c

TABLE I
TARGET POSES AND RANGES OF VARIABILITY (α) USED IN THE FEEDBACK SYSTEM OF THE *Robot-Master* PROTOCOL

Movements	Target Pose	Final Target Angles (rad)				α (rad)
		Elbow Roll				
Waving	Inward Rotation	1.19				0.3
	Outward Rotation	0.29				
		Shoulder Pitch		Shoulder Roll		
		Right	Left	Right	Left	
Dragging	Hands up	-0.54	-0.63	-0.54	0.10	0.3
	Hands down	1.2	1.3	-0.2	0.42	
		Shoulder Pitch		Shoulder Roll		
Picking	Right Hand up	-0.87		0.14		0.3
	Right Hand down	1.37		0.14		
	Left Hand up	-0.91		-0.37		
	Left Hand Down	1.34		-0.16		
		Shoulder Pitch				
Pointing	Hand up	-0.96				0.2

reached the target angular position p_t in each subphase of the movement, for the most significant degrees of freedom, with a certain range of variability, α (Equation 1):

$$||p_t - p_c||_\infty \leqslant \alpha \tag{1}$$

The target angle values and degrees of freedom, specific for each movement and for each movement phase, were chosen based on a database of signals from the same movements, previously recorded in a group of 28 healthy adults. Each subject executed the 4 movements of the *Robot-Master* protocol with three repetitions. The target angular pose was obtained by averaging all the individual target poses. The range of variability (α), with respect to the target, was defined for each movement considering that all the individual target poses should have positive feedback, since the subjects did the movement correctly (Table I). Both target pose and range of variability have an angular representation, being invariant to the body ratio. Thus the values determined for adults could be applied also to children.

For the completion of the exercise and reception of the final vocal feedback, the subject had to accomplish 6 subphases in the case of the waving (3 inward rotations and 3 outward rotations), 2 subphases in the case of the dragging and 4 in the case of the picking, as described in Table I. The pointing exercise contained one single subphase.

3) Data Analysis:

a) Signal processing: In order to extract parameters for evaluating the designed protocol, the Kinect's keypoints and the corresponding robot joint angles were analysed. Processing of the movements started after the "go" signal with a short delay (mean±SD: 6.20 ± 0.03 s).

b) Outcome measures: In the *Robot-Master* protocol, the extracted parameters evaluated both the movement temporal properties and the difficulty of execution, thus quantifying the exercises' complexity.

Specifically, the movement *latency* was computed from the 3D coordinates of the hand and wrist keypoints, as the time-interval between the robot "go" signal and the time instant when the person starts the exercise. Then, the control angle signals were used to extract the other parameters. The exercise's *duration* was computed as a global indicator of the complexity of the chosen movements. Only those exercises completed in less than 20 seconds were considered for the calculation of the overall duration of each exercise. The number of uncompleted exercises was accounted as the *number of failures*. To evaluate movement difficulty, we also extracted the *number of attempts* a person had to do before reaching the first target position. They were the number of saddle points in both limb angular signals, before the first feedback. For each exercise, different most significant degrees of freedom were chosen, as described in subsection II-C.2. A parallel measure of the difficulty of the exercises was the *first positive feedback time (FPFT)*, i.e. the time interval until receiving the first feedback.

For all parameters, values higher than three scaled median absolute deviations from the overall median were identified as outliers and removed from the analysis.

D. Adult-Master Protocol

In the second protocol, the adult is the master: first he/she demonstrates the movement for a certain time interval (7s), asking NAO to replicate the movement. Finally, it becomes the child's turn to perform the movement, while NAO is mirroring the child (Figure 1(c)). In this protocol, no feedback is given by the robot, because the adult, e.g. the therapist, is supposed to provide it, leading the interaction.

1) Movements: This protocol was set within the theme of "Sports" in order to teach NAO the sports from the Earth. The chosen sports were: basket, bowling, swimming, tennis and skiing, each one associated to a representative movement. Within a 7-second time window, each adult was allowed to repeat the movement multiple times to show to the robot. The children could also repeat the movement several times but without any time limitation. For the basket, the movement was throwing up a ball with both hands above the head; for bowling, throwing a ball forward with one arm; for swimming, breaststroke movements; for tennis, hitting a ball laterally; for skiing, sliding the arms. With this protocol, we wanted to evaluate which sports are more standard across subjects and could be used for robotic mirroring coaching. In this case, movements are more complex than the "Space" ones, being multi-joint global movements. Therefore, novel meta-parameters based on multiple signals features were designed.

2) Data Analysis:

a) Signal processing: As explained in the previous paragraph, multiple movement repetitions were collected for the same exercise. Each recording included a sequence of joint angle measurements, containing the exercise execution, which we call movement repetition, and other non-functional movements.

To extract the movement repetitions, angle signals of the most significant degrees of freedom were selected, i.e. the right shoulder roll for tennis and swimming, the right shoulder pitch for bowling, ski and basket. Then, the angle trajectories were filtered with a median or a moving-average filter (for tennis task), in order to reduce noise.

