Quantifying patterns of joint attention during human-robot interactions: An application for autism spectrum disorder assessment

Salvatore Maria Anzalone, Jean Xavier, Sofiane Boucenna, Lucia Billeci, Antonio Narzisi, Filippo Muratori, David Cohen, Mohamed Chetouani

Pattern Recognition Letters
journal homepage: www.elsevier.com

Quantifying Patterns of Joint Attention during Human-Robot Interactions: an Application for Autism Spectrum Disorder Assessment

Salvatore Maria Anzalone[a,**], Jean Xavier[b], Sofiane Boucenna[d], Lucia Billeci[e], Antonio Narzisi[f], Filippo Muratori[f], David Cohen[b], Mohamed Chetouani[c]

[a] *CHArt Laboratory, Université Paris 8 Vincennes - Saint Denis, 2 Rue de la Liberté, 93526 Saint-Denis, France*
[b] *Department of Child and Adolescent Psychiatry, Groupe Hospitalier Pitié-Salpêtrière, 47 bd. de l'Hôpital, 75013 Paris, France*
[c] *Institute of Intelligent Systems and Robotics, Université Pierre et Marie Curie, 4 place Jussieu, 75005 Paris, France*
[d] *Equipes traitement de l'information et systémes (ETIS), Université de Cergy-Pontoise, 33, bvd. du Port, 95014 Cergy, France*
[e] *Institute of Clinical Physiology, National Research Council, Via Moruzzi 1, 56124, Pisa, Italia*
[f] *Fondazione Stella Maris, Viale del Tirreno, 331, 56018, Calambrone (Pisa), Italia*

ABSTRACT

In this paper we explore the dynamics of Joint Attention (JA) in children with Autism Spectrum Disorder (ASD) during an interaction task with a small humanoid robot. While this robot elicits JA in children, a coupled perception system based on RGB-D sensors is able to capture their behaviours. The proposed system shows the feasibility and the practical benefits of the use of social robots as assessment tools of ASD. We propose a set of measures to describe the behaviour of the children in terms of body and head movements, gazing magnitude, gazing directions (left vs. front vs. right) and kinetic energies. We assessed these metrics by comparing 42 children with ASD and 16 children with typical development (TD) during the JA task with the robot, highlighting significant differences between the two groups. Employing the same metrics, we also assess a subgroup of 14 children with ASD after 6-month of JA training with a serious game. The longitudinal data confirms the relevance of the proposed metrics as they reveal the improvements of children behaviours after several months of training.

1. Introduction

Autism Spectrum Disorder (ASD) is characterized by deficits in social communication and social interaction and by restricted and repetitive patterns of behaviours, interests and activities (Xavier et al., 2015). Although heterogeneous in term of severity, ASD symptoms can significantly impair normal, daily-like activities, in particular social activities. Major genetic risk factors have been found (Betancur, 2011). Yet, in many cases, scientists still do not know the exact cause of the disorder and a combination of genetic and environmental risk factors has been supposed (Tordjman et al., 2014; Guinchat et al., 2012b,a). The assessment of ASD is possible since the early childhood, typically during the first two years of life (Ouss et al., 2014). Early diagnosis is fundamental to limit the disorder's effects, allowing clinicians to deploy early intensive behavioural intervention. These protocols take advantage of the learning potential that the children's brain has, focusing in particular on the learning prerequisites that infants should develop for the acquisition of new skills. There is an ongoing evidence suggesting that Applied Behavioural Analysis (ABA) teaching methods and developmental approach such as the Early Start Denver Model (ESDM) are able to help children on remediating areas of weakness by improving their Intelligent Quotient (IQ), their language abilities and their social interaction skills (Narzisi et al., 2014; Reichow and Wolery, 2009; Ospina et al., 2008). In particular, the ESDM approach, suitable since the early childhood, tries to integrate into the ABA practices a relationship-focused developmental model in which the toddlers' development is view as an interpersonal process strongly enhanced by the sensitivity and responsivity to children's cues (Dawson et al., 2010). According to this framework, the child preferences, choices and motivations will guide the intervention of the therapists and

**Corresponding author: Tel.: +0-000-000-0000; fax: +0-000-000-0000;
e-mail: sanzalone@univ-paris8.fr (Salvatore Maria Anzalone)

of the parents. The ESDM approach proposes a set of activities in naturalistic scenarios in which positive reinforcements and affect-based relationship are used to the enhancement of interpersonal skills: the child will be able to develop and reinforce social-emotional skills, cognitive competences, and language by playing with toys in a natural playground, guided by an interdisciplinary team of therapists composed by special education teachers, developmental and clinical psychologists, speech-language pathologists, occupational therapists and behaviour analysts, with the key contribution of the parents.

Together with the difficulties of a correct assessment of Autism Spectrum Disorders (ASD) in toddlers (Xavier et al., 2015), recent studies have shown that several factors can limit the effectiveness of therapeutic interventions (Narzisi et al., 2014; Reichow and Wolery, 2009; Ospina et al., 2008). One of such weaknesses is the artificiality of the context in which the therapeutic activities are usually performed, a hospital or other "laboratory setup". Activities accomplished in such contexts could entail a lack of generalisation. At the same time, activities performed at the hospital are provided just few hours during the week, drawing another important weakness in the intensity of the therapy. To overcome such limitations researches have focused on the integration of information and communication technologies (ICT) in the "classic" protocols for both assessment and treatment of ASD (Boucenna et al., 2014b). Recently, an increasing number of research teams have focused on the use of social robots in ASD treatment (Pennisi et al., 2015; Scassellati et al., 2012; Diehl et al., 2012). Although most research in this field may be regarded as preliminary by clinicians, those kind of robots emerged as an important tool for children with ASD because of the various benefits their use could bring to the therapy, regarding:

- **Complexity**: social robots in ASD therapy could simplify the inner complexity of the social interactions. While interacting, people exchange an enormous amount of information in both verbal and non-verbal forms (speech, words, prosody, facial expressions, emotions, proxemics, and so on). As social robots are entirely controlled by robot programmers, the behaviours they can express and the interaction proposed could be very simple and predictable. Clinicians and robotics engineers can take advantage of this, developing new experimental protocols using social robots, focusing just on one or few aspects of the interaction, or just on the building blocks of the sociality, simplifying the cognitive load required to "decode" such interactions.

- **Embodiment**: social robots can communicate and interact in a multimodal way with children, but, unlikely to serious games, avatars, or other software agents, they have their own "physical presence" in the real world. The embodiment of social robots will permit physical explorations and interactions with the environment (Kozima et al., 2005) as well as a communication with people based also on gestures and touch, widening the possibility of their employment in therapeutic protocols (Boucenna et al., 2016). In addition, there is a growing evidence that ASD patients may perceive a humanoid robot as a social partner (Chaminade et al., 2012).

