

Learning of social signatures through imitation game between a robot and a human partner

Sofiane Boucenna, Salvatore Anzalone, Elodie Tilmont, David Cohen, Mohamed Chetouani

Abstract—In this paper, a robot learns different postures by imitating several partners. We assessed the effect of the type of partners, i.e., adults, typically developing (TD) children and children with autism spectrum disorder (ASD), on robot learning during an imitation game. The experimental protocol was divided into two phases: (1) a learning phase, during which the robot produced a random posture and the partner imitated the robot; and (2) a phase in which the roles were reversed and the robot had to imitate the posture of the human partner. Robot learning was based on a sensory-motor architecture whereby neural networks (N.N.) enabled the robot to associate what it did with what it saw. Several metrics (i.e., annotation, the number of neurons needed to learn, and normalized mutual information) were used to show that the partners affected robot learning. The first result obtained was that learning was easier with adults than with both groups of children, indicating a developmental effect. Second, learning was more complex with children with ASD compared to both adults and TD children. Third, learning with the more complex partner first (i.e., children with ASD) enabled learning to be more easily generalized.

Index Terms—Human-Robot interaction, sensory-motor architecture, imitation, autism spectrum disorder

I. INTRODUCTION

The parent-infant interaction is a highly dynamic face-to-face interaction that requires the continuous mutual adaptation of behaviors [20], [47], [54]. Imitation is a specific dyadic behavior that plays an important role in infant development [17], [44], [45], [81] and core deficit in autism spectrum disorder (ASD) [23]. Imitation is a central issue in many disciplines, such as psychology, neurobiology and robotics. Several definitions of imitation have been proposed in the literature; however, no definition has been universally accepted [6], [63], [80], [82], [84]. In this study, we define imitation as the process by which a learner learns behavioral characteristics from a teacher (i.e., an interactive partner or model). We focus on learning new behaviors through an imitation game between a human and a robot (Fig. 1) with the specific aim of investigating the impact of the participant on robot learning.

Cognitive developmental robotics attempts to link developmental psychology with robotics. In this field, a new understanding of how higher human cognitive functions develop is sought using a synthetic approach [4]. The study of cognitive development can motivate the construction of autonomous robots. Robots can also be used as a tool to investigate cognitive models [51]. We believe that robotic behavior for rich interactions must be investigated from a developmental perspective to avoid the symbol grounding problem [40] (i.e., a human expert must provide knowledge to the system). In this study, we focus on learning new behaviors through the imitation interaction between a human and a robot (Fig. 1)

with the specific aim of investigating the impact of the participant on robot learning.



Fig. 1. Example of a typical human / robot interaction game (here, the human imitates the robot)

Several considerations motivated our experimental design. First, we chose three different interactive partners to maximize the complexity of interactive behavior: children with ASD, TD children, and healthy adults. Indeed, several studies from developmental psychology have shown that mental disorders in the mother (or the infant) can negatively affect the quality of the infant-mother interaction [55], [56], [73]. Second, in the context of learning, imitation reduces the search space of the learner and facilitates the participant-infant (or teacher-learner) interaction. Thus, in the field of robotics, imitation is often considered to be a powerful behavior that corresponds to the ability to learn by observation [14], [24], [25], [49], [76]. For example, a robot can combine known actions to reproduce a behavior that a participant performs. This concept of learning by observation is very close to "deferred imitation", which has been extensively described in developmental psychology [63], i.e., the ability of a child to reproduce an observed task but in a different context (e.g., spatial or temporal). Third, we chose a common "naive" system to learn imitating another person's posture during a natural interaction to study how learning "develops" by defining metrics that exhibit the influence of participants on robot learning. In this paper, we call Social Signature the pattern of interaction between a partner and the robot during the imitation task. The social signatures can be behavioral characteristics such as gestures, actions, sounds and/or words useful for communicating with the social environment. We show that robust and distinctive Social Signatures are captured during experiments with children with ASD, TD children, and healthy adults. These metrics characterized the quality of the imitation based on visual features

and reflected how the robot learned to recognize the postures of different participants (i.e., an adult, a TD child or a child with ASD). The naive system can be summarized as follows. Learning was performed without an explicit teaching signal that associated a specific posture with the robot's internal motor state. During the learning phase, the robot produced a random posture, and the participant imitated the robot: the robot associated what it did with what it saw via a sensory-motor architecture based on a neural network (N.N.). After this first phase, the roles were reversed, and the robot imitated the posture of the participant.

