Research in Autism Spectrum Disorders 8 (2014) 814–826

Contents lists available at ScienceDirect

Research in Autism Spectrum Disorders

Journal homepage: http://ees.elsevier.com/RASD/default.asp

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

Salvatore Maria Anzalone [a,*], Elodie Tilmont [a,b], Sofiane Boucenna [a], Jean Xavier [b], Anne-Lise Jouen [a], Nicolas Bodeau [b], Koushik Maharatna [c], Mohamed Chetouani [a], David Cohen [a,b] the MICHELANGELO Study Group [1]

[a] Institute of Intelligent Systems and Robotics, University Pierre and Marie Curie, 75005 Paris, France
[b] Department of Child and Adolescent Psychiatry, APHP, Groupe Hospitalier Pitié-Salpêtrière et University Pierre and Marie Curie, 75013 Paris, France
[c] University of Southampton, Southampton, UK

ARTICLE INFO

Article history:
Received 17 November 2013
Received in revised form 28 February 2014
Accepted 8 March 2014

Keywords:
Autism spectrum disorder
Development
Social engagement
Joint attention
Social robots

ABSTRACT

We aimed to compare, during a joint attention (JA) elicitation task, how children with autism spectrum disorder (ASD) and children with typical development (TD) behave and explore their 4 dimensional (meaning spatial 3D + time) when interacting with a human or with a robotic agent.

We built a system that employed a Nao robot and a perception system based on a RGB-D sensor (Kinect) to capture social engagement cues. A JA induction experiment was performed in which children with ASD ($N = 16$) and matched TD children ($N = 16$) had a 3-min interaction with the robot or with a therapist. Nao induced JA by gazing; by gazing and pointing; and by gazing, pointing and vocalizing at pictures. Both groups of children performed well with the therapist. However, with Nao, both groups had lower JA scores, and the children with ASD had a significantly lower score than the TD children. We found that (i) multimodal JA induction was more efficient in both groups; (ii) the 3D spatial world gaze exploration showed less accuracy; and (iii) the trunk position in ASD showed less stability in the 4 dimensions compared to TD controls.

We conclude that, in ASD, JA skill depends on the interaction partner, and implies a higher motor and cognitive cost.

1. Introduction

The purposes of the work presented in this paper is: (1) to compute a robotic platform able to elicit joint attention (JA) during an interaction task; (2) to compare, during the JA elicitation task, how children with autism spectrum disorder (ASD) and children with typical development (TD) behave when interacting with a human or with a robotic agent; and (3) to assess

* Corresponding author at: Institut des Systemes Intelligents et de Robotique, Universite Pierre et Marie Curie, Pyramide – Tour 55, 4 Place Jussieu, 75005 Paris, France.
E-mail addresses: anzalone@isir.upmc.fr, Dcohen55@noos.fr (S.M. Anzalone).

[1] The members of the MICHELANGELO Study Group are listed in Appendix A.

http://dx.doi.org/10.1016/j.rasd.2014.03.002
1750-9467/

how children with ASD explored their 4 dimensional (meaning spatial 3D + time) environment compared to children with TD.

1.1. ASD and JA

ASD is a developmental syndrome that implies impaired social interaction, communication and language as well as stereotyped and/or restricted behaviors. Despite evidence that some symptoms of ASD are present early in life (Guinchat et al., 2012; Saint-Georges et al., 2011), autism diagnosis is generally made between 3 and 5 years of age (Saint-Georges et al., 2013; Cohen, 2012). Achieving efficient interaction between humans and autistic children is a difficult task for their families as well as for well-trained therapists (e.g. Saint-Georges et al., 2011; Cohen et al., 2013). Although ASD remains a devastating disorder with a poor outcome in adult life (Roux et al., 2013; Howlin, Moss, Savage, & Rutter, 2013), there have been important improvements in the condition with the development of various therapeutic approaches. The literature on interventions in ASD has become quite extensive, with increasing convergence between behavioral and developmental methods (Matson et al., 2012; Ospina et al., 2008). The focus of early intervention is directed toward the development of skills that are considered to be ''pivotal'', such as JA and imitation as well as communication, symbolic play, cognitive abilities, sharing emotions and regulation (Toth, Munson, Meltzoff, & Dawson, 2006) (e.g. the Early Start Denver Model; Rogers & Dawson, 2009). One of the main problems when interacting with children with ASD is their deficit of social interaction. While playing a game or conducting other activities with a social partner, these children tend to not concentrate on what they are actually doing, switching to other, repetitive, stereotypical behaviors that are of interest to them but that usually have no or few relations with the actual social context. In other words, children with ASD can display concerted attention to toys or objects that they like, but they have difficulties in sharing attention or interests with others (Rogers & Dawson, 2009). For example, maintaining eye contact with the caregiver is especially complicated (Maestro et al., 2005; Saint-Georges et al., 2010). Specifically, they lack JA, which is a key element of social cognition. JA teaches us much about social relationships, and it is a critical precursor of theory of mind (Premack & Woodruff, 1978) and language acquisition (Dominey & Dodane, 2004). Emery defined JA as a triadic interaction that showed that both agents focus on a single object (Emery, 2000). Agent 1 detects that the gaze of agent 2 is not directed at him/her and, therefore, follows the direction of the gaze to look at the ''object'' of attention of agent 2. This definition highlights a unidirectional process, unlike shared attention, which appears to be a coupling between mutual attention and JA. In shared attention, the attention of both agents concerns not only the object but the other agent as well (''I know that you are looking at the object, and you know that I am looking at the object''). Some authors (e.g., Tomasello, 1995) have argued that JA implies viewing the behavior of other agents as intentionally driven. In that sense, JA is much more than gaze following or simultaneous looking. By 12 months of age, TD infants display all aspects of JA (Carpenter, Nagell, & Tomasello, 1998).

In children with ASD, JA has been studied mainly by the annotation of video-recorded interaction: in a natural context (e.g., home movies, Saint-Georges et al., 2010) or in a laboratory context that uses JA induction during interactive play (e.g., early social communication scales, (Mundy et al., 2003). Children with ASD showed impairment in social orienting compared to children with intellectual disability (ID) and with TD (e.g., Dawson, Meltzoff, Osterling, Rinaldi, & Brown, 1998). They also showed impairment in sharing and proto-imperative JA (such as requesting) (Sigman, Mundy, Sherman, & Ungerer, 1986). Additionally, JA abilities at preschool ages predict language ability at the age of four years (Toth et al., 2006). In more recent years, social attention was explored using more sophisticated methods, including Information Communication Technology (ICT): 2-year-old toddlers with ASD showed the absence of preferential looking into the eyes of approaching adults, which predicted the level of social disability (Jones, Carr, & Klin, 2008), and there was a limited attention bias for faces (Chawarska, Volkmar, & Klin, 2010). Additionally, Klin, Lin, Gorrindo, Ramsay, and Jones (2009) showed that toddlers with ASD preferentially oriented visually to non-social contingencies rather than to biological motion.

1.2. ASD and robotics

ICT-based approaches and methods have been used for the therapy and special education of children with ASD. ICT research has explored several approaches for the treatment of persons with ASD, which are: (i) counteracting the impact of autistic sensory and cognitive impairments on daily life (assistive technologies, e.g., Crittendon, Murdock, & Ganz, 2013); (ii) trying to modify and improve the core deficit in social cognition (cognitive rehabilitation/remediation, e.g., Serret, 2012); and (iii) bypassing ASD impairments to help children acquire social and academic skills (special education, e.g., Lanyi & Tilinger, 2004). Nonetheless, much has yet to be improved to attain significant success in treating individuals with ASD. From a practical perspective, many of the existing technologies have limited capabilities in their performance, which limits the success of ICT treatment in persons with ASD. Clinically, most ICT proposals have not been validated outside the context of proof of concept studies (Boucenna et al., 2014a). Because most ICTs have limitations (e.g., the interaction is not natural, intuitive, or physical), emerging research in the field of autism is aimed at the integration of social robotics (Diehl, Schmitt, Villano, & Crowell, 2012; Kozima, Michalowski, & Nakagawa, 2009; Welch, Lahiri, Warren, & Sarkar, 2010). Social robots are used to communicate, display and recognize the ''emotion'' and develop social competencies and maintain social relationships (Fong, Nourbakhsh, & Dautenhahn, 2003).