- **Shape**: the shapes of social robots used in ASD treatment are different, according to their role in the interaction and to the goal of the interaction itself. Android, human-like, animal-shaped, non-anthropomorphic coloured toy: in any case, the shape of the robot should contribute to the reduction of the stress of the children during the experiment, making them comfortable and at ease (Scassellati et al., 2012).

1.1. The Michelangelo Project

The Michelangelo Project, funded by the European Commission, proposed several, cost-effective, technology tools to bring the ASD assessment and therapy to the home setting. As part of this project, the Michelangelo Study Group developed a network made up by different sensors (Ghidoni et al., 2014), such as cameras, RGB-D sensors, microphones, wearable systems for electroencephalographic (EEG) (Cester et al., 2008) and electrocardiogram (ECG) (Billeci et al., 2015) signals recording, to capture the fine detail of the behaviour of the children in controlled environments (Anzalone and Chetouani, 2013; Cruciani et al., 2010).

This network of sensors has been used to assess imitation and JA skills in children with ASD. In particular, the current study exploits cameras and RGB-D sensors during an interactive game developed by the Michelangelo Study Group that involves the use of a small humanoid robot (Anzalone et al., 2014b; Boucenna et al., 2014a). Focusing on JA, the game tracks the behaviour of the child while exploring the world. JA is a key element of social cognition. It involves a triadic interaction and can be defined as a process in which two individuals share their gaze over the same focus of attention (Emery, 2000). One agent alerts a second one of his attention towards an object by eye gazing, by pointing, or by using other verbal or non-verbal indication. Then, the second agent follows the suggested direction towards the same object. It is important to highlight that this definition slightly differs from the shared attention, which implies a coupling between mutual attention and JA, making both the agents aware of the attention of the other: "I know that you are looking at the object, and you know that I am looking at the object" (Tomasello, 1995).

At the same time, the study group developed a Gaming Open Library for Intervention in Autism at Home, GOLIAH (Bono et al., 2016), a therapeutic game based on the Early Start Denver Model (Narzisi et al., 2014; Reichow and Wolery, 2009; Ospina et al., 2008). The game presents a set of activities that can be done by the child with the help of a clinician in the hospital or with a parent at home. Each activity focuses on the training of specific abilities, in particular imitation and joint attention (JA). The focus on such skills is justified by their importance as "building blocks" for the development of social cognition (Toth et al., 2006; Nadel, 2006).

As in Fig. 1, children have been followed for 6 months of training sessions using GOLIAH, at the hospital but also at home. At the beginning and at the end of the treatment, ses-

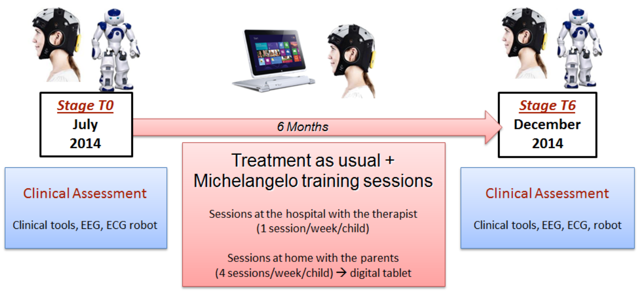


Fig. 1: Michelangelo Project's ASD treatment protocol.

sions of assessment have been performed employing the robot game.

1.2. Hypothesis

In a previous pilot study (Anzalone et al., 2014b), we already explored the behaviour of children with ASD during a JA elicitation task involving a Nao robot, the Softbank Robotics's small humanoid robot. The proposed protocol take advantage of a RGB-D sensor to capture the movements of the child, while the robot is employed as powerful tool to induce JA. The robot is able to engage TD children by exchanging simple, multimodal, social signals, taking advantage of its simplified but communicative shape, able to reduce the complexity of the interpersonal interactions. In this study we show that, while the robot was able to engage TD children, the response to the JA induction was lower in ASD children than in Typical Development (TD) children, despite we selected and matched them on developmental age and on their ability to perform JA with a human partner.

In this paper, we expand the group of children with ASD and, we introduce new behavioural metrics (Anzalone et al., 2015) exploiting the RGB-D sensor data, based on the analysis of the child's displacement, his gazing and of his kinematic energy. In the following, we illustrate the details of the JA elicitation experimental protocol using a small humanoid robot and of the features employed to describe the children's behavioural response. Then, we propose a statistical analysis of the data obtained in several experiments involving children in typical development (N=42) and children with ASD (N=16). Finally, we assess through the same metrics a subgroup of children with ASD (N=14) that has been trained using GOLIAH. In particular, we compare their behaviours at baseline and after 6 month of intensive training, four time a week at home and one time a week at hospital.

Given the results of our previous pilot study (Anzalone et al., 2014b; Boucenna et al., 2014a), we hypothesized that:

1. the response to the robot's JA elicitation task is lower in ASD children;
2. the postural stability induced by the engagement to the joint activity with the robot, expressed in terms of:
    - displacement in the space,
    - kinematic energy,

   is lower in ASD children;
3. the children showing improvement after GOLIAH training, improve also their performance in the JA elicitation task with Nao.

2. Methods

The JA induction experiment we proposed involves the interaction of a child with a Nao robot used to elicit JA behaviours in a laboratory setup. Data collected from a RGB-D sensor during the experiment are elaborated offline to extract a set of descriptors able to represent behavioural information of each child.

2.1. Participants

Table 1 summarizes participants' socio-demographics and clinical characteristics. Children were recruited by two specialized clinics for autism in the Pitié-Salpétrière hospital (Paris, France) and the Stella Maris Foundation (Pisa, Italy). Both teams agreed on participants' inclusion and exclusion criteria. ASD children selected suffered from various social impairments, including language disabilities and poor communicative skills. Assessment of ASD symptoms has been performed using the Autism Diagnostic Interview-Revised (Lord et al., 1994). The psychiatric assessments and parental interviews were conducted by three clinicians who specialized in autism (AN, JX, DC). The developmental age was assessed using a cognitive assessment. Depending on the children abilities and ages, we used either the Wechsler Intelligence scales (Wechsler, 1949), the Kaufman-ABC (Kaufman and Kaufman, 1983) or the Psychoeducational Profile Revised (PEP-3) (Schopler and Reichler, 1976).

TD participants were recruited from several schools in the Paris area. TD participants met the following inclusion criteria: no verbal communication impairment, no intellectual disability (ID), and no motor, sensory or neurological disorders. TD participants were matched to the children with ASD with respect to their developmental ages and genders. For the TD group, the developmental and chronological ages were considered to be the same.