Each movement could be described with the sequence of values of the most significant degree of freedom, e.g. swimming motion could be seen as the evolution of the right-shoulder roll angle, from an initial value, towards a peak and back. This concept is used to identify movement repetitions within the filtered signal, by finding the frames where the signal goes above a calculated baseline, has a peak in between and returns under the baseline value. The baseline value was calculated as a percentage of the range between the minimum and maximum value of the angle movement. Specifically, for each exercise this was between 20 and 55 % of the range.

b) Features: Since these gestures were more complex than those in the *Robot-Master* protocol, we chose to extract a set of main features from the segmented signals in order to characterize the movement in a more holistic manner. Then, meta-parameters to evaluate the *Adult-Master* protocol were designed and computed from the distribution of the observed/measured features.

Features were computed from both single subject repetitions and paired adult-child repetitions. For each subject repetition, we computed the half-width signal duration and the amplitude of each repetition. The other (paired) features were obtained from a combination of the child-adult pair signals to obtain the maximum correlation and lag between the child's repetition and the adult's repetition. To compute the correlation between the two (child/adult) repetitions, both signals were first resampled to the shortest repetition's size. Then, the correlation was calculated and the maximum value and respective lag were extracted.

We used the half-width duration and the amplitude as features to characterize the observed movement for each single subject, as illustrated in Figure 2(a). Each data point represents a repetition of a subject, characterized by the corresponding half-width duration and amplitude of that repetition. Therefore, for child C_j and adult A_k, the set of observations can be represented as follows:

$$C_j = \{R_1^{C_j}, \ldots, R_{N_j}^{C_j}\} \text{ and } A_k = \{R_1^{A_k}, \ldots, R_{N_k}^{A_k}\} \quad (2)$$

where N_j is the number of repetitions R of the movement for child C_j and N_k is the number of repetitions of the movement for adult A_k. Both j and k, go from 1 to 15, since there are fifteen children and fifteen adults.

Other sets of observations were constituted by the combinations of each repetition of one child and each repetition of one adult, using the correlation and the lag between the repetitions as features (Equation 3 and Figure 2(b)).

$$P_{kj} = \{R_1^{C_j} R_1^{A_k}, R_1^{C_j} R_2^{A_k}, \ldots, R_{N_j}^{C_j} R_{N_k}^{A_k}\} \quad (3)$$

c) Outcome measures: After forming the different sets, the exercises' repeatability within categories was analysed, using the single subject features. We considered all the data points for healthy children and adults to evaluate movement variability in the two populations. Thus, we joined all the repetitions of the children and compared with all the repetitions of the adults. When forming the two sets, the amplitude and the half-width duration were normalized by the maximum value for each sport due to different range of motions of corresponding movements. Then, the variability of each set (Δ) was given by the mean of distances of the several elements R_i, to the centroid of the data set ($centr_T$), as calculated in Equation 4. We named this metric *intraset distance*.

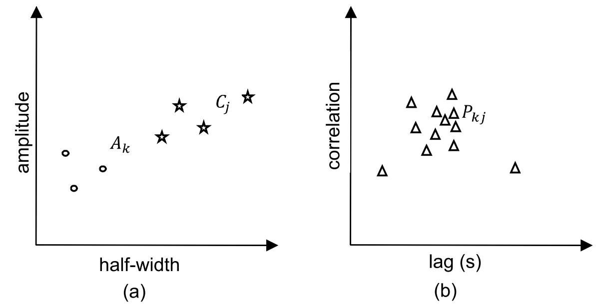

Fig. 2. Example of set of observations using (a) single subject analysis and (b) paired analysis. In (a), circles and stars indicate the data points for one example adult and child, respectively; in (b) triangles indicate data points for one example pair.

$$\Delta = \frac{\sum_i dist(centr_T, R_i)}{|T|} \quad (4)$$

where T represents a generic data set (which can be the set of the adults A, or children C), $|T|$ is the number of elements of T and the values of i depend on the number of elements of the set. We chose $dist$ to be the Euclidean distance. Lower is this distance, less variable is the set, more repeatable is the exercise among the categories that have tried it.

Secondly, the repeatability among pairs in each exercise was quantified. To this aim, just the paired repetitions of children and adults in the same session were used to form the different sets ($k = j$ in Equation 3). In total, for each exercise there were fifteen sets, which corresponded to the fifteen pairs. Again, the *intraset distance* was calculated, but in this case, T represented a pair of adult-child and R_i was represented by the correlation and lag between the child and adult. The mean of the *intraset distances* of the healthy pairs for each sport was computed to understand which exercise was less variable among the different pairs.

Finally, we evaluated how the mirroring between the child and his/her respective adult changed with the exercises. The analysis was divided by each child. There were fifteen sets, for the fifteen adults combined with one child. The sets were formed by the correlations and lags of each repetition of the child (j in Equation 3 is fixed) and the several repetitions of each adult ($k = 1, \ldots, 15$). Then, the *intraset distance* between the child and respective adult was calculated and compared with the mean of the *intraset distances* between the child and all the other adults. When the lowest value of *intraset distance* was between the child and respective adult, this meant that they had a lower variability of correlation and lag between them compared to the other combinations of the same child and the other adults. They were mirroring each

TABLE II
PARAMETERS EXTRACTED FOR THE TWO PROTOCOLS AND ASSOCIATED RESEARCH QUESTIONS

Protocol	Parameters (features)	Research Question
Robot-Master	Latency Duration FPFT # attempts # failures	RQ1: Movement complexity RQ2: Role of the feedback
Adult-Master	Single intraset distance (half-width duration, amplitude)	RQ3: Movement standardization
	Paired intraset distance (lag, correlation peak)	RQ1: Movement 'mirrorability'

other, since the child movement was more similar to his/her respective adult than to the other adults.