In recent years an increasing number of studies have focused on the use of robots with individuals who have ASD (Diehl et al., 2012; Scassellati, Admoni, & Mataric, 2012). These studies involve the robots mainly in two roles of intervention:

practice and reinforcement (Duquette, Michaud, & Mercier, 2008). Some studies have attempted to evaluate the reaction of children involved in interaction with robots, according to robot-like characteristics (Pioggia et al., 2005; Pioggia et al., 2008; Feil-Seifer & Mataric, 2011a,b), and to emotional stimuli (Nadel, 2006; Nadel et al., 2006). Other studies have attempted to use robots for diagnostic purposes (Scassellati, 2007; Tapus, Mataric, & Scassellati, 2007) or as a tool to elicit behaviors (Feil-Seifer & Mataric, 2011a,b). In other cases, robots have been used as simplified tools that can facilitate social interactions (Duquette et al., 2008) and thereby teach or practice certain skills (Dautenhahn, 2003). Finally, robots have been also used as an agent that can provide feedback and encouragement (Dautenhahn, 2003; Picard, 2010) thus acting as a social mediator during activities between children with ASD and their partners.

1.3. Hypotheses

In this paper, we used a small humanoid robot, Aldebaran's Nao, in conjunction with an intelligent perception system, during a trial of experiments that involved both children with ASD and matched children with TD. In these experiments, the robot acted as an autonomous interactive partner, proposing to each participant a small set of activities that focus on stimulating JA. The perception system makes it possible to track and register the behaviors of the child, allowing their off-line analysis. To perform the experiments, we decided to use Aldebaran's Nao because it is a humanoid robot with a toy-like simplified shape, it has been used previously with children with ASD and it appeared to be attractive to them (Boucenna, Anzalone, Tilmont, Cohen, & Chetouani, 2014). Also, to explore JA as a cognitive activity per se, we decided to focus on older children with ASD assuming that they developed JA abilities through development and treatment. Exploring them would prevent comparison with TD children to be a consequence of poor JA abilities as show in young children with ASD. We had the following 3 main hypotheses: (H1) our system which employed a Nao robot to interact with children and a perception system based on a RGB-D sensor (Kinect) to capture social engagement cues should be able to elicit JA during interaction with a child. (H2) JA performance of older children with ASD should be close or similar to performance of TD children when interacting with a human partner and with Nao. (H3) Despite similar JA performance, we expected differences between children with ASD and children with TD in the way they explore their 4D (spatial 3D +time) environment during the JA induction task. More specifically, we expected less time spent on target and less movement smoothness during interaction.

2. Materials and methods

The aim of the system presented here was to provide a set of relevant measures about the social engagement capabilities of children. The collected information should help the understanding of how children with ASD explore their interactive world during JA activities with the Nao partner. A Nao robot was used in conjunction with a perception system based on a RGB-D sensor as a playground mate for children. During the interactions the system could track the position of the child and of his/her gaze and react accordingly. To capture the relevant information, a simple JA task was prepared: the robot attempted to induce the child to look toward some figures placed on the sides of the experiment room while the perception system recorded his/her posture and gaze. The data were analyzed off-line to extract relevant features. In particular, the system was capable of extracting the histogram of the child's gaze by using the cumulative distribution of the head poses and the displacement of his/her posture.

The system was implemented by developing a real-time system, which could collect the data during the experiment, and an off-line system, which could analyze the data and extract relevant information. While the off-line data processing was performed with the use of GNU Octave (see http://www.octave.org), the whole real-time system was implemented by means of the Robot Operative System (ROS), from Willow Garage (see Willow Garage Robot Operative System: http://www.ros.org), which is a convenient framework for robotics that easily allows the creation, management and coordination of distributed, highly decoupled software modules. Through this framework it was possible to easily develop several highly specialized software modules that could communicate via the network. The main parts of the real-time system are the robot control system and the perception system.

2.1. Robotic platform

The robot used in the presented system was Nao, which is a small humanoid robot made by Aldebaran. As shown in Fig. 1, it has 25 degrees of freedom distributed along the body: 4 joints for each arm; 2 for each hand; 5 for each leg; 2 for the head; and one to control the hip. Nao has several sensors to capture information about the environment, such as force sensors and acceleration sensors. The platform carries stereo microphones that can be used to recognize speech and to locate sound sources during direct interaction. Moreover, a set of colored light emitting diodes (LEDs) over the body are present to assure a certain degree of communication in a non-verbal way. Its software control system (NAOqi SDK) is available at Aldebaran and is bridged with ROS (Aldebaran Robotics NAOqi SDK: http://www.aldebaran-robotics.com).

The choice of this platform has been guided by Nao's capability of arousing empathy from children. The cute shape that resembles a child is easy to anthropomorphize. Moreover, human faces have very large informative content, such as attention, emotions, lip movements and other facial mimics that need to be deciphered and interpreted by the children; in the case of autistic children, this deciphering could be a very difficult task that could result in a deficit in their attention

Fig. 1. The robotic platform used in the present system: Nao from Aldebaran Robotics.

toward the caregiver. Thus, the simplified face of a humanoid robot could allow the children to easily interpret it and to pay more attention to its non-verbal communicative cues.

2.2. People perception

The perception system mainly used a RGB-D sensor that allowed capturing real-time images from the environment and their depth information. Human activity was detected by the use of the 3D information perceived by the RGB-D data by means of a multiple skeleton tracking system provided by OpenNi (Shotton et al., 2011). In particular, each person was tracked in the space in terms of their joint information. These data permitted the retrieving of posture, limb and arm movements and gesture information. Depth information acquired by the sensor was analyzed to distinguish the body of each person from the environment. This distinction was achieved through a background subtraction technique applied to the depth image. The depth image which remained was then segmented and classified according to a per-point approach, labeling each depth point of the body as a specific body part. As shown in Fig. 2, a total of 31 patches distributed around the body were considered. The body patch labeling uses depth invariants and 3D translation-invariant features that attempt to describe to which part of the body each depth pixel belongs. The classification process is based on a randomized decision forest of such features, which results in a dense probabilistic skeleton that has body parts appropriately labeled (http://www.codamotion.com/uploads/files). The training process of such a classifier employed a database of 500k labeled frames captured through motion capture in hundreds of different scenarios, such as dancing, kicking, and running. Each estimated patch is finally employed, according to its density, to extract the position of each joint of the body, which corresponds to each patch.