2.2. Experimental Setup

The Michelangelo Room is where the therapy and the experiments with the robot take place. It is composed by an experimental area and a hidden control room. The experimental area has been customized for the experiment of joint attention induction. A small, thin, squared carpet has been placed in the position that should be occupied by the child. In front of it, a robot has been placed at around 1mt of distance, as in Fig. 2. The robot used is a Nao robot, from Aldebaran Robotics, a small humanoid able to communicate through simple body movements and verbal language. On its feet, a RGB-D sensor, a Microsoft Kinect is conveniently placed in order to capture the full body of the child and in particular his face. The robot and the RGB-D sensor are placed over a small table to keep its head at the same height of the child. On the two sides, two Focus of Attention (FoA) are placed: an image of a dog and an image of a cat. According to the protocol, the robot will try to induce JA over them. Several cameras allow a researcher hidden in the

Participants at baseline to select pertinent metrics		
	ASD (N=42)	Typical Development (N=16)
Age, mean (± SD), year Male – Female	7.94 (± 1.67) 37 – 5	8.06 (± 2.49) 10 – 6
Diagnosis	Autism: N=10 Asperger: N=4 ASD: N=28	No diagnosis
ADI-R, current, mean (± SD) Social impairment score Communication score Repetitive interest score Developmental score	 12.26 (± 5.02) 10.54 (± 5.85) 3.24 (± 2.55) 3 (± 1.41)	Not administered
Developmental age IQ	7.47 (± 2.9) 89 (± 18.2)	8.06 (± 2.49) All controls ¿ 80
Participants with ASD trained with GOLIAH for 6 months (N=14)		
Age, mean (± SD), year Male – Female	6.85 (± 1.34) 14 – 0	Not appropriate
Diagnosis	Autism: N=3 Asperger: N=2 ASD: N=9	Not appropriate
ADI-R, current, mean (± SD) Social impairment score Communication score Repetitive interest score Developmental score	 14.4 (± 4.58) 10 (± 5.82) 4 (± 2.91) 3 (± 1.36)	Not appropriate
Developmental age IQ	6.67 (± 3.2) 98.8 (± 20.1)	Not appropriate

Developmental age and IQ assessed with the Vineland Developmental Score, the Psycho-Educational Profile-Revised, the Kaufman Assessment Battery for Children or the Wechsler Intelligence Scale for Children.
ASD = Autism Spectrum Disorder; SD = Standard Deviation; ADI-R=Autism Diagnostic Interview-Revised; GAF =Global Assessment Functioning.

Table 1: Socio-demographic and clinical characteristics of the participants.

control room to have a complete knowledge about what is going on in the experimental area while managing and controlling the robot. For security reasons, he can always take the complete control of the robot. The control room has been equipped with two computers, one to control the robot and acquire the raw data from the RGB-D sensor, one as Michelangelo base station that centralizes all the other cameras displaced on the environment. Through this second computer, the hidden operator can monitor and record video data and events from the experimental room. As the experiments were conducted in the two clinics involved in the project, in Pisa and in Paris, identical Michelangelo Rooms have been set up in the two cities.

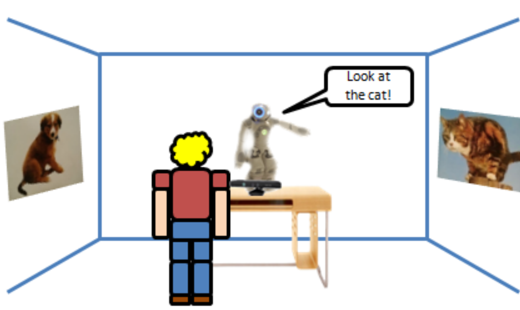


Fig. 2: The experimental area of the Michelangelo Room.

2.3. Protocol

The design of the protocol conceived for this experiment focuses on the induction of children spontaneous reaction by the small humanoid robot Nao. Each child is introduced in the Michelangelo room by a clinician. The child is guided to establish a first contact with the robot. This stage gives to the child some time for getting used to the robot as a new toy. A second researcher is hidden in a separate control room, following the interaction, able to intervene in case of unexpected behaviors of the child that may either endanger the robot or the child himself. Then, the child is invited to stay over a small, thin carpet, in front of the robot, and then the experiment starts. In the joint attention experiment proposed, the robot tries to induce attention over the two focuses placed on the side of the room. The robot increases in three times the amount of information passed to the child by using more modalities to communicate its attention, as in Fig. 3: just by gazing, using head movements; by gazing and by pointing with the hands; by gazing, pointing and vocalizing, saying "look at the cat", "look at the dog". The behaviours of the robot are offline scripted, being designed offline by mimicking the interpersonal interplay between humans. Through the implementation of simple, recognizable, stereotyped behaviors the dynamics of the interaction becomes simplifyed and its complexity reduced. The child is free to decide to stay in place or go away, to move in place, to respond to the induction of the robot by gazing or by hand gestures. Behavioural data from each child is recorded and analysed offline.

The protocol presented in this paper was approved by the Pitié-Salpêtrière hospital ethics committee for French participants and by the Stella Maris hospital ethics committee for Italian participants. All of the parents received information on the experiment and gave written consent before the participation of their child.

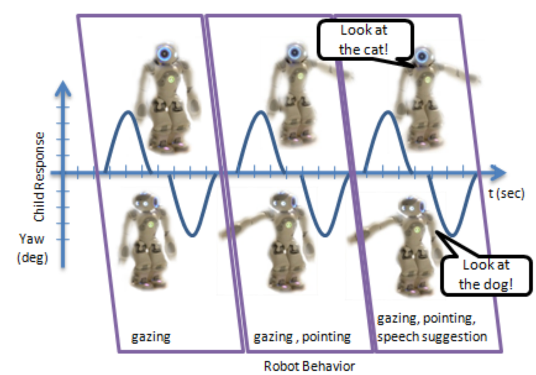


Fig. 3: The joint attention elicitation protocol followed by the robot.

3. Children Behaviour Metrics

The behaviours of each child, coupled with the robot behaviour, are analysed in order to extract a set of descriptors able to measure in a detailed way the response of each child to the JA elicitation. Such descriptors, introduced in the following, will be able to depict the JA response in terms of gazing behaviour, body movements and the kinetic energy.