In general, the design of both protocols and analysis tried to answer the research questions proposed in section I (Table II). The choice of the best protocols for robotic mirroring coaching was addressed in both cases, while to understand the better robotic feedback just the *Robot-Master* protocol was required. The standardization of the movements was explored in the *Adult-Master* protocol, because this involved repeated movements initiated by the different subjects, thus the different execution and mirroring could be evaluated.

III. RESULTS

The two protocols were correctly executed by the participants. All the child-adult pairs enjoyed the experience and demonstrated availability to collaborate in future acquisitions. The majority of them wanted to explore the mirroring system freely after the acquisition to test its capabilities. While during the acquisitions with healthy children the exercises order was always the same within *Robot-Master* and *Adult-Master* protocol, in the clinical acquisition, the therapist chose the best order based on the child needs. In the end, the child completed the two protocols and was engaged during the whole process.

A. Robot-Master Protocol

1) System Demonstration: In the *Robot-Master* protocol, we verified the correctness of the feedback system throughout the acquisitions. Figure 3 shows example signals for three cases (waving, dragging and pointing), with exercise-specific criteria required to give a positive feedback. For the dragging (plot (b)), the movement was considered correct when all the four degrees of freedom reached their target angle region. In the pointing exercise (plot (c)), the positive feedback was only provided three seconds after the subject maintained the target position. In the case of the waving (plot (a)), the child was already in the target position at the beginning of the analysis, so a positive feedback was recorded at the initial time. The negative feedback (red line) appeared because the exercise was not concluded after 20 seconds.

2) Outcome Measures: In order to explore the complexity/difficulty of the different exercises, the values of five parameters described in Section II are here reported.

a) Latency: The latency values for both adults and children can be found in Table III. The latencies of the children resulted bigger than the adults', except for the waving. The biggest

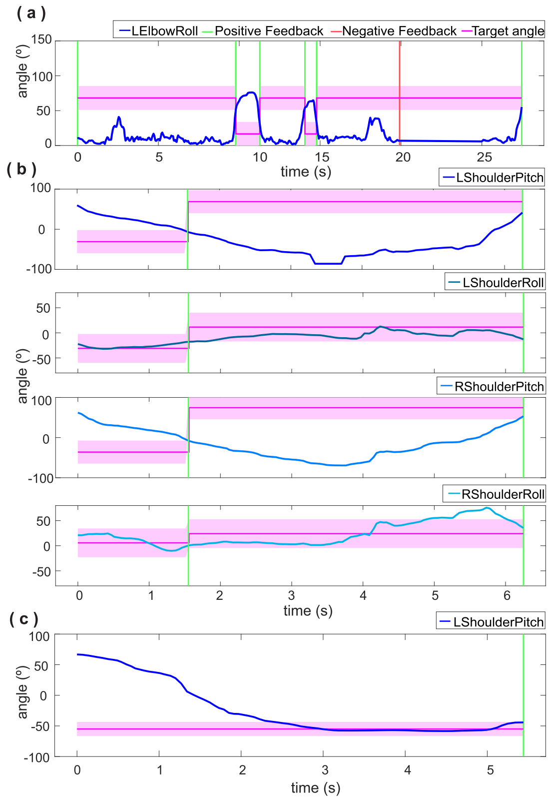

Fig. 3. Signals of the (a) waving, (b) dragging and (c) pointing and respective feedback signals for one child subject during the *Robot-Master* protocol. The zero is associated to the beginning of the analysis by the robot. The green lines correspond to the several positive feedback messages given by the robot and the red line corresponds to the negative feedback. The pink line represents the target pose and the pink shade the range of variability α of each movement.

TABLE III
LATENCY TIMES OF THE THERAPIST AND ASD CHILD AND MEAN LATENCIES OF THE HEALTHY ADULTS AND CHILDREN FOR THE EXERCISES IN THE *ROBOT-MASTER* PROTOCOL. THE ZERO VALUES REPRESENT SITUATIONS IN WHICH THE SUBJECT WAS ALREADY IN THE STARTING POSITION AT THE ONSET OF THE ANALYSIS

	Mean±SD Adults	Therapist	Mean±SD Children	ASD child
Waving (s)	4.45±5.10	0.00	3.76±3.77	31.25
Dragging (s)	0.79±0.91	0.00	1.18±1.67	3.57
Picking (s)	1.06±1.43	0.26	1.69±1.20	11.20
Pointing (s)	1.21±1.39	0.00	4.26±4.22	5.08

latency corresponded to the adults' waving, the first group doing the first exercise of the system. The high standard deviation of the latency, present in both adults and children, reflects the variability on how each person understands the execution of each exercise. In general, the ASD child had bigger latencies compared to the mean of the healthy children, while the therapist, by contrast, had very small values, a sign of her experience on this type of platforms. In some cases, the subject started the exercises before the ending of the short delay intrinsic to the platform. This resulted in a latency value of zero (Tables A1, A2), which was not considered in the average computation.