Once located, the system estimates the 3D position of the head of each human in the environment and projects it to the RGB image space, to select a section in that space in which the head of each subject should appear. A face tracker algorithm is then applied to estimate a model in terms of the pitch and yaw of each head. Two algorithms have been used: the Constrained Local Models (CLM) algorithm and the Generalized Adaptive View-based Appearance Model (GAVAM). Both of these models can be used with three-dimensional features, depth-map-based data. However, due to the high level of noise of the sensors used, the use of RGB images has been chosen. The CLM algorithm follows the Active Appearance Model (AAM) approach: both attempts to model faces via a statistical description that is based on a set of landmarks. The shapes of the faces are deformed iteratively according to the landmark positions in order to find a best fit with the actual face image. In particular, a standard AAM algorithm attempts to compute a model of faces according to their shape and their appearance. The shape model is computed from a set of key points spread over a face, such as its contour, the border of the lips, the nose and the eyes. In particular, a data set of labeled faces has been chosen as the training set of the algorithm. The shape model is obtained from the mean and the variance of the Principal Component Analysis (PCA) of the facial key points of all the faces in the training set. It will be described as the mean shape parameterized by the variance. The appearance model will be built by normalizing the grayscale image of the face, wrapping it over its mean shape. Additionally, in this case, the model was obtained from the PCA transformation of the wrapped faces in the training set, described by its mean and parameterized according to the variance. The shape model and the appearance model were fused to obtain a full face model using another PCA transformation: this action resulted in a parameterized model that could account for both the shape and appearance of the faces. During its normal usage, the AAM algorithm calculates the error between the current model and its actual appearance. The error controls the change in the model parameters iteratively in order to better approximate the actual appearance (error minimization). In particular, the error encodes how the parameters of the model should be changed: this relation is learned in the training step and is used iteratively, providing high-speed performance to the whole system.

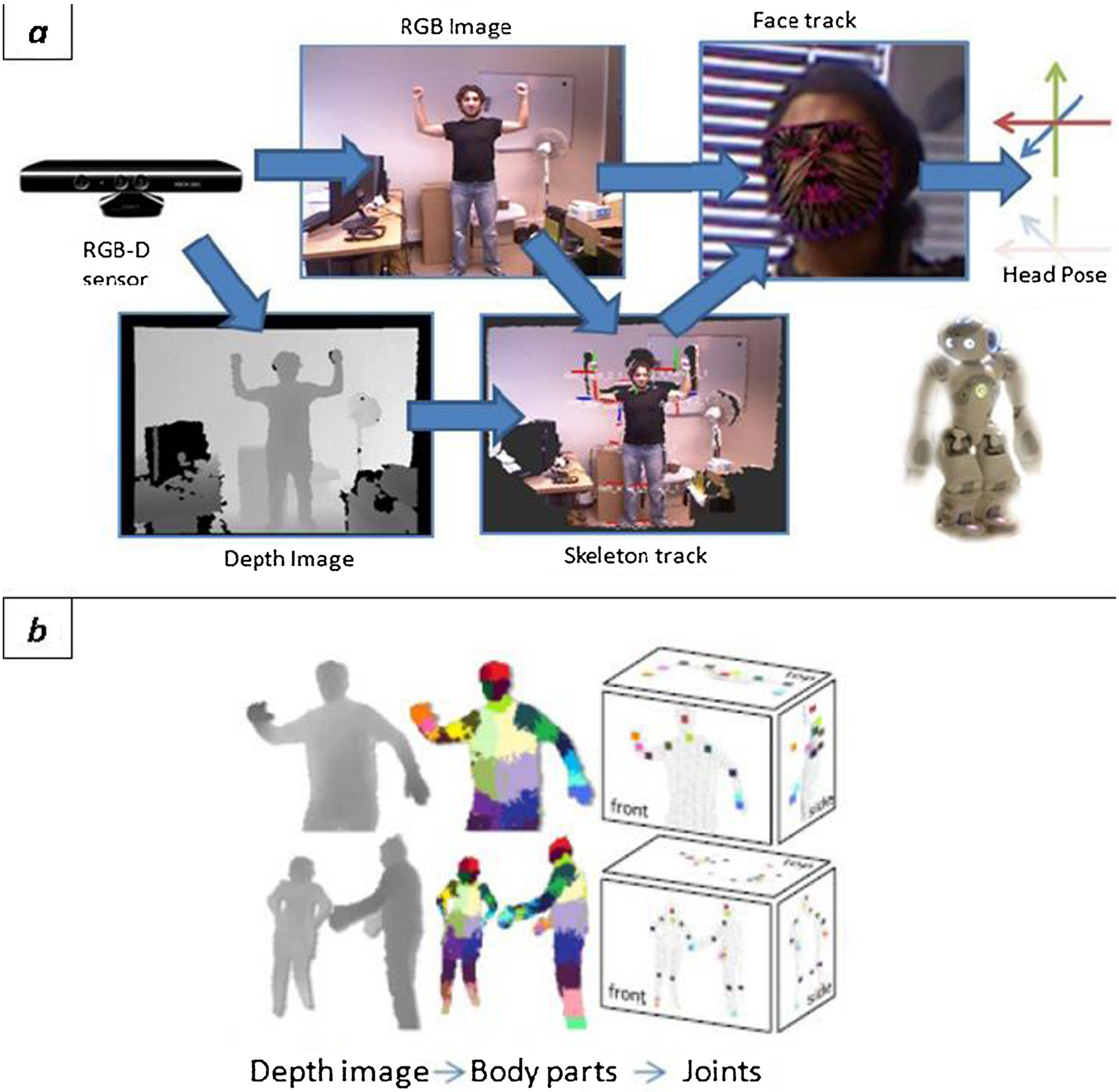

Fig. 2. People perception. (a) Gaze recognition pipeline: skeleton tracking, head detection and gaze estimation. (b) The process of extraction and tracking of people joints.
Source: Images from Shotton et al. (2011).

The CLM algorithm can be seen as a slight variation of the AAM algorithm: in this case, the face model is composed of a conjunction of a space model and an appearance model. However, in this case, the appearance model is built by employing local features: a patch of pixels around each key point is considered instead of using the whole wrapped face as in the AAM algorithm. Moreover, the model fitting is different: to adapt the parameters to the actual face, the Nelder-Meade simplex algorithm is employed.

The GAVAM algorithm tracker follows a different path: it attempts to integrate several state-of-the-art approaches to obtain a reliable head pose estimation. In particular, a static pose estimator has been fused with a derivative tracker and with a key point-based pose estimator. The static pose estimator can recognize faces in a single frame but neglects all of the useful temporal information. This issue is addressed by a derivative approach: the head position and orientation is tracked through the frame sequence. However, while this last system gives high precision in short time scales its accuracy diminishes significantly over the time. This approach is then integrated with a local key point approach that uses templates, in a similar way as AAM and CLM, to track the head over the time.