3.1. Data Acquisition Pipeline

The information of children's movements are acquired through the elaboration of the data perceived by a RGB-D sensor (Anzalone et al., 2014b; Anzalone and Chetouani, 2013; Anzalone et al., 2015, 2014a). As shown in Fig. 4, the body of the participant is firstly distinguished from the background through a subtraction algorithm, then the skeleton in world coordinate is calculated (Shotton et al., 2013; Han et al., 2013). At the same time, the RGB image is cropped around the position in which, according to the skeleton information, the head should appear, then a face pose tracking algorithm is applied (Xiong and De la Torre, 2013). High level descriptors, detailed in the following, are finally extracted from such data to characterize gazing, body movements, and the kinetic energies from the body and from the head movements.

Fig. 4: The RGB-D data processing pipeline able to estimate people gaze, body position and posture.

3.2. Response Events

During the experiment, the robot tries to induce joint attention towards the two focuses placed on the left and right sides of the room. The child can reply or not by gazing towards them or by hand gestures. It is possible to retrieve the presence of those responses by verifying whether or not the child moves his head or his arms. The detection of such movements can be performed through a short time spectrum analysis of the energy of the head movements and of the arms movements of the child (Godfrey et al., 2008). In such spectra, each movement appears as a linear combination of components at different frequency. Slow movements will be characterized by peaks in low frequencies, while fast movements will produce peaks in high frequencies.

After an empirical analysis of the videos of children movements, slow, significant movements, are isolated and selected by thresholding the components at 1Hz. Such slices are, then, fused together using the DBScan algorithm (Ester et al., 1996), obtaining clusters of temporal slices. When they exist, the center of the cluster closest temporally to the induction event of the robot is considered as the response event. The number of the response events that took place during the experiment can be considered as a measure of the effectiveness of the JA elicitation. A higher response of TD children is expected.

3.3. Displacement

The protocol proposed in this work requests the children to spend some time in front of the robot. A measure of displacement around each one's average position could be informative, together with gaze direction, about the engagement towards the current activity. Children focused on the activity will move in the space less than inattentive children. A two-dimensional histogram of the displacement is a convenient way to depict the participant's behaviour (Fig. 5). The magnitude of the displacement and the displacement along two preferential directions, left-right and forward-backward axis, have been selected as descriptors able to resume the displacement information. A higher displacement in the space of ASD children is expected.

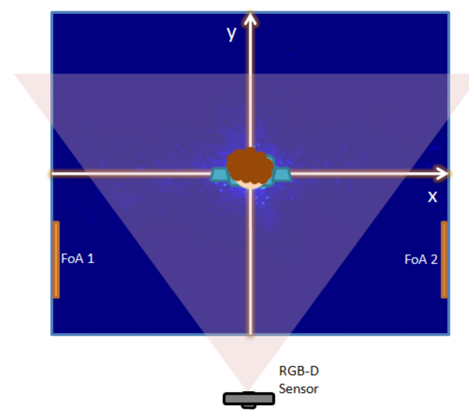


Fig. 5: Histogram of the displacement of the child in the environment represented as heatmap. Colors indicate the areas around their own average in which the child moved.

3.4. Gazing

The robot elicit JA behaviours that is mainly expressed by children through a gazing response. In the considered scenario the position of the focus of attentions are chosen with the explicit goal of forcing head movements towards them. In these terms, such movements can be seen as an approximation of the gazing. Those movements are captured in terms of pitch and yaw, as shown in Figure 6, then modelled in different ways.

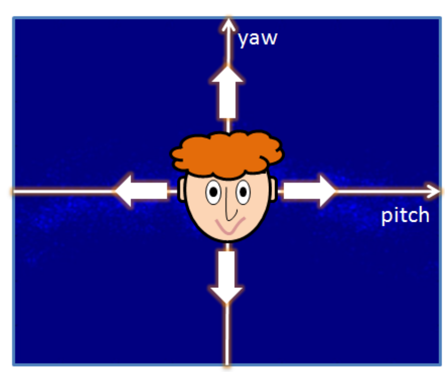
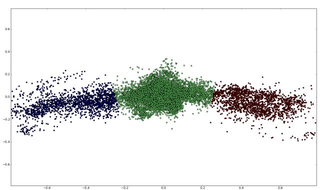


Fig. 6: Histogram of the gazing movement of the child represented as heatmap. Colors indicate the areas towards the child looked at.

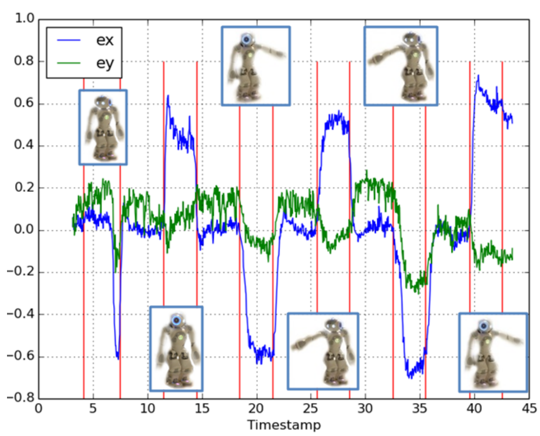
A two-dimensional histogram of the head movement can easily illustrate how children explore the environment. In particular, the displacement of the gaze from each one's average position has been considered in terms of: (i) the magnitude of the head displacement, to depict the movement of the head without a particular prevalent direction; (ii) the head displacement along the pitch axis and yaw axis.

Head movement data from children in typical development has been employed to compute a model of their gazing through machine learning. This model has been employed, then, to evaluate the ASD gazing behaviour. Here we employed k-Means to find three clusters (K=3) corresponding to the three focus of attention: looking towards the robot, looking towards the left and the right focus of attention. Data from each participant has been categorized according to such model. Then, the numbers of samples belonging to each cluster, representing how much the child spent looking towards the cluster's direction, and their displacement measures, both in terms of magnitude and along the pitch and the yaw axis, can be calculated for each cluster. Differences in clusters dimensions and in the amplitude of the head movements are expected in TD children, highlighting the effectiveness of joint attention induction.

(a) The child interacting.



(b) The head movements clusters



(c) The child's head trajectory.

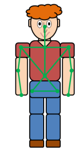
Fig. 7: The robot elicits JA behaviours to the child in front of him (a); gaze clusters generated using k-means to locate the three main focuses of attention of the Michelangelo Room (b); child head movements in terms of yaw (x-axis) and pitch (y-axis) retrieved through the RGB-D data processing, while in red the beginning and the end of each JA induction event performed by the robot (by gazing, by gazing and pointing, by gazing, pointing and vocalizing) (c).

3.5. Energy

To describe the quantity of the movement of each participant, the energy of the upper body as well as the energy of the head have been calculated. This has been expressed in terms of total kinetic energy, the sum of the translational and of the angular energy, as in Eq. 1:

$$E(t) = \sum_{k=limbs} \left(\frac{1}{2} M_k v_k^2(t) + \frac{1}{2} I_k \omega_k^2(t) \right) \quad (1)$$

where k is the identifier of a limb, M_k is its mass, I_k is its inertia, $v_k(t)$ is its translational speed and $\omega(t)$ is its angular speed.