b) First positive feedback time (FPFT) and number of attempts: From the *number of attempts*, the waving was clearly a difficult exercise for both children and adults (Table IV). In some

TABLE IV

FIRST POSITIVE FEEDBACK TIMES (FPFT) AND NUMBER OF ATTEMPTS UNTIL THE FIRST FEEDBACK OF THE THERAPIST, ASD CHILD, HEALTHY ADULTS AND CHILDREN FOR THE EXERCISES IN THE *Robot-Master* PROTOCOL. THE EMPTY CELLS ARE OUTLIERS

	Mean±SD Adults	Therapist	Mean±SD Children	ASD child
Waving				
FPFT (s)	5.37±4.64	0.30	5.32±3.91	6.31
# of attempts	2.93±3.43	1	2.79±1.89	5
Dragging				
FPFT (s)	2.21±1.15	0.00	2.22±1.89	3.98
# of attempts	1.79±1.25	1	2.71±4.07	2
Picking				
FPFT (s)	0.89±1.29	1.05	2.33±1.37	11.54
# of attempts	1.50±1.02	1	1.36±0.74	6
Pointing				
FPFT (s)	6.64±3.00	3.02	10.86±6.96	8.12
# of attempts	1.36±0.63	1	2.79±1.85	3

cases, the subjects started with the ipsilateral arm, when they should have used the controlateral arm, since they were doing a mirroring exercise. Moreover, it was the first exercise of the protocol to be executed, so the higher number of attempts is expectable. Consequently, the first positive feedback times were also bigger. For the children, the pointing was the most difficult exercise to achieve a positive feedback, since they had to keep the arm at the pointing position for three seconds. For the ASD child, the same difficulty on the pointing was verified, although the most difficult exercise was the dragging.

Moreover, the therapist had some very low values of *FPFT* which is justified by the fact that she initiated the movement soon after the robot's "go" signal. Therefore, she achieved the target angle at the time of the start of the analysis, as the subject in Figure 3(a).

c) Duration and number of failures: The majority of children and adults completed all the exercises within the settled time. The exercise with the bigger number of failures was the waving of children (Table V). As in the *latency* parameter, the biggest duration was verified for the waving in the case of the adult (Table V). This is justified by the several target angles that should be sequentially completed (a total of six subphases). Therefore this exercise out-stands by its increased difficulty. The picking exercise was an intrinsically slow exercise since it required each arm to be completely extended, which the robot took some time to mirror. The pointing exercise was not included in this analysis since its duration was equal to the first positive feedback time. Regarding the ASD child, she did not finish the picking and waving exercises within the settled time. However, further acquisitions are expected to train this skill, achieving movement durations within healthy ranges.

In summary, the most difficult exercise was the waving for the adults and pointing for the children. Other differences between children and adults were not so visible, and were just given by the fact that each adult executed the exercise before each child.

B. Adult Master Protocol

1) System Demonstration: After experiencing the *Robot-Master* protocol, the pairs adult-child were already familiar

TABLE V

DURATION OF SUCCESSFUL TRIALS AND NUMBER OF FAILURES FOR THE THERAPIST, ASD CHILD, HEALTHY ADULTS AND CHILDREN IN THE EXERCISES OF THE *Robot-Master* PROTOCOL. THE EMPTY CELLS ARE OUTLIERS

	Total/ Mean±SD Adults	Therapist	Total/ Mean±SD Children	ASD child
Waving				
# failures	7	0	6	1
Duration(s)	12.72±5.94	2.75	10.27±3.84	-
Dragging				
# failures	0	0	3	1
Duration(s)	6.79±4.25	3.57	6.50±3.24	8.71
Picking				
# failures	5	0	3	1
Duration(s)	9.81±3.90	6.77	12.08±3.89	-

TABLE VI

Intraset Distance IN THE CHILDREN SET, ADULTS SET, THERAPIST AND ASD CHILD SETS FOR EACH SPORT OF THE *Adult-Master* PROTOCOL

	Tennis	Swimming	Bowling	Ski	Basket
Adult set	0.17	0.16	0.27	0.14	0.26
Child set	0.30	0.19	0.27	0.21	0.28
Therapist	0.16	0.18	0.00	0.09	0.05
ASD child	0.26	0.24	0.15	0.08	0.10

with the mirroring system and the robot's "go" signal. Since the execution of the adult had a specific time interval and the execution of the child did not, the repetitions done by the adult were in general less than the repetitions done by the child. Moreover, in the case of the child, the robot was mirroring him/her, while in the case of the adult, it was not, thus the movements' characteristics of each population tended to be different.

2) Outcome Measures: In this section the results related to standardization of the different sport movements (within children and adults and between the correspondent pairs) are presented. The exercises are also analysed in terms of their possibility to be easily mirrored by the participants.

a) Repeatability among categories: Figure 4 reports the comparison between the sets of the children and of the adults. The set of the therapist and of the ASD child are presented separately. In all sports, the centroid of the children was at a higher amplitude and half-width duration than the centroid of the adult, except for the swimming, in which both child and adult had a very similar value.