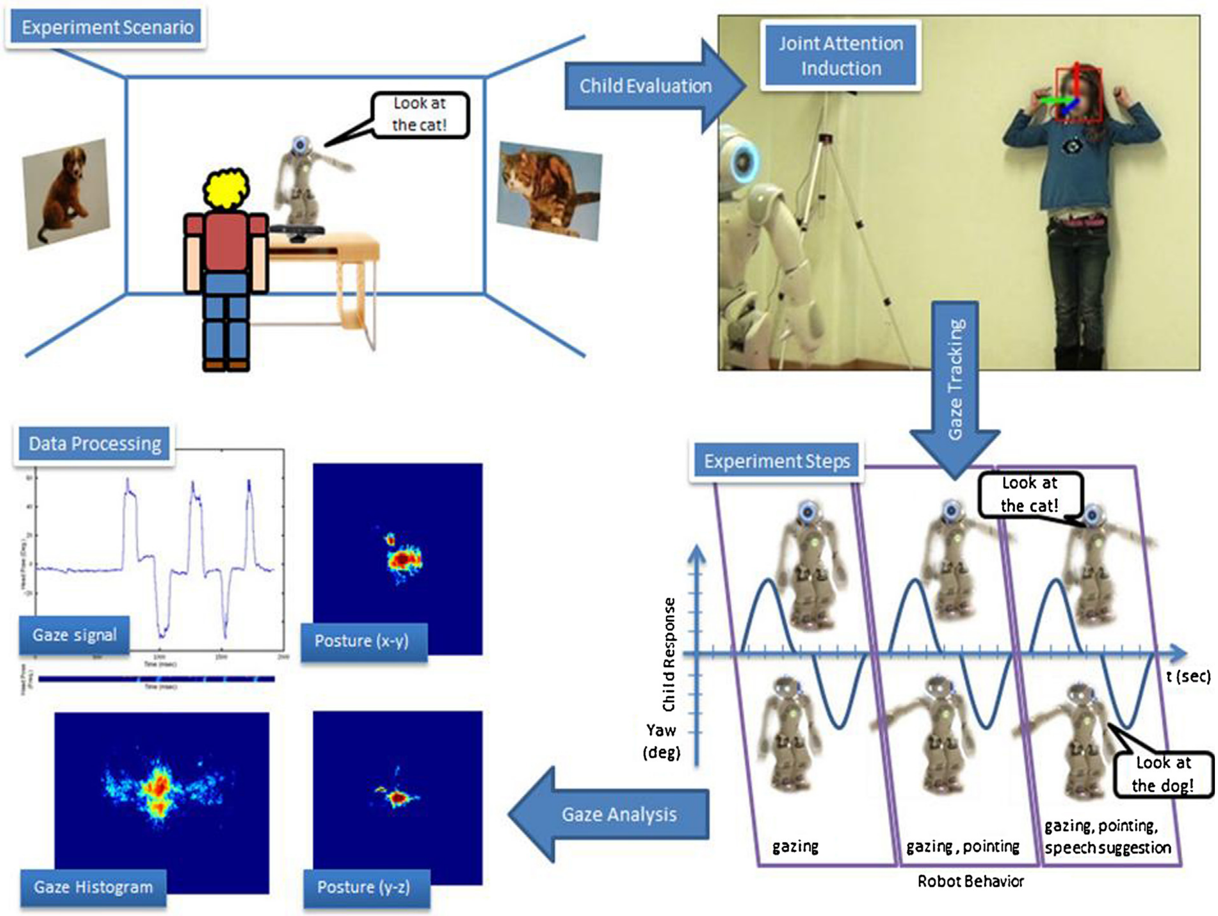

Fig. 3. The joint attention induction experiment: scheme of the experimental room, illustration of the data recorded and their off-line processing.

2.3. JA induction

While the system can be used in different scenarios, in order to capture information that can be relevant for understanding how children with ASD interact with a partner we defined a simple JA task within an experimental context: the robot placed in front of the child attempts to induce JA by alternatively gazing toward the child's face and toward some figures on the side of the room; at the same time, all of the perceptual cues were logged and stored for further analysis. A careful analysis of these records was conducted off-line.

In the JA induction experiment, an experimental room was arranged conveniently: as shown in Fig. 3, the robot and the RGB-D sensor were placed in front of the patient, in a suitable position, within a distance of 1.5/2.0 meters. In this way, the children could see the robot and its movements, while the sensor could perceive the behaviors of the child correctly. On the left and right side of the robot, two images, a cat and a dog, were placed and used as a focus of attention by the system.

During the experiment, each child was invited to stand in front of the system. The robot attempted to induce JA by, in a random order, alternatively gazing toward the child and then toward the image of the cat, on one side of the room; then, it gazed again alternatively toward the child and then toward the image of the dog, on the opposite side of the room. This process was repeated three times, each time adding more social information to the gazing, such as gestures and speech. According to the protocol, during a second phase, the robot combined a head movement with a pointing gesture toward the focus of attention: Nao randomly gazed and pointed to the cat and then to the dog. Finally, during the third stage, the robot combined a head movement, a pointing gesture and speech: Nao randomly gazed and pointed to the cat, while saying, "look at the cat"; then, Nao gazed and pointed at the dog, while saying, "look at the dog". The described experiment was implemented as a simple finite-state automaton that executes different actions according to the script described.

During the experiment, we collected and stored the information about the actual behavior of the robot in terms of the joint states, the information about the behavior of the child in terms of his/her body model (including the head's roll, pitch and yaw), and the RGB-D information.

2.4. Participants

The protocol was approved by the Pitié-Salpêtrière Hospital Ethics Committee. All of the parents received information on the experiment and gave written consent before the participation of their child. Thirty-two children participated in the study: 16 of the children were followed in the day-care setting for ASD of la Pitié-Salpétrière hospital. Those children suffered from various social impairments, including language disabilities and poor communicative skills. Sixteen children with TD were recruited from several schools in the Paris area. The controls met the following inclusion criteria: no verbal communication impairment, no ID, and no motor, sensory or neurological disorders. The controls were matched to the children with ASD with respect to their developmental ages and genders. For the control group, the developmental and chronological ages were considered to be the same. Children with ASD were assessed with the Autism Diagnostic Interview-Revised (ADI-R: Lord, Rutter & Le Couteur, 1994), to assess ASD symptoms, and the Gobal Assessment Functioning, to assess the current severity. The psychiatric assessments and parental interviews were conducted by three child psychiatrist/psychologists who specialized in autism (ET, JX, DC). The developmental age was assessed using a cognitive assessment. Depending on the children's abilities and ages, we used either the Wechsler Intelligence scales, the Kaufman-ABC or the Psycho-Educational Profile, third version (PEP-III). With both children with ASD and controls, we also performed a child-therapist interactive play session in which a JA task was incorporated to assess the ability of each child to interact using JA. The task was similar to that implemented for the Nao interaction, except that the child and therapist were seated at a table. It also included 3 types of induction. The child-therapist interactive play session always occurred before the experiment with Nao, which enabled it to offer some training to the child. All of the sessions were video-recorded for the annotation of JA using the ANVIL system. Each JA event was rated a 1 (success) or a 0 (failure). A total score was produced by a simple addition (maximum score = 6). An inter-rater reliability study was conducted on a subsample using 10 videos and 3 raters. The kappa was 0.98, which shows perfect agreement.

2.5. Statistical and computational analysis

Presentation of the data was based on the x, y, z axis position histograms for the head, pose and posture. To study the main direction (head yaw), we trained a K-means classifier on the raw data, and given the experimental procedure, we expected a three-class model (head gazing on the right to the cat, on the left to the dog and in front to Nao). The K-means classifier trained on TD children was applied to children with ASD to explore how their head yaw distributed across the three classes. Statistical analyses were performed with R statistical software. First, we compared JA annotation scores using the Exact Wilcoxon Mann Whitney rank sum test, given the variable distribution and the number of equal values. Second, we computed a comparison based on the variance of the yaw and pitch position of the head and the x, y, z axis position for the pose and posture. Depending on the distribution and type of each variable, we used either a multivariate regression or a Linear Mixed Model (LMM) with or without a Box-Cox transformation. For each dependent variable to be explored, the age and sex were always included in the models as possible covariates together with group memberships (ASD vs. TD). Third, to compare the K-means distribution across the classes, we used Fisher's exact test.

3. Results

3.1. Performance of children during the JA task

For the experiment, we recruited 16 children with ASD and 16 children with TD as the control group matched for developmental age and sex. Each subject performed the JA induction with the therapist during a play session first and then with the robot (see methods). Because of technical issues with recording, two recordings were not exploitable in the TD group. We therefore compared 16 children with ASD and 14 TD children. The characteristics and JA annotation scores of the participants are summarized in Table 1. With the therapist, the children with ASD performed as well as the TD children (Exact Wilcoxon Mann Whitney rank sum test: $p = 0.338$). With the Nao robotic platform, both groups of children showed lower performances that increased during the three phases of the tasks. During Nao interaction, we found that the children with ASD had a significantly lower JA score than the TD children (Exact Wilcoxon Mann Whitney rank sum test: $p = 0.001$). With Nao, the lower performances of children with ASD were particularly evident in the gazing and gazing + pointing conditions as opposed to the gazing + pointing + vocalizing condition. Regarding H1, we conclude that H1 is verified since our system was able to elicit JA during interaction. In contrast, the results only partly support H2: we did find that JA performances were similar in children with ASD and TD controls. However, performances with Nao were not similar when compared to interaction with a human partner and within groups of children. This indicates that Nao is less engaging than a human partner and even less for children with ASD, and it is particularly true for the most complex condition of the JA elicitation task (gazing only). Before assessing how the children behave in 4-D (space and time) movements and gazing during the JA task with Nao, we first present the perception performances of our system and the choice that we made to optimize it.