	Middle Trunk	Upper Trunk	Head
Trunk	14.65	15.45	6.68
Arm	Upper Arm	Forearm	Hand
Left	2.55	1.38	0.5
Right	2.55	1.38	0.5

Fig. 8: Body skeleton and percentage masses for each upper body part (De Leva, 1996).

For the upper body, only the translational energy has been calculated, while the angular energy has been used just for the head movements. As a measure of the quantity of movement of the body limbs, the contribution of the locomotion of the body on the ground has been subtracted, calculating the energy from the displacement of the body around the torso.

Masses of each limb are calculated using the Zatsiorsky's model adjusted by de Leva (De Leva, 1996), shown in Figure 8, considering an average total body mass of 25kg as the average weight of children in the considered population (Kuczmarski et al., 2002). Inertia of the head has been calculated by modelling it as an solid sphere of 8.35cm of radius, as 52.5cm head circumference (Rollins et al., 2010), ($I = \frac{2}{5}MR^2$). The energy is described in terms of its median. According to previous results (Anzalone et al., 2014b), higher energies are expected in movement of children with ASD.

4. Results

4.1. Joint attention elicitation effects at baseline

Statistical analysis were conducted to compare the behaviour of the TD population with the behaviour of the ASD population. Because of technical issues during recording, full data were available for 37 participants (ASD: N=25 vs. TD: N=12). We used the Mann-Whitney-Wilcoxon test for independent samples to compare the two groups according to the modality employed by the robot to elicit behaviours (gazing, pointing, vocalizing). Table 2 summarizes the main extracted features. The comparison between the behaviour of TD children and children with ASD revealed several statistically significant differences. As expected, the analysis shows that children with TD respond more than children with ASD to the JA induction performed by the robot, in terms of head movements responses to JA induction.

By focusing on the gazing movement clustering, it is possible to observe a significant difference on how much time children spend on gazing towards the two focus of attention: children with ASD spend less time gazing towards the focus of attention than TD children, while they spend more time looking in front. This confirms the first hypothesis.

From the analysis of the displacement of children from their average position, it emerges a significant difference in the behaviour of the two studied groups: the amplitude of the displacement on the ground is higher for children with ASD than for TD children. Moreover, the analysis shows a preferential direction along the left-right axis. In addition, the analysis shows also an important difference on the kinematic energy of the body: ASD children employ a higher amount of energy than that of TD children. The analysis of children head's movements

Feature	$\overline{TD}$	$\sigma(TD)$	$\overline{ASD}$	$\sigma(ASD)$	Variation
Head movement response to JA induction	75%	29%	42%	28%	↘*
Gazing std magnitude	10.2°	3.9°	2.9°	7.3°	↘
Gazing std yaw	14.9°	6.0°	9.7°	3.8°	↘*
Gazing std pitch	4.4°	1.5°	4.7°	2.2°	
Gazing frequency towards the front	73%	14%	88%	8%	↗*
Gazing frequency towards the left FoA	13%	8%	7%	4%	↘*
Gazing frequency towards the right FoA	14%	7%	5%	5%	↘*
Displacement std magnitude	0.9 cm	0.4 cm	2.8 cm	3.8 cm	↗
Displacement std left-right	1.4 cm	1.2 cm	3.9 cm	4.9 cm	↗
Displacement std front-back	1.2 cm	0.4 cm	2.3 cm	2.0 cm	
Body energy median	0.162 mJ	0.095 mJ	0.775 mJ	0.925 mJ	↗*
Head energy median	0.01 mJ	0.003 mJ	0.022 mJ	0.025 mJ	↗

↗= ASD>TD, ↘= ASD<TD, while $p \leq 0.05$; * if $p \leq 0.01$.

Table 2: Features comparison results for the JA experiment between TD (N=12) and ASD children (N=25) and their statistical significance.

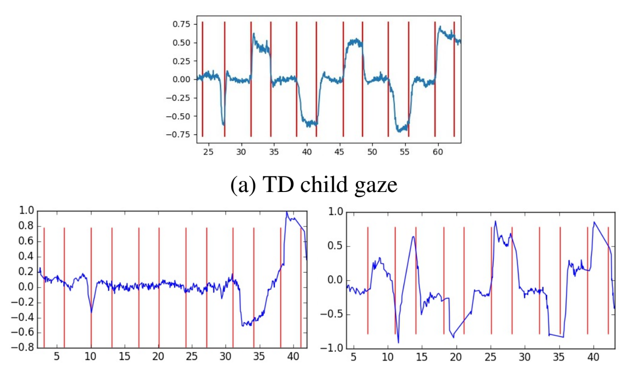
energy shows a similar behaviour: in the case of ASD group, the energy employed for head's movement is higher than in TD children Those results can be interpreted as a measure of the stability of children's body, confirming the second hypothesis of this work: children with ASD are less stable in terms of displacement in the environment, as well as in terms of their body movements.

4.2. 6-month follow-up assessment of children with ASD exposed to GOLIAH

Among the 14 individuals with ASD trained with GOLIAH, we only had 8 individuals with full data available at both baseline and 6-month follow up. Due to the low number of GOLIAH users, no statistical test was conducted. However, a qualitative description of the results is given in Table 3. We only present the features that were significantly different at baseline between ASD and TD children (see Table 3). As shown in the middle column, it seems that features extracted during the JA task at 6 months from children with ASD tend to change in a direction that is closer to TD value.

In particular, the response to JA increases in ASD children after 6 months of GOLIAH training, showing an incremented reaction of the children to the behaviours elicited by the robot. Gazing movements in terms of standard deviation tend towards the results obtained in TD group. The same tendency is confirmed by a detailed analysis of the gaze, in terms of focus of attention clusters: here, the time spent looking frontally, towards the robot, decrements after 6 months of GOLIAH training, turning towards the TD children behaviour; at the same time, the time spent looking the focus of attention on the two side of the room increments. Despite of this, children body displacement and kinematic energy of this experimental group still remains significant and higher compared to the TD group. While those results seem paradoxical, they could evidence the existence of impairments in postural control described in children with ASD, with increased sways and meaningful variability in posture associated to a micro-instability in their movement's trajectories (for a review, see Lim et al. (2017)), named micro-movements by Torres et al. (2013). Such micro-instability should emerge in the features describing the child's motor control and, in particular, in the energies. At the baseline, such children do not use to respond to the JA induction; despite this inactivity, elevated energies reveal such behavioural micro-instability. After six months from the baseline, children start to respond to the behaviour induction: in this case, the higher energies are explainable as the mutual contribution of JA behaviour and micro-instability.