In order to evaluate the variability of each set, *intraset distances* were calculated (Table VI). The mean of these distances represents the size of the set. In general, the two sets had a comparable size although the children set had always a bigger variability than the adults set. The same pattern of variability was verified for the ASD child and therapist, except for the ski movement. Overall, the ski and the swimming were the least variable exercises, in which the sum of the *intraset distances* from both sets was the smallest, which is also visible in Figure 4.

b) Repeatability among pairs: In order to evaluate the repeatability of the paired movement, just the combinations between correspondent adult and child repetitions were considered (Figure 5). The pair ASD child-therapist was in the same region of the other pairs of healthy children. The most

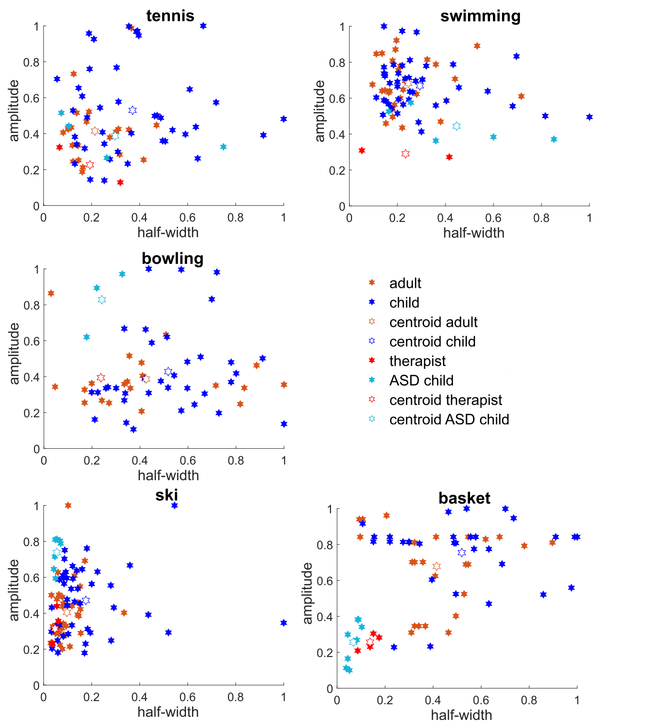

Fig. 4. Child (blue) and adult (orange) observation sets for each sport in the *Adult-Master* protocol. Empty markers indicate the centroids of each set.

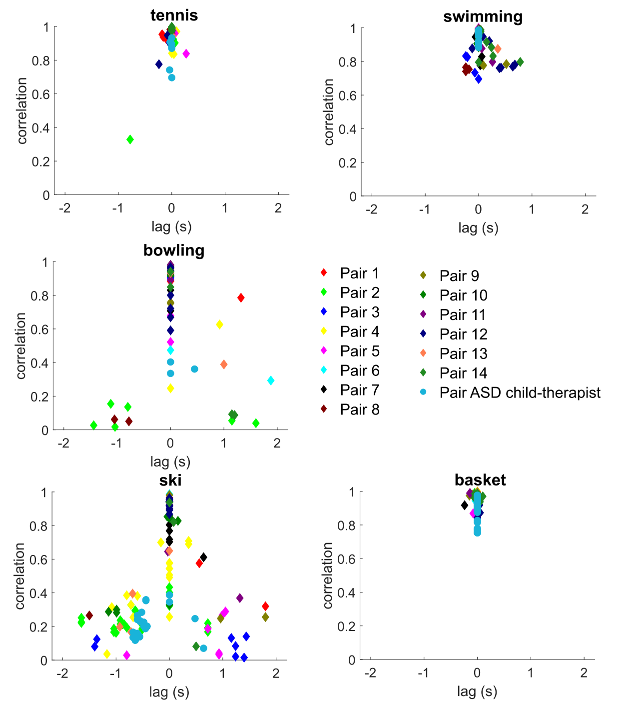

Fig. 5. Observation sets for the different pairs (different colours) in each sport of the *Adult-Master* protocol.

TABLE VII

MEAN OF THE *INTRASET DISTANCES* FOR HEALTHY SUBJECTS AND *INTRASET DISTANCE* FOR ASD CHILD-THERAPIST PAIR FOR EACH TYPE OF EXERCISE IN THE *ADULT-MASTER* PROTOCOL

	Tennis	Swimming	Bowling	Ski	Basket
Mean of intraset distance healthy	0.07	0.10	0.40	0.53	0.04
Pair therapist ASD child	0.07	0.03	0.20	0.23	0.06

compact sets were the ones of tennis, swimming and basket, while bowling and ski presented higher variability. The same conclusion is evident from Table VII. The highest variability of the bowling and ski exercises was caused by a variability of the lags between paired child-adult signals, which is not so present in the other exercises.

c) Mirroring in the different exercises: After, the data of each child was combined with the data of the several adults in order to understand if the child was mirroring or not the correspondent adult, depending on the exercise. The results of calculating the *intraset distance* between the child and the respective adult and the mean of the *intraset distances* between the child and all the other adults are reported in Table VIII. As explained in section II, when the lower value of *intraset distance* is between the child and respective adult, this means that they have a lower variability of correlation and lag between them compared to the other combinations of the same child and the other adults. This condition represents a closer execution between the mirroring pairs with respect to all other possible pairs. Instead, when the lower value is with the other adults, this means it is impossible to say that child is mirroring the paired adult.