3.2. Perception system performances

A pre-evaluation of the performances of the perception system was first conducted to better understand and highlight its reliability and its limitations. Experiments were conducted to evaluate the performances of the perception system in real

Fig. 4. Data capture session to assess perception system performance: motion capture markers are placed on the helmet and the arms.

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

	ASD ($N = 16$)	Typical development ($N = 16$)
Age, mean (±SD), year	9.25 (±1.87)	8.06 (±2.49)
Male – Female	13–5	9–6
ADI-R, current, mean (±SD)		
Social impairment score	10.77 (±5.3)	Not administered
Verbal communication score	7.72 (±4.22)	Not administered
Non verbal communication score	4.3 (±3.5)	Not administered
Repetitive interest score	2.5 (±1.88)	Not administered
Developmental score	3.3 (±1.5)	Not administered
Total score	31.1 (±5.46)	
ADI-R, 4–5 years, mean (± SD)		
Social impairment score	17.33 (±8.47)	Not administered
Communication verb score	13.75 (±5.72)	Not administered
Communication non-verb score	8.08 (±4.4)	Not administered
Repetitive interest score	5.25 (±3.52)	Not administered
Developmental score	3.83 (±1.47)	Not administered
Total score	48.25 (±7.34)	
Developmental age[a]	7.47 (±2.9)	8.06 (±2.49)
IQ[a]	73 (±14)	All controls > 80
GAF score	40.27 (±9.44)	All controls > 90
Joint Attention Score/Therapist	4.67 (±1.63)	5.2 (±1.32)
Gazing	1 (±1)	1.4 (±0.9)
Gazing and pointing	1.73 (±0.59)	1.86 (±0.35)
Gazing, pointing and vocalizing	1.86 (±0.51)	1.86 (±0.5)
Joint Attention Score/Nao	1.47 (±1.19)	3.8 (±1.97)
Gazing	0.13 (±0.52)	1 (±0.85)
Gazing and pointing	0 (±0)	1 (±1)
Gazing, pointing and vocalizing	1.33 (±0.97)	1.8 (±0.56)

ASD = Autism Spectrum Disorder; SD = Standard Deviation; ADI-R = Autism Diagnostic Interview-Revised; GAF = Global Assessment Functioning.
[a] Assessed with the Vineland Developmental Score, the PsychoEducational Profile-Revised, the Kaufman Assessment Battery for Children or the Wechsler Intelligence Scale for Children.

human-robot interaction scenarios. A motion capture system composed of 3 CodaMotion CX1-800 units was chosen to obtain a good estimation of the posture of the human partners (Nadel et al., 2006). In particular, this system was used to track the position of the head of three subjects with three markers placed on the head, one over the left and one over the right ear and one on the forehead. Such information has been employed as ground truth to evaluate the presented system. Each human partner stood up in front of the robot at an approximately 1.5 m distance from the RGB-D sensor. The experimenters asked the experimental subjects, as in Fig. 4, to gaze on the left, on the right, and then upward and downward.

The experiment was repeated three times for each subject. As shown in Table 2, two types of tests that were related to the head's posture were conducted: the head's pitch and yaw were evaluated, correlating them to the ground truth, employing the GAVAM algorithm or the CLM algorithm. According to their overall performances, the two algorithms are similar, but accounting for the performances of each single recognition task, the pitch recognition and the yaw recognition, it is clear that the best performances could be achieved by using both of the algorithms together in a convenient way: GAVAM to recognize the head's pitch and CLM to recognize the head's yaw.

3.3. Children with ASD versus TD children exploring the 4D world during the JA task

During the experiment, the gaze and posture information of each child were recorded and then, were analyzed off-line. In particular, gazing information was collected using the GAVAM algorithm to retrieve the head's pitch and using the CLM

Table 2
Head pose estimation performances using the GAVAM algorithm and using the CLM algorithm.

Head pose	GAVAM	CLM
Pitch	61%	49%
Yaw	79%	93%
Overall	70%	71%

GAVAM = Generalized Adaptive View-based Appearance Model; CLM = Constrained Local Model.

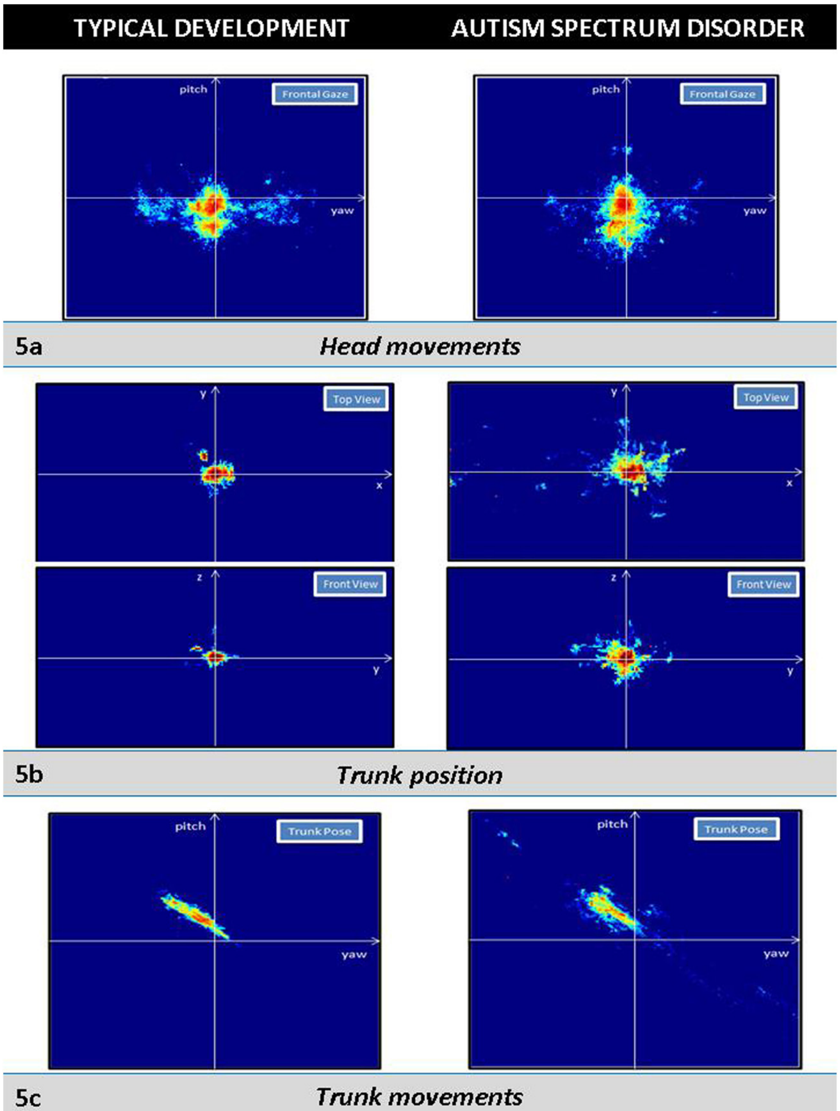

Fig. 5. The average histogram of movements of children with typical development and of children with autism spectrum disorder: (a) Head movements (yaw and pitch); (b) Trunk position (x, y and z); (c) Trunk movements (yaw and pitch).