In the following section, we detail a single case with a rather positive evolution. The behaviour of a child at the beginning of GOLIAH training is compared with the behaviour of the same child after 6-months. The behaviour of a TD child is employed as basis for comparison. The gaze of the ASD child at baseline reveals an effective response to the JA induction elicited by the robot just on the vocalization stage (Figure 9.b). After 6 months (Figure 9.c), the child respond to almost all the elicitation behaviours. Gazing trace clearly shows this change: the behaviour at 6 months (Figure 9.c) tends to the behaviour of TD children (Figure 9.a).



(a) TD child gaze

(b) ASD gaze baseline

(c) ASD gaze followup

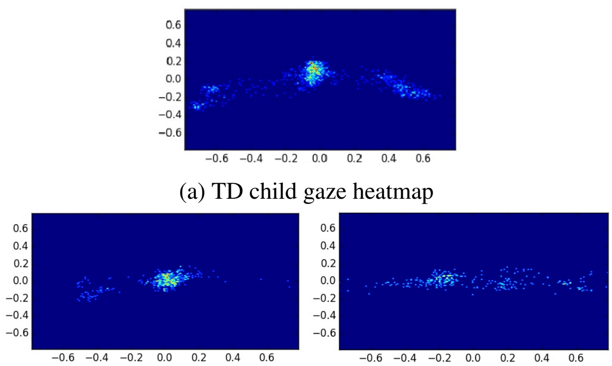
Fig. 9: Gazing movement trajectory for a sample TD child and for a sample ASD child, at baseline and after 6 months.

The result is confirmed by the analysis of gazing heat maps. At the beginning of the therapy the child spent the most part of his time by looking in front of him, towards the robot: the histogram presents a peak on the central cluster (Figure 10.b). After 6 month, the gazing is distributed in a uniform way over the environment, along the left-right axis (Figure 10.c). In any case, the histogram is still quite different compared to the behaviours of TD children (Figure 10.a).

The displacement histogram at baseline shows an almost complete inactivity of the child 11.b). At 6 months from the

Feature	ASD Baseline	σ	ASD after 6 months	σ	TD baseline	σ
Head movement response to JA induction	37%	2%	58%	3%	75%	29%
Gazing std magnitude	8.2°	2.4°	9.2°	3.6°	10.2°	3.9°
Gazing std yaw	10.5°	3.0°	13°	5.5°	14.9°	6.0°
Gazing std pitch	4.8°	2.7°	4.5°	1.3°	4.4°	1.5°
Gazing frequency towards the front	87%	7%	80%	13%	73%	14%
Gazing frequency towards the left FoA	8%	4%	10%	6%	13%	8%
Gazing frequency towards the right FoA	5%	4%	10%	7%	14%	7%
Displacement std magnitude	1.6 cm	1.59 cm	1.7 cm	0.7 cm	0.9 cm	0.4 cm
Displacement std left-right	2.2 cm	2.5 cm	2.4 cm	1.5 cm	1.4 cm	1.2 cm
Displacement std front-back	1.6 cm	1.0 cm	1.9 cm	0.9 cm	1.2 cm	0.4 cm
Body energy median	0.473 mJ	0.527 mJ	0.689 mJ	0.54 mJ	0.162 mJ	0.095 mJ
Head energy median	0.023 mJ	0.02 mJ	0.039 mJ	0.039 mJ	0.01 mJ	0.003 mJ

Table 3: Features extracted during the JA induction task at base line and 6 month follow-up for individuals with ASD trained with GOLIAH (N=8) and at baseline for TD children (N=12).

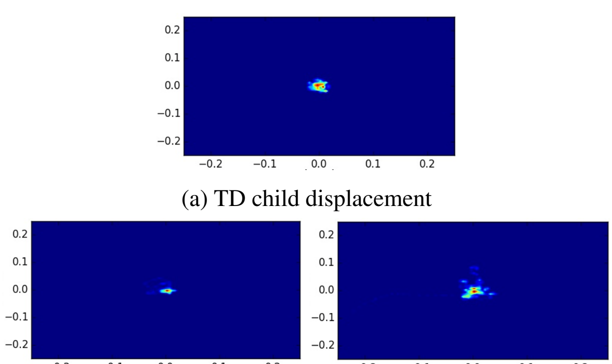


(a) TD child gaze heatmap

(b) ASD gaze heatmap baseline (c) ASD gaze followup

Fig. 10: Gazing heatmaps for a sample TD child and for a sample ASD child, at baseline and after 6 months.

beginning of the therapy, instead, the histogram shows more mobility 11.c), reflecting the incremented response to the behaviour elicitation. In any case, the histogram presents more mobility in this case than in TD children 11.a): this is compatible with the previously introduced results, for which ASD children are generally less stable than TD children.

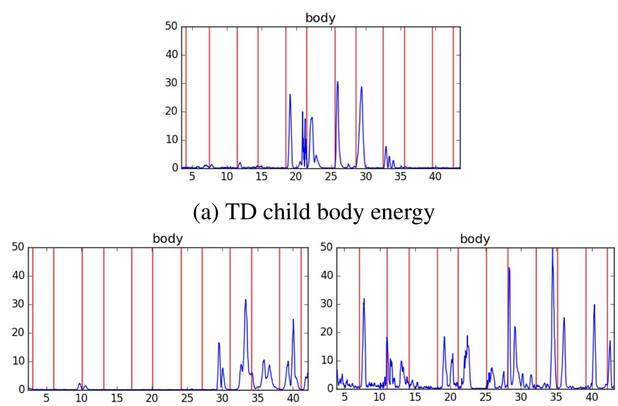


(a) TD child displacement

(b) ASD displacement baseline (c) ASD displacement followup

Fig. 11: Displacement heatmap for a sample TD child and for a sample ASD child, at baseline and after 6 months.

Variation between the behaviour of the sample ASD child at the beginning of the therapy (Figure 12.b) and after 6 months (Figure 12.c) are also highlighted by the kinematic energy of his body: the child responds more to JA in terms of his body movements after 6 months of training. Moreover, the energy employed by the ASD child is always higher than the one employed by the TD child, highlighting how the joint activity with the robot induces a lower postural stability in the ASD child than in the TD child. More in detail, the energy employed by the ASD child during the proposed task is lower than the one measured after 6 months, supporting the hypothesis of the contribution of micro-instability to the JA behaviour.



(a) TD child body energy

(b) ASD body energy baseline (c) ASD body energy followup

Fig. 12: Body movement energy for a sample TD child and for a sample ASD child, at baseline and after 6 months.