The exercises easiest to mirror were the basket and the tennis, with significantly lower distances between the respective adult in comparison with the other adults, based on the Wilcoxon Signed Test ($p < 0.05$). On the bowling and the swimming eight children had a closer execution to their respective adult compared with other adults. The ski was the exercise most difficult to mirror, with significantly higher distances between the respective adult in comparison with the other adults ($p < 0.05$).

In summary, among pairs, tennis and basket were the exercises with a higher repeatability and consequently a higher standardization. Furthermore, these sports were the ones with a lower degree of mirroring effort between child and adult (more 'mirrorable').

IV. DISCUSSION

In this paper, we designed a platform and two game protocols for robotic mirroring coaching of motor, social and cognitive skills, based on a triadic interaction in a semantic framework. Design choices were driven by clinical experience and needs, considering minimal intrusiveness of the hardware and high engagement of the end-user. The exercises were chosen among meaningful gestures within a story (*Robot-Master*), or sport gestures (*Adult-Master*).

The system and protocols were tested on healthy subjects, showing ease of use, transparency and flexibility. In addition, several parameters were defined to evaluate the different exercises and the subjects performance, providing metrics for the characterization of movements and evaluation of execution.

A. Robot-Master Protocol

In this protocol, time-related parameters were used to estimate the complexity of different exercises. The waving and

TABLE VIII
FOR EACH SPORT IN THE *ADULT-MASTER* PROTOCOL, *INTRASET DISTANCES* OF THE PAIR CHILD-RESPECTIVE ADULT (LEFT COLUMN) COMPARED TO THE MEAN OF *INTRASET DISTANCES* FOR THE PAIRS CHILD-OTHER ADULTS (RIGHT COLUMN). THE COLOURED CELLS MARK THE LOWER DISTANCES, BEING GREEN WHEN VERIFIED BY THE COUPLE CHILD-RESPECTIVE ADULT (MIRRORING VERIFIED), AND RED, OTHERWISE

	Tennis		Swimming		Bowling		Ski		Basket	
	Respective adult	Other adults	Respective adult	Other adults	Respective adult	Other adults	Respective adult	Other adults	Respective adult	Other adults
Child1	0.03	0.04	0.004	0.06	0.66	0.21	0.67	0.34	0.03	0.15
Child2	0.38	0.39	0.004	0.02	1.10	1.05	0.52	0.45	0.02	0.02
Child3	0.04	0.05	0.14	0.18	0.02	0.08	1.41	0.77	0.00	0.00
Child4	0.06	0.06	0.01	0.06	0.52	0.58	0.56	0.54	0.01	0.05
Child5	0.12	0.04	0.00	0.07	0.08	0.07	0.50	0.12	0.07	0.09
Child6	0.00	0.00	0.01	0.04	0.94	0.14	0.07	0.14	0.03	0.03
Child7	0.02	0.05	0.09	0.03	0.06	0.60	0.20	0.26	0.12	0.04
Child8	0.02	0.01	0.16	0.25	0.70	0.48	0.80	0.43	0.02	0.05
Child9	0.02	0.03	0.22	0.16	0.00	0.34	0.79	0.09	0.06	0.04
Child10	0.01	0.01	0.02	0.07	0.03	0.05	0.57	0.46	0.00	0.02
Child11	0.00	0.01	0.11	0.06	0.04	0.15	0.54	0.27	0.07	0.08
Child12	0.10	0.10	0.27	0.21	0.13	0.13	0.02	0.16	0.04	0.06
Child13	0.00	0.01	0.16	0.11	0.57	0.48	0.34	0.24	0.02	0.03
Child14	0.01	0.01	0.17	0.07	0.71	0.86	0.41	0.67	0.06	0.09
ASD child	0.07	0.18	0.03	0.14	0.20	0.52	0.23	0.24	0.06	0.19

the pointing exercises were evaluated as the most complex for the current mirroring system. These parameter values were also influenced by the information provided before the session beginning. On one hand, some subjects initiated an anatomical imitation exercise instead of a mirroring imitation exercise. On the other hand, the delay between the robot "go" signal and the analysis start was not considered by the subjects in the first repetitions; this had minor impact on the obtained values of the first positive feedback time. Nevertheless, future versions of the platform could use a more instantaneous "go" signal, e.g. a sound.

Since the conclusion of the exercises depended on reaching several consecutive target poses, if the first one failed due to erroneous reaching, this would lead to a postponed feedback. Answering our second research question, a feedback system able to recognize the gestures as a whole and not just sub-phases of it, would be more efficient. As a possible future direction we are considering machine learning algorithms that could be more robust for this purpose [27]. The feedback computation would also become independent from the selected most significant degree of freedom, allowing to consider the trajectory of all involved joints at the same time.