II. RELATED WORK

A. What is imitation?

Imitation is a central issue in many disciplines such as psychology, neurobiology and robotics. Several definitions of imitation exist. However, no definition is universally agreed upon: (1) Thorndike [80] offered a definition based on visual aspects: "learning to do an action by watching someone doing it". The visual aspect is highlighted in this definition. However, a full definition of imitation must consider multi-sensory aspects. (2) Wallon [82] defined imitation as a learning technique without reward (without reinforcement). (3) Whiten [84] defined imitation as the process by which the imitator learns some behavioral characteristics of the model. (4) Baldwin [6] proposed one of the first theories linking Darwinian evolution and imitation. If we consider imitation to be a mechanism of individual adaptation, it can allow individuals to improve some abilities that are not efficient enough at birth to ensure survival. In order to obtain a reliable definition of imitation, we should answer the following questions: imitate what, when, how and why. Imitate what? This question is probably the most important. Is imitation defined only by the reproduction of new actions by the imitator? Must we distinguish the new actions and the familiar actions? Moreover, actions or gestures can be imitated but different capabilities are required. Gestures involve only the body whereas actions involve both the body and the environment (e.g. object). Also, we can only imitate actions which are in our motor stock [15]. Why imitate? When a child is learning some social behaviors by imitating a social partner, this allows associating what he sees with what he does. Moreover, imitation can be seen as a way to communicate [57]. It establishes a social link and allows communicating before language acquisition. The aim is to interact with the partner. How to imitate? The quality of imitation must also be considered to define the imitation: correct or approximate, complet or incomplete, with or without corrections. This can inform us about the sensori-motor aspects. In addition, the correspondance problem must be adressed [59]. For example, an observer must identify a mapping between the demonstrator and the imitator in order to judge if the behavior is correctly imitated. However, this correspondance is easier if the demonstrator and the imitator have similar bodies. When to imitate? We must distinguish three time scales: immediate, shifted and differed. This involves several levels of memory and has different social consequences.

B. Robot imitation skills

Beginning with Kuniyoshi's studies [5], [49], learning by observation has been shown to proceed in three phases: (1) observation, which is watching an action performed by a human, e.g., a human grasps an object and then moves it to another position; (2) understanding, which involves the construction and memorization of an internal representation of the observed task; and (3) reproduction of the observed task. This approach has been used in several studies in different contexts, such as household environments [30], labyrinth [41] and learning sequences [7], [8]. In other studies, imitation has been used to reproduce an observed gesture (i.e., a low-level gesture).

Several research questions are thus motivated on movement recognition (can the robot identify the human arm and characterize the human arm trajectory?), the form of the gesture (what should the robot imitate?), and the perspective being considered. A solution to the the last issue might be to perform the gestures directly with a robotic forelimb, e.g., using a remote control [16] to manipulate the hand [14] or by fitting a robot model with sensors [1], [52] or an exoskeleton [43].

Imitation that involves interaction with the environment is more complex. The difficulty is in determining the relationships among the hands, arms and different objects. However, humans can aid the robot by specifying the relationships among the objects. The robot can also be endowed with primitive movements such as grasping an object. These primitive movements provide a vocabulary of actions for the robot, which the robot must then learn to combine to perform complex tasks [60], [61]. However, an important limitation is that the robot can only learn to perform tasks that require this primitive repertoire. Consequently, robotic systems with these designs have developed the following capabilities: (1) learning primitive movements such as grasping an object [16] or putting it in a box [42] and (2) performing gestures by adapting to the environment [16], [39], [77].