5. Conclusions

This paper presents an experimental protocol focusing on the use of a small humanoid robot as useful and natural tool to elicit behaviours in a JA taks. A set of metrics based on the analysis of body and head movements, gazing magnitude, gazing directions (left vs. front vs. right) and kinetic energies, able to describe children's behaviour during the task with the robot has been introduced. Results from experiments with TD children and children with ASD show the usefulness and the benefits of the presented protocol as well as of the informativeness of the metrics introduced. Such metrics show potential to measure JA characteristics during natural interaction since we were able (i) to distinguish TD children and children with ASD; (ii) to show improvements of children towards the behaviour of TD group after 6-month training of JA using GOLIAH. The proposed measures, however, do not have the ambition of becoming a standard that can be used in different institutions and hospitals: they should be seen as dependents to the particular scenario in which they are employed. A standardization of such metrics would involve their validation across different scenarios, employing bigger populations than the one that participated to our experiment. On the contrary, it is possible to generalize the proposed methodology in similar scenarios to evaluate JA. In this case, measures should be adapted to the specific context and to the specific protocol.

Nevertheless, the protocol proposed arouses several questions. In particular, the proposed protocol is implemented in a laboratory setup, a strongly controlled environment. The degrees of freedom of the proposed interaction with the robot are still very limited. Thus, the behaviour of the child is in some way very constrained by such factors. Those constraints come mainly from the limitations of the technology employed. The sensor employed to monitor the child activity is a single RGB-D sensor that has a limited field of view and is not able to track people in the entire environment (Anzalone et al., 2011). The implementation of the protocol in a space equipped with a network of synchronised RGB-D sensors will permit to the researcher to follow and analyse a wider range of movements of the child, as well as objects in the environment and other people, as clinicians (Ghidoni et al., 2014; Anzalone and Chetouani, 2013). In such situation, the robot would be able not only to induce more complex behaviours from the children, but would also actively participate to free interactions with the child. In an ideal scenario, the robot would be involved as a real partner to the ESDM activities with the child and the clinician.

The robot employed in the presented protocol is a small humanoid robot Nao. It has been chosen for his simplified aspect as well as for his ability of establishing a natural, emphatic communication with children (Ivaldi et al., 2014). However, a similar protocol could be proposed using other kind of robots, humanoids, androids, animal-like, or with more abstract shapes. Each robot will be able to accomplish the protocol in different way and induce a different degree of behaviour elicitation.

Presented results also support the existence of postural control impairments and atypicalities in the kinematics aspects of movements in ASD. Their exploration, requiring further studies, poses significant pragmatic challenges for researchers and clinicians alike. In this regard, computational modeling involving human-machine interaction may be promising.

Acknowledgements

Authors would like to thank all the MICHELANGELO Study group: Silvio Bonfiglio (FIMI, Italy), Fabio Apicella, Federico Sicca, Chiara Lucentini, Francesca Fulceri (Fondazione Stella Maris, Italy), Giovanni Pioggia, Alessandro Tonacci (CNR, Italy), Federico Cruciani, Cristiano Paggetti (I+, Italy), Angele Giuliano, Maryrose Francisa (Accros Limit, Malta), Koushik Maharatna, Valentina Bono, Wasifa Jamal (University of Southampton, UK), Leo Galway, Mark Donnelly (University of Ulster, UK), Anne-Lise Jouen, Elodie Tilmont (ISIR-UPMC). The authors would also like to thank all the patients and families who participated in the current study. They are also very grateful for the support from the hospitals staff. Authors would give many thanks to A. Arrigo, R. Grassia and G. Varni for their kind support and collaboration. The current study was supported by a grant from the European Commission (FP7: Michelangelo under grant agreement No. 288241), the fund “Entreprendre pour aider”, the Laboratory of Excellence SMART (ANR-11-LABX-65) supported by French State funds managed by the ANR within the Investissements d’Avenir programme under reference ANR-11-IDEX-0004-02. The funding agencies and the University were not involved in the study design, collection, analysis and interpretation of the data, the writing of the paper, or the decision to submit it for publication.

References

Anzalone, S.M., Boucenna, S., Cohen, D., Chetouani, M., 2014a. Autism assessment through a small humanoid robot. Proc. HRI: a bridge between Robotics and Neuroscience, Workshop of the 9th ACM/IEEE Int. Conf. on Human-robot interaction , 1–2.

Anzalone, S.M., Boucenna, S., Ivaldi, S., Chetouani, M., 2015. Evaluating the engagement with social robots. International Journal of Social Robotics , 1–14.

Anzalone, S.M., Chetouani, M., 2013. Tracking posture and head movements of impaired people during interactions with robots, in: New Trends in Image Analysis and Processing–ICIAP 2013. Springer Berlin Heidelberg, pp. 41–49.

Anzalone, S.M., Menegatti, E., Pagello, E., Yoshikawa, Y., Ishiguro, H., Chella, A., 2011. Audio-video people recognition system for an intelligent environment, in: Human System Interactions (HSI), 2011 4th International Conference on, IEEE. pp. 237–244.

Anzalone, S.M., Tilmont, E., Boucenna, S., Xavier, J., Jouen, A.L., Bodeau, N., Maharatna, K., Chetouani, M., Cohen, D., Group, M.S., et al., 2014b. 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, 814–826.

Betancur, C., 2011. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain research 1380, 42–77.

Billeci, L., Tartarisco, G., Brunori, E., Crifaci, G., Scardigli, S., Balocchi, R., Pioggia, G., Maestro, S., Morales, M.A., 2015. The role of wearable sensors and wireless technologies for the assessment of heart rate variability in anorexia nervosa. Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity 20, 23–31.

Bono, V., Narzisi, A., Jouen, A.L., Tilmont, E., Hommel, S., Jamal, W., Xavier, J., Billeci, L., Maharatna, K., Wald, M., et al., 2016. Goliah: A gaming platform for home-based intervention in autism–principles and design. Frontiers in Psychiatry 7.

Boucenna, S., Anzalone, S., Tilmont, E., Cohen, D., Chetouani, M., 2014a. Learning of social signatures through imitation game between a robot and a human partner. Autonomous Mental Development, IEEE Transactions on 6, 213–225.

Boucenna, S., Cohen, D., Meltzoff, A.N., Gaussier, P., Chetouani, M., 2016. Robots learn to recognize individuals from imitative encounters with people and avatars. Scientific reports 6.

Boucenna, S., Narzisi, A., Tilmont, E., Muratori, F., Pioggia, G., Cohen, D., Chetouani, M., 2014b. Interactive technologies for autistic children: a review. Cognitive Computation 6, 722–740.