B. Adult-Master Protocol

In this protocol, measures of the variability and ease of mirroring for the exercises were extracted. Based on point-wise features of the movements, swimming and ski were the most standardized movements, since they had the lowest sum of the *intraset distances* of children and adults, in the single subject analysis. From this analysis, the difference between the movements of children and adults became also clear: children tended to do bigger and longer gestures in most exercises.

When using the correlation and lag, to evaluate each mirroring pair, the tennis, swimming and basket turn out to be the most repeatable movements, while the ski was the most variable. Overall, the sport gestures had different levels of intra-pair stereotipation: some were very similar in between users, while others were very variable. The selection of the type of gesture should be dependent on the therapy goal.

The analysis on pairs was more appropriate than the single subject analysis to evaluate child-adult mirroring, since it considered the whole signal and not just point-wise features. The mirroring effect was verified in all exercises for the majority of children, except in the ski exercise. This was expected, since this movement included the hyperextension of the arm, occluding the wrist keypoint from the Kinect view. Consequently, the robot did not mirror correctly the movement demonstrated by the adults. Indeed, the correctness of the children to imitate the gesture was also dependent on the capability of the robot. However, after noticing this phenomenon, children adjusted the exercises done, so that the robot was able to do it, by slowing down the movements or by putting the arms always in front of the trunk. Basket and tennis were the easiest to mirror. Thus, answering our first research question, protocols for robot mirroring coaching should involve global movements with high range of motion.

In relation to the clinical acquisition, the ASD child was engaged during the two protocols. In the *Robot-Master* protocol, she completed two out of the four exercises within the settled time of 20 seconds. This parameter could be changed in the future to better cope with the durations of the target population. In the *Adult-Master* protocol the similarity between therapist and ASD child movements became evident through the computed parameters. Although the comparison with healthy children would not be robust due to the very different number of subjects, these measurements could be used in the future to monitor the progresses of ASD children during therapy.

V. CONCLUSIONS AND FUTURE WORK

The pilot study presented in this work showed that the two designed protocols of robotic mirroring coaching can be suitable for turn-taking games between child and adult. Despite some minor criticalities in the feedback system, all the subjects were able to conclude the exercises and, specially the children, wanted to experiment the protocol freely after the session. The analysis demonstrated the importance of balancing movements complexity with technical characteristics of the robot.

This mirroring scenario will also be further tested in a therapeutic framework for children with autism, in which robots have been recently introduced with proven benefits [28]. This is the only framework where the clinical efficacy of the current strategy can be fully assessed. This will require the evaluation of ASD children heterogeneity and the definition of inclusion criteria and will be the focus of a dedicated clinical study.

Nevertheless, the NAO robot chosen for this study is being widely used in social therapies for ASD, with promising results [5]. As a humanoid robot, it looks like a human without being one, and it can provide audio and visual stimuli, all features favourable for interaction with ASD children, who tend to prefer simplified stimuli to avoid focusing on details [28]. Moreover, the Kinect camera is an optimal tracking system for this application, since transparent setups are recommended for use with patients, especially children.

The results obtained here provide insights for the clinicians to design flexible protocols (subset of exercises or different games) that can be adapted and tailored to the target population, addressing specific levels of disability. Future versions could implement an adaptive protocol, taking into account the variability of ASD. The difficulty and order of exercises could become incremental, adapted to different ASD levels, and eventually personalized by the therapist based on specific child needs.

Finally, in a therapeutic scenario, the platform described here could provide measurements and quantitative parameters to evaluate therapy, including the ones designed and calculated in this study. This kind of quantitative analysis is missing in research on robotic therapies for autism [28], and would be fundamental to evaluate the progress of the therapy.

References

[1] I. Papadopoulos, C. Koulouglioti, R. Lazzarino, and S. Ali, "Enablers and barriers to the implementation of socially assistive humanoid robots in health and social care: A systematic review," *BMJ Open*, vol. 10, no. 1, Jan. 2020, Art. no. e033096.

[2] G. Hofree, P. Ruvolo, M. S. Bartlett, and P. Winkielman, "Bridging the mechanical and the human mind: Spontaneous mimicry of a physically present android," *PLoS ONE*, vol. 9, no. 7, Jul. 2014, Art. no. e99934.

[3] J. Fasola and M. Mataric, "A socially assistive robot exercise coach for the elderly," *J. Hum.-Robot Interact.*, vol. 2, no. 2, pp. 3–32, Jun. 2013.

[4] B. Borovac, M. Gnjatović, S. Savić, M. Raković, and M. Nikolić, "Human-like robot marko in the rehabilitation of children with cerebral palsy," in *New Trends in Medical and Service Robots*. Cham, Switzerland: Springer, 2016, pp. 191–203.

[5] P. Pennisi *et al.*, "Autism and social robotics: A systematic review," *Autism Res.*, vol. 9, no. 2, pp. 165–183, Feb. 2016.

[6] E. Biffi *et al.*, "Gait pattern and motor performance during discrete gait perturbation in children with autism spectrum disorders," *Frontiers Psychol.*, vol. 9, p. 2530, Dec. 2018.