algorithm to retrieve the head's yaw. The behavioral differences between the children with ASD and the TD children were explored by their yaw and pitch position histogram. Fig. 5A shows as a heat map the average bi-dimensional histogram of occurrences using the yaw and the pitch, for the TD population on the left and for the population affected by ASD on the right. Following the yaw, it can be seen how the gaze of the TD children focused on the left, on the right and on the center, depicting spots that correspond to the presence of the two focuses of attention, the animal figures on the two sides, and of the robot, which was placed exactly in front of the child. On the other hand, in children with ASD, the gazes were less concentrated on

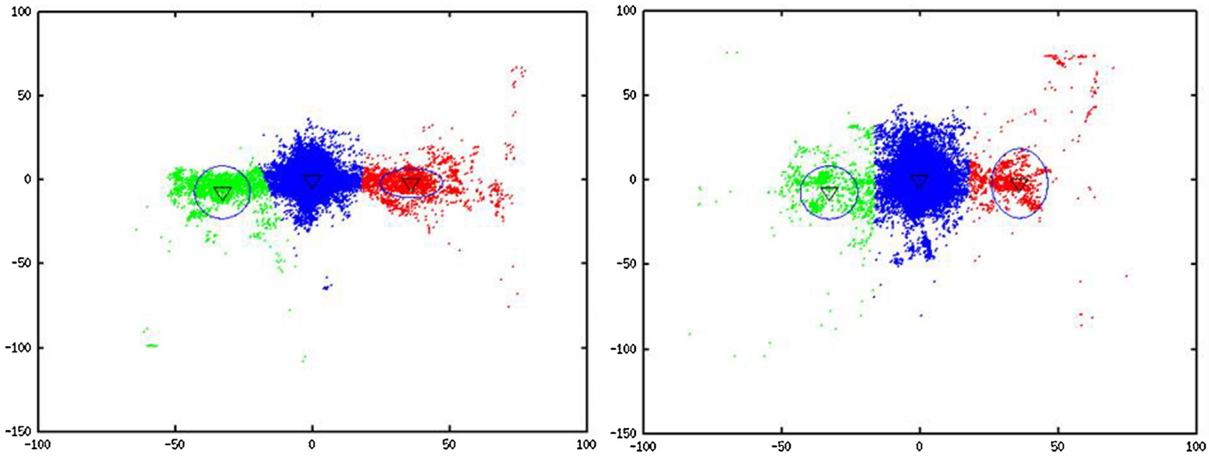

Fig. 6. K-means classification of head movements (yaw) in children with typical development and children with autism spectrum disorder.

the focuses. Using LMM to explore the yaw and pitch variance, we found that the yaw variance was significantly higher in the TD children than in the children with ASD ($\beta = 1.66$, $p = 0.002$), with no significant effect of age and sex. Additionally, there was a significant effect of JA induction modality with a higher variance in JA with vocalizing + pointing compared to JA with pointing ($\beta = 1.52$, $p < 0.001$) and compared to JA with gazing only ($\beta = 1.55$, $p < 0.001$). Pitch variance was significantly lower in TD children than in children with ASD ($\beta = -0.84$, $p = 0.019$) and in girls than in boys ($\beta = -0.89$, $p = 0.03$). There was no significant effect of age and JA modality.

The same analysis was performed for the pose and posture of the children. Fig. 5B shows the occurrences histogram of the displacements from the zero position for both the TD children (on the left) and the children with ASD (on the right). Two views are given, from the top and from the front. In this histogram, we also see differences that are similar to those reported for the gaze: the pose of the TD children resulted in a more stable circumstance among all of the axes compared to the pose of the children with ASD. Using multivariate regression, the pose variance was significantly lower in TD children than in children with ASD, within all axes (x, estimates = -28.1, $p = 0.001$; y, estimates = -7, $p = 0.006$; z, estimates = -12, $p = 0.009$).

A similar behavior was found for the posture of the children. Fig. 5C shows the heat map of the occurrences histogram for the pitch and the yaw angles of the trunk for both the TD children (left) and for the children with ASD (right). Here, the posture of the TD children is also more stable for all of the axes compared with the posture of the children with ASD. Using multivariate regression, posture variance was significantly lower in the TD children than in the children with ASD, within all axes (x, estimates = -13.9, $p = 0.0016$; y, estimates = -9.2, $p = 0.016$; z, estimates = -1.6, $p = 0.003$).

3.4. K-means classification and Nao perception of children with ASD versus TD children

To explore the regularities in behavior during the experiment (Fig. 6), we also ran a k-means algorithm to classify the head direction of the TD children. As expected, the classification distinguished three classes, which correspond to the two visual targets on the right (cat) and on the left (dog) and the interactive partner (Nao) in front. The left and right directions gathered 30.2% of all of the occurrences. In the children with ASD, the same classifier, when trained on the TD children, found that the left and right directions represented 8.72% of all of the occurrences (Fisher's exact test, $p = 2.2 \times 10^{-16}$). The TD children gazed on the right or on the left 4.6 times more frequently during the JA task than the children with ASD (95% Confidence Interval: 4.4–4.6).

4. Discussion

In this paper, we built a 4-D (spatial 3-D and time) interactive environment system to explore the responses of children during a JA task. This system was mainly composed of a small humanoid robot, called Nao, that could arouse empathy and interact with children, and a perception system that could track and model in real-time a child's body and head movements. The performances shown by the perception system assure that the information is extracted with a high level of reliability, especially under controlled conditions. During the experiment that was based on JA induction, the system could retrieve social engagement cues and capture significant data from both TD children and those with ASD. H1 is verified. Regarding our second hypothesis, we did find that JA performance of older children with ASD was similar to performance of TD children when interacting with a human partner. In spite of similar performances during the JA task with the therapist, interaction with Nao was not as easy, and children with ASD had a significant decrease in their JA score with Nao, which means that interaction also depends on the partner. Also, for both TD and ASD children, the JA task was easier when several social cues were given (gazing + pointing + vocalizing). H2 is only partially verified. This result is similar to that of a recent exploratory study that also used Nao to elicit JA (Bekele, Crittendon, Swanson, Sarkar, & Warren, 2013; Warren et al., 2000). Regarding

H3, a set of measures using the collected data has been presented showing significant differences between the two groups in terms of the patterns of behavior. During all JA induction conditions, gazing exploration showed less accuracy in children with ASD compared with the TD children. Finally, the trunk position and movement showed less stability both within the 3D spatial world and over time in children with ASD compared with the matched TD controls.

The current results trigger several comments. First, it has been shown that infants use prior experience with robot interactions as evidence that the robot is a psychological agent and, therefore, want to engage in JA with the robot if it changes its gaze direction by directing it to an object (Meltzoff, Brooks, Shon, & Rao, 2010). The current results suggest, however, that this response might be not as fluent as with a human partner (Bekele et al., 2013; Warren et al., 2000). Several hypotheses may be proposed to understand the meaning of this: (i) children with ASD exhibit many impairment in starting spontaneous (meaning without instruction) interaction (e.g., they show immutability, including hesitancy to try a new partner); (ii) some of the robotics platform characteristics may be involved (e.g., Nao cannot "really" gaze but only orient its head toward a target); (iii) the nature of the task itself (JA induction without explicit instruction) may be involved since in another imitation task with Nao we found that children with ASD performed as well with a human as with Nao (Boucena et al. in revision/b).