Cester, I., Dunne, S., Riera, A., Ruffini, G., 2008. Enobio: Wearable, wireless, 4-channel electrophysiology recording system optimized for dry electrodes, in: Proceedings of the Health Conference, Valencia, Spain, Citeseer.

Chaminade, T., Fonseca, D.D., Rosset, D., Lutcher, E., Cheng, G., Deruelle, C., 2012. Fmri study of young adults with autism interacting with a humanoid robot, in: RO-MAN, 2012 IEEE, IEEE. pp. 380–385.

Cruciani, F., Donnelly, M.P., Nugent, C.D., Parente, G., Paggetti, C., Burns, W., 2010. Dante: A video based annotation tool for smart environments, in: Sensor Systems and Software. Springer, pp. 179–188.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., Donaldson, A., Varley, J., 2010. Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model. Pediatrics 125, e17–e23.

De Leva, P., 1996. Adjustments to zatsiorsky-seluyanov's segment inertia parameters. Journal of biomechanics 29, 1223–1230.

Diehl, J.J., Schmitt, L.M., Villano, M., Crowell, C.R., 2012. The clinical use of robots for individuals with autism spectrum disorders: A critical review. Research in autism spectrum disorders 6, 249–262.

Emery, N., 2000. The eyes have it: the neuroethology, function and evolution of social gaze. Neuroscience & Biobehavioral Reviews 24, 581–604.

Ester, M., Kriegel, H.P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise., in: Kdd, pp. 226–231.

Ghidoni, S., Anzalone, S.M., Munaro, M., Michieletto, S., Menegatti, E., 2014. A distributed perception infrastructure for robot assisted living. Robotics and Autonomous Systems 62, 1316–1328.

Godfrey, A., Conway, R., Meagher, D., ÓLaighin, G., 2008. Direct measurement of human movement by accelerometry. Medical engineering & physics 30, 1364–1386.

Guinchat, V., Chamak, B., Bonniau, B., Bodeau, N., Perisse, D., Cohen, D., Danion, A., 2012a. Very early signs of autism reported by parents include many concerns not specific to autism criteria. Research in Autism Spectrum Disorders 6, 589–601.

Guinchat, V., Thorsen, P., Laurent, C., Cans, C., Bodeau, N., Cohen, D., 2012b. Pre-, peri-and neonatal risk factors for autism. Acta obstetricia et gynecologica Scandinavica 91, 287–300.

Han, J., Shao, L., Xu, D., Shotton, J., 2013. Enhanced computer vision with microsoft kinect sensor: A review. Cybernetics, IEEE Transactions on 43, 1318–1334.

Ivaldi, S., Anzalone, S.M., Rousseau, W., Sigaud, O., Chetouani, M., 2014. Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement. Frontiers in neurorobotics 8.

Kaufman, A.S., Kaufman, N.L., 1983. Kaufman assessment battery for children. Wiley Online Library.

Kozima, H., Nakagawa, C., Yasuda, Y., 2005. Interactive robots for communication-care: A case-study in autism therapy, in: Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on, IEEE. pp. 341–346.

Kuczmarski, R.J., Ogden, C.L., Guo, S.S., Grummer-Strawn, L.M., Flegal, K.M., Mei, Z., Wei, R., Curtin, L.R., Roche, A.F., Johnson, C.L., 2002. 2000 cdc growth charts for the united states: methods and development. Vital and health statistics. Series 11, Data from the national health survey , 1–190.

Lim, Y.H., Partridge, K., Girdler, S., Morris, S.L., 2017. Standing postural control in individuals with autism spectrum disorder: Systematic review and meta-analysis. Journal of Autism and Developmental Disorders , 1–16.

Lord, C., Rutter, M., Le Couteur, A., 1994. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of autism and developmental disorders 24, 659–685.

Nadel, J., 2006. Does imitation matter to children with autism. Imitation and the social mind: Autism and typical development , 118–137.

Narzisi, A., Costanza, C., Umberto, B., Filippo, M., 2014. Non-pharmacological treatments in autism spectrum disorders: an overview on early interventions for pre-schoolers. Current clinical pharmacology 9, 17–26.

Ospina, M.B., Krebs Seida, J., Clark, B., Karkhaneh, M., Hartling, L., Tjosvold, L., Vandermeer, B., Smith, V., 2008. Behavioural and developmental interventions for autism spectrum disorder: a clinical systematic review. PloS one 3, e3755.

Ouss, L., Saint-Georges, C., Robel, L., Bodeau, N., Laznik, M.C., Crespin, G.C., Chetouani, M., Bursztejn, C., Golse, B., Nabbout, R., et al., 2014. Infant's engagement and emotion as predictors of autism or intellectual disability in west syndrome. European child & adolescent psychiatry 23, 143–149.

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., Pioggia, G., 2015. Autism and social robotics: A systematic review. Autism Research .

Reichow, B., Wolery, M., 2009. Comprehensive synthesis of early intensive behavioral interventions for young children with autism based on the ucla young autism project model. Journal of autism and developmental disorders 39, 23–41.

Rollins, J.D., Collins, J.S., Holden, K.R., 2010. United states head circumference growth reference charts: birth to 21 years. The Journal of pediatrics 156, 907–913.

Scassellati, B., Admoni, H., Mataric, M., 2012. Robots for use in autism research. Annual review of biomedical engineering 14, 275–294.

Schopler, E., Reichler, R.J., 1976. Psychoeducational profile. .

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

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

Tordjman, S., Somogyi, E., Coulon, N., Kermarrec, S., Cohen, D., Bronsard, G., Bonnot, O., Weismann-Arcache, C., Botbol, M., Lauth, B., et al., 2014. Gene× environment interactions in autism spectrum disorders: Role of epigenetic mechanisms. Frontiers in psychiatry 5.

Torres, E.B., Brincker, M., Isenhower, R.W., Yanovich, P., Stigler, K.A., Nurnberger, J.I., Metaxas, D.N., José, J.V., 2013. Autism: the micro-movement perspective. Frontiers in integrative neuroscience 7.

Toth, K., Munson, J., Meltzoff, A.N., Dawson, G., 2006. Early predictors of communication development in young children with autism spectrum disorder: Joint attention, imitation, and toy play. Journal of autism and developmental disorders 36, 993–1005.

Wechsler, D., 1949. Wechsler intelligence scale for children. .

Xavier, J., Bursztejn, C., Stiskin, M., Canitano, R., Cohen, D., 2015. Autism spectrum disorders: An historical synthesis and a multidimensional assessment toward a tailored therapeutic program. Research in Autism Spectrum Disorders 18, 21–33.

Xiong, X., De la Torre, F., 2013. Supervised descent method and its applications to face alignment, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE. pp. 532–539.