[7] L. R. Watson, E. R. Crais, G. T. Baranek, J. R. Dykstra, and K. P. Wilson, "Communicative gesture use in infants with and without autism: A retrospective home video study," *Amer. J. Speech-Lang. Pathol.*, vol. 22, no. 1, pp. 25–39, Feb. 2013.

[8] M. Jouaiti and P. Hénaff, "Robot-based motor rehabilitation in autism: A systematic review," *Int. J. Social Robot.*, vol. 11, no. 5, pp. 753–764, Dec. 2019.

[9] J. Nadel, A. Revel, P. Andry, and P. Gaussier, "Toward communication: First imitations in infants, low-functioning children with autism and robots," *Interact. Stud.*, vol. 5, no. 1, pp. 45–74, May 2004.

[10] I. B. Wijayasinghe, I. Ranatunga, N. Balakrishnan, N. Bugnariu, and D. O. Popa, "Human–robot gesture analysis for objective assessment of autism spectrum disorder," *Int. J. Social Robot.*, vol. 8, no. 5, pp. 695–707, Nov. 2016.

[11] Z. Zheng, E. M. Young, A. R. Swanson, A. S. Weitlauf, Z. E. Warren, and N. Sarkar, "Robot-mediated imitation skill training for children with autism," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 6, pp. 682–691, Jun. 2016.

[12] B. Görer, A. A. Salah, and H. L. Akin, "A robotic fitness coach for the elderly," in *Proc. Int. Joint Conf. Ambient Intell.*, 2013, pp. 124–139.

[13] A. Di Nuovo, D. Conti, G. Trubia, S. Buono, and S. Di Nuovo, "Deep learning systems for estimating visual attention in robot-assisted therapy of children with autism and intellectual disability," *Robotics*, vol. 7, no. 2, p. 25, Jun. 2018.

[14] J.-J. Cabibihan, H. Javed, M. Ang, and S. M. Aljunied, "Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism," *Int. J. Social Robot.*, vol. 5, no. 4, pp. 593–618, Nov. 2013.

[15] T. Sudo, T. Herai, and K. Mogi, "Egocentric mental transformation of self: Effects of spatial relationship in mirror-image and anatomic imitations," *Exp. Brain Res.*, vol. 221, no. 1, pp. 27–32, Aug. 2012.

[16] C. Pierpaoli, L. Ferrante, T. Manzoni, and M. Fabri, "Anatomical or mirror mode imitation? A behavioral approach," *Arch. Italiennes Biologie*, vol. 152, pp. 20–31, Mar. 2014.

[17] S. Spasojević, T. V. Ilić, S. Milanović, V. Potkonjak, A. Rodić, and J. Santos-Victor, "Combined vision and wearable sensors-based system for movement analysis in rehabilitation," *Methods Inf. Med.*, vol. 56, no. 2, pp. 95–111, 2017.

[18] S. Ali *et al.*, "An adaptive multi-robot therapy for improving joint attention and imitation of ASD children," *IEEE Access*, vol. 7, pp. 81808–81825, 2019.

[19] F. Mehmood, Y. Ayaz, S. Ali, R. De Cassia Amadeu, and H. Sadia, "Dominance in visual space of ASD children using multi-robot joint attention integrated distributed imitation system," *IEEE Access*, vol. 7, pp. 168815–168827, 2019.

[20] J. Greczek, E. Kaszubski, A. Atrash, and M. J. Mataric, "Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders," in *Proc. 23rd IEEE Int. Symp. Robot Hum. Interact. Commun.*, Aug. 2014, pp. 561–566.

[21] S. M. Nguyen, P. Tanguy, and O. Remy-Neris, "Computational architecture of a robot coach for physical exercises in kinaesthetic rehabilitation," in *Proc. 25th IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2016, pp. 1138–1143.

[22] W.-C. So *et al.*, "Using a social robot to teach gestural recognition and production in children with autism spectrum disorders," *Disab. Rehabil., Assistive Technol.*, vol. 13, no. 6, pp. 527–539, Aug. 2018.

[23] L. Santos, A. Geminiani, I. Olivieri, J. Santos-Victor, and A. Pedrocchi, "Copyrobot: Interactive mirroring robotics game for asd children," in *15th Mediterranean Conference on Medical and Biological Engineering and Computing*, J. Henriques, N. Neves, and P. de Carvalho, Eds. Cham, Switzerland: Springer, 2020, pp. 2014–2027.

[24] A. Geminiani *et al.*, "Design and validation of two embodied mirroring setups for interactive games with autistic children using the NAO humanoid robot," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 1641–1644.

[25] *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Arlington, VA, USA: American Psychiatric Association, 2013.

[26] V. Hus and C. Lord, "The autism diagnostic observation schedule, module 4: Revised algorithm and standardized severity scores," *J. Autism Develop. Disorders*, vol. 44, no. 8, pp. 1996–2012, Aug. 2014.

[27] Y. Hou, Z. Li, P. Wang, and W. Li, "Skeleton optical spectra-based action recognition using convolutional neural networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 3, pp. 807–811, Mar. 2018.

[28] B. Scassellati, H. Admoni, and M. Matarić, "Robots for use in autism research," *Annu. Rev. Biomed. Eng.*, vol. 14, no. 1, pp. 275–294, Aug. 2012.