Second, the facilitation of the JA task by reinforcing the gaze with pointing and/or vocalizing is clinically not surprising (see intro). However, the relevance of this reinforcement was recently illustrated in an experiment on developmental robotics in which pointing gestures were exploited to rapidly and efficiently learn JA behavior (Doniec, Sun, & Scassellati, 2006). Third, an important point is to not misunderstand the meaning of what we described as less motor stability, which requires an understanding of the time scale. Indeed, the current motion analysis examined movement at the level of the video frame, i.e. 50 ms, considering that there were 15 images per second. None of the children with ASD had behavioral motor hyperactivity. However, we considered the current results to be meaningful in terms of the spatial and temporal micro-stability. We propose this "subtle instability" to be the consequence of an increased cognitive cost during the JA tasks. Fourth, one can argue that the differences evidenced between children with ASD and those with TD may be related to lower performance in children with ASD when interacting with Nao. However, we believe that the current findings have some value since the same 3D spatial world gaze exploration and trunk position characteristics were found during the gazing + pointing + vocalizing JA task despite similar performance between the two groups when interacting with Nao (see Table 2).

Regarding the system presented here, we think that there might be interesting clinical and research implications. Its real-time capabilities should help us to use the system as part of therapeutic activities that monitor in real-time the reaction of the child and inform the therapist of the JA events (Fujimoto et al., 2011). In a very similar JA exploratory study with Nao, Warren et al. (2000) showed that the 6 children with ASD included in the experiment improved in their ability to orient to prompts administered by the robotic system across a series of four sessions. Moreover, the robot itself can use the information that is perceived in order to adapt its own behaviors to the reaction of the child and act autonomously as a therapist or educational assistant, thus developing a profile on each child, recording preferences, the most and the least productive types of activities, to attempt to compose personalized therapeutic activities. Szafir and Mutlu (2012) proposed a fascinating design for an adaptive agent that could monitor EEG attention correlates and improve user engagement by employing behavioral techniques to regain attention. Although the system was tested only in TD adults, its potential in children with attention impairment has yet to be explored. Very recently, Jamal, Das, & Maharatna (2013) observed that, when using a wavelet transform, there exists a small unique set of phase synchronization topographies (in the range of 3–6) over the scalp from the time-varying EEG signal during a face-perception task in adults, each topography being stable on the order of milliseconds. During the execution of the task, these topographies, called Synchrostates, exhibit abrupt inter-state switching to construct a well-behaved time-sequence. Considering that the phase synchronization phenomenon is related to the functional connectivity of the underlying information processing network in the brain, this observation implies the existence of a small set of well-defined functional connectivity networks that repeat themselves in a well-defined temporal manner. More importantly, although the set of unique topographies is the same for normal and scrambled face perception tasks (both being in the general category of face-perception), the time-course or temporal stability that each of these topographies exhibit is markedly different in these two cases, which provides an indication toward possible stimulus-dependent stability of the functional connectivity network. This observation, together with Catarino et al. (2013) showing that interhemispheric wavelet transform coherence was reduced in people with ASD during visual perception and categorization, potentially opens up new directions of person-specific and stimulus-specific characterization of information processing capabilities in the brain, which could pave a new way of developing a person-centric therapeutic protocol for autistic children by characterizing their cognitive ability with objective measures that originate from neurophysiology. However, it is too early to draw a conclusion on this observation because a more rigorous study on a different population, in particular the autistic population, must be conducted for ascertaining the generalized nature of the existence of such synchrostates. Nevertheless, if proven unequivocally, such a phenomenon and its associated objective measures could well be utilized for developing therapy with the system proposed in this paper. In this case, EEG signals over the scalp of the subject could be captured during his/her interaction with the robot while performing a JA task. The captured EEG signal would undergo synchrostate analysis to evaluate the JA-specific cognitive characteristics of the subject which informs the development of appropriate therapy. In this case, the system not only acts as a tool for cognitive ability characterization but also acts as a facilitator for translating it into effective therapy.

The results should be interpreted in the context of several limitations. First, we attempted to depict a real use-case scenario of direct interaction between humans and robots. However, while in such a case the system was able to correctly track humans, in free interaction contexts, the gazing recognition performances became weaker, due to the rapid variation in the gazing of the people involved in the experiments and due to the low resolution of the camera. Moreover, because in both of these systems, the gaze recognition relies on RGB information, their performances are strongly influenced by the lighting conditions in the environment. Second, the performance of our system when following head movements that correspond to gazing was lower for pitch (61%) than for yaw (93%, see Table 2). However, it is likely that this lower performance had no effect on our results because the experiment that we built to induce JA was on the yaw axis given the position of the dog and cat pictures on the right and on the left of the participants. Third, due to the wide spectra of ASD pathologies, which involve a very large number of different signs and symptoms, we have chosen to refrain from the explicit classification of healthy and non-healthy children. Finally, because the results are validated on a small group of children, the authors cannot claim the character of generality of the results.

We conclude that, when children with ASD develop the JA skill, the ability depends on the interaction partner. However, it is better to use multimodalities to induce JA, which implies more motor and cognitive cost compared to TD children.

Conflict of interest

The authors declare that they have no competing interests regarding this manuscript.

Authors' Contributions

MC, DC, and the MICHELANGELO Study Group designed the study; SMA, EA, SB performed the experiment; ET, JX and DC assessed the children; SAM, SB, MC performed the computational analyses; NB, KM and DC performed the statistical analyses; and SMA and DC wrote a preliminary draft. All of the authors read, critically modified and approved the final manuscript.

Acknowledgments

The authors would like to thank all of the patients and families who participated in the current study. They are also very grateful for the support from the hospital staff, in particular Julie Brunelle, MD. The authors would like to thank to A. Carbone and T. Luiz for their kind 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 "*Fondation Serge Dassault*" and Valentina Bono to the group from University of Southampton. 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.

Appendix A. MICHELANGELO Study Group

Silvio Bonfiglio (FIMI, Italy), Fabio Apicella, Filippo Muratori, Antonio Narzisi, Federico Sicca (Fondazione Stella Maris, Italy), Lucia Billeci, Giovanni Pioggia (CNR, Italy), Federico Cruciani, Cristiano Paggetti (I+, Italy), Angele Giuliano, Maryrose Francisa (Accros Limit, Malta), Saptarshi Das, Wasifa Jamal (University of Southampton, UK), Leo Galway, Mark Donnelly (University of Ulster, UK).

References

Bekele, E., Crittendon, J. A., Swanson, A., Sarkar, N., & Warren, Z. E. (2013). Pilot clinical application of an adaptive robotic system for young children with autism. *Autism*. http://dx.doi.org/10.1177/1362361313479454

Boucenna, S., Narzisi, A., Tilmont, E., Muratori, F., Pioggia, G., Cohen, D., et al. Michelangelo Study Group. (2014). Information Communication Technology (ICT) and autism: Overview and Focus on Early Developmental Issues and Social Robotics. *Cognitive Computation*.

Boucenna, S., Anzalone, S., Tilmont, E., Cohen, D., & Chetouani, M. Michelangelo Study Group. (2014). Extraction of social signatures through imitation learning between a robot and a human partner. *IEEE Transactions on Autonomous Mental Development*.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*, 1–143.

Catarino, A., Andrade, A., Churches, O., Wagner, A. P., Baron-Cohen, S., & Ring, H. (2013). Task-related functional connectivity in autism spectrum conditions: An EEG study using wavelet transform coherence. *Molecular Autism, 4*, 1.

Chawarska, K., Volkmar, F., & Klein, A. (2010). Limited attentional bias for faces in toddlers with autism spectrum disorders. *Archives of General Psychiatry, 67*, 178–185.

Cohen, D. (2012). Controverses actuelles dans le champ de l'autisme. *Annales Médico-Psychologiques, 170*, 517–525.

Cohen, D., Cassel, R. S., Saint-Georges, C., Mahdhaoui, A., Laznik, M. C., Apicella, F., et al. (2013). Do motherese prosody and fathers' commitment facilitate social interaction in infants who will later develop autism? *PLoS ONE, 8*(5), e61402.

Crittendon, J., Murdock, L. C., & Ganz, J. (2013). Use of an ipadplay story to increase play dialogue of preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorder*.

Dautenhahn, K. (2003). Roles and functions of robots in human society: Implications from research in autism therapy. *Robotica, 21*, 443–452.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to naturally orient occurring social stimuli. *Journal of Autism and Developmental Disorder, 28*, 479–485.

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.
Dominey, P., & Dodane, C. (2004). Indeterminacy in language acquisition: The role of child directed speech and joint attention. *Jurnal of Neurolinguistics, 17*, 121–145.
Doniec, M. W., Sun, G., & Scassellati, B. (2006). Active learning of joint attention. *6th IEEE-RAS international conference on humanoid robots* (pp. 34–39).
Duquette, A., Michaud, F., & Mercier, H. (2008). Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots, 24*, 147–157.
Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience Biobehavioral Review, 24*, 581–604.
Feil-Seifer, D., & Mataric, M. J. (2011a). Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features. *2011 6th ACM/IEEE international conference on human-robot interaction (HRI)* (pp. 323–330).
Feil-Seifer, D., & Mataric, M. J. (2011b). Toward socially assistive robotics for augmenting interventions for children with ASD. *In Experimental Robotics Springer Berlin*.
Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems, 42*, 143–166.
Fujimoto, I., Matsumoto, T., Ravindra, P., De Silva, S., Kobayashi, M., & Higashi, M. (2011). Mimicking and evaluating human motion to improve the imitation skill of children with autism through a robot. *International Journal of Social Robotics, 3*, 349–357.
Guinchat, V., Chamak, B., Bonniau, B., Bodeau, N., Perisse, D., Cohen, D., et al. (2012). Very early signs of autism reported by parents include many concerns not specific to autism criteria. *Research in Autism Spectrum Disorders, 6*, 589–601.
Howlin, P., Moss, P., Savage, S., & Rutter, M. (2013). Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*, 572–581. http://www.codamotion.com/uploads/files/Codamotion%20C-X1%20Unit%20v2.pdf
Jamal, W., Das, S., & Maharatna, K. (2013). Existence of millisecond-order stable states in time-varying phase synchronization measure in EEG signals. *35th Annual international conference of the IEEE engineering in medicine and biology society (EMBC'13)* (pp. 2539–2542).
Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of General Psychiatry, 65*, 946–954.
Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature, 459*, 257–261.
Kozima, H., Michalowski, M. P., & Nakagawa, C. (2009). Keepon: A playful robot for research, therapy, and entertainment. *International Journal of Social Robotics, 1*, 3–18.
Lanyi, C. S., & Tilinger, A. (2004). Multimedia and virtual reality in the rehabilitation of autistic children. *Computers Helping People with Special Needs*, 625.
Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revision version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorder, 24*, 659–685.
Maestro, S., Muratori, F., Cavallaro, M. C., Pecini, C., Cesari, A., Paziente, A., et al. (2005). How young children treat objects and people: An empirical study of the first year of life in autism. *Child Psychiatry Human Development, 35*, 383–396.
Matson, J. L., Turygin, N. C., Beighley, J., Rieske, R., Tureck, K., & Matson, M. L. (2012). Applied behavior analysis in Autism Spectrum Disorders: Recent developments, strengths, and pitfalls. *Research in Autism Spectrum Disorders, 6*, 573–577.
Meltzoff, A. N., Brooks, R., Shon, A. P., & Rao, R. P. N. (2010). Social robots are psychological agents for infants: A test of gaze following. *Neural Networks, 23*, 966–972.
Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). *Early social communication scales (ESCS)*. Coral Gables, FL: University of Miammi.
Nadel, J. (2006). Does imitation matter to children with autism. *Imitation and the Social Mind*, 118–134.
Nadel, J., Simon, M., Canet, P., Soussignan, R., Blancard, P., Canamero, L., et al. (2006). Human responses to an expressive robot. *Proceedings of the sixth international workshop on epigenetic robotics*. Lund University.
Ospina, M. B., Seida, J. K., Clark, B., Karkhaneh, M., Hartling, L., Tjosvold, L., et al. (2008). Behavioral and developmental interventions for autism spectrum disorder: A clinical systematic review. *PLoS ONE, 3*, e3755.
Picard, R. W. (2010). Emotion research by the people, for the people. *Emotion Review, 2*, 250–254.
Pioggia, G., Igliozzi, R., Ferro, M., Ahluwalia, A., Muratori, F., & De Rossi, D. (2005). An android for enhancing social skills and emotion recognition in people with autism. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 13*, 507–515.
Pioggia, G., Igliozzi, R., Sica, M. L., Ferro, M., Muratori, F., Ahluwalia, A., et al. (2008). Exploring emotional and imitational android-based interactions in autistic spectrum disorders. *Journal of CyberTherapy & Rehabilitation, 1*, 49–61.
Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral & Brain Sciences, 1*, 515–526.
Rogers, S. J., & Dawson, G. (2009). *Early Sart Denver Model curriculum checklist for young children with autism*. Guilford Press.
Roux, A. M., Shattuck, P. T., Cooper, B. P., Anderson, K. A., Wagner, M., & Narendorf, S. C. (2013). Postsecondary employment experiences among young adults with an autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*, 931–939.
Saint-Georges, C., Cassel, R. S., Cohen, D., Chetouani, M., Laznik, M. S., Maestro, S., et al. (2010). What studies of family home movies can teach us about autistic infants. *Research in Autism Spectrum Disorders, 4*, 355–366.
Saint-Georges, C., Mahdhaoui, A., Chetouani, M., Cassel, R. S., Laznik, M. C., Apicella, F., et al. (2011). Do parents recognize autistic deviant behavior long before diagnosis? Taking into account interaction using computational methods. *PLoS ONE, 6*, e22393.
Saint-Georges, C., Guinchat, V., Chamak, B., Apicella, F., Muratori, F., & Cohen, D. (2013). Signes précoces d'autisme: D'où vient-on? Où va-t-on?. *Neuropsychiatrie de l'Enfance et de l'Adolescence, 61*, 400–408.
Scassellati, B. (2007). How social robots will help us to diagnose, treat, and understand autism. In *Robotics research* (pp. 552–563). Springer.
Scassellati, B., Admoni, H., & Mataric, M. J. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering, 14*, 275–294.
Serret, S. J. (2012). A serious game for autism spectrum disorders. *Neuropsychiatrie de l'Enfance et de l'Adolescence, 60*, 59.
Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., et al. (2011). Real-time human pose recognition in parts from single depth images. *Proceedings of the 2011 IEEE conference on computer vision and pattern recognition (CVPR'11)* (pp. 1297–).
Sigman, M. D., Mundy, P., Sherman, T., & Ungerer, J. (1986). Social interactions of autistic, mentally retarded and normal children and their caregivers. *Journal of Child Psychology and Psychiatry, 27*, 647–656.
Szafir, D., & Mutlu, B. (2012). Pay attention! Designing adaptive agents that monitor and improve user engagement *ACM:CHI*, 11–20.
Tapus, A., Mataric, M. J., & Scassellati, B. (2007). Socially assistive robotics. *IEEE Robotics and Automation Magazine, 14*, 35.
Tomasello, M. (1995). Joint attention as social cognition. *Joint Attention: Its Origins and Role in Development*, 103–130.
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 Disorder, 36*, 993–1005.
Warren, Z. E., Zheng, Z., Swanson, A. R., Bekele, E., Zhang, L., Crittendon, J. A., et al. (2000). Can robotic interaction improve joint attention skills? *Journal of Autism and Developmental Disorder*. http://dx.doi.org/10.1007/s10803-013-1918-4
Welch, K. C., Lahiri, U., Warren, Z., & Sarkar, N. (2010). An approach to the design of socially acceptable robots for children with autism spectrum disorders. *International Journal of Social Robotics, 2*, 391–403.