In other studies, a sensory motor architecture based on neural network has been designed to exhibit learning and communication capabilities via imitation [3], [2], [24]. An artificial system does not need to incorporate an internal model to perform real-time and low-level imitations of human movements, despite the related correspondence problem between humans and robots. These sensory motor architectures and this paradigm are useful because robots can learn online and autonomously, thus creating a real interaction between a human partner and a robot. In this case, the human partner communicates with the robot through imitation. In previous studies, we showed that imitation can be used as a communication tool in online facial expression learning [10], [11] and to bootstrap complex capabilities such as social referencing or joint attention [13], [12]. These experiments showed that a robotic head can autonomously develop facial expression recognition without being given a teaching signal. Our starting point was a computational model that showed that if an infant used sensory motor architecture to recognize a facial expression, its participants needed to imitate the infant's facial expression to enable online learning [36]. In

this present study, we used the same cognitive developmental robotics (CDR) paradigm. The infant-participant interaction is usually considered to be a relevant framework. An infant has a set of expressions that are linked with his/her own internal state, such as crying and a sad face when he/she needs food or a happy face after being fed. Meltzoff [53] showed that infants between 12 and 21 days of age can imitate both facial expressions and gestures. It is challenging to learn to recognize facial expressions without a teaching signal that enables what is perceived by the robot to be associated with its internal state (e.g., associating the vision of a "happy face" with an internal emotional state of happiness [37]). This issue has been investigated in robotic experiments [13]. In this present study, we show that as in the facial expression recognition problem, posture recognition can be autonomously learned using sensory-motor architecture.

C. Robotics and children with autism

As previously mentioned, we investigated the role of a partner on imitation learning using interactive sessions with different groups: adults, typically developing children and children with ASD. Robotics studies have been developed within the context of clinical studies [29], [75] since 2000. Diehl distinguished four different categories of studies [29]. The first category consists of studies on the behavior of individuals with ASD when they interact with robots based on external appearance, i.e., robot-like characteristics, human characteristics or non-robotic objects. For example, in [66], [67], a child with ASD was compared with a child typically developing control over his/her behavioral and physiological responses to a robotic face. Nadel [58], explored the responses of children to emotional expressions produced by a robot or a human actor. Interaction with a robot resulted in an overall increase in performance with age, and human expressions were recognized more easily. Studies [22], [70] have shown that some children with ASD prefer to interact with robots over non-robotic toys or human partners. The following studies [9], [65] found that adults with ASD exhibited a speed advantage when imitating robotic hand movements compared to human hand movements.

In the second category, robots are used to elicit behavior [74], [79]. For example, a robot can give a set of social cues designed to elicit social responses for which the presence, absence, or quality of the response aids diagnostic assessment. Stribling [78] showed that the interaction between a robot and a child could be used to elicit and analyze perseverative speech in an individual with high-functioning ASD. In other studies, the robot motivated desirable or prosocial behavior [21], [32]. In studies such as [34], [71], prosocial behavior, such as joint attention and imitation, was elicited. Ravindra [68] showed that individuals with ASD could follow social referencing behaviors performed by a robot. However, these studies should be considered exploratory because the sample sizes prevent generalization.

In the third category, robots are used to model, teach or practice a skill [21], [74]. The aim is to teach a skill that a child can imitate or learn and eventually transfer to

interactions with humans. For example, [31] explored whether a mobile robot toy could facilitate reciprocal social interaction. Fujimoto [35] used techniques for mimicking and evaluating human motions in real time using a therapeutic humanoid robot. These practical experiments were performed to test the interaction of ASD children with robots and to evaluate the improvement in children's imitation skills.

In the final category, robots are used to provide feedback and encouragement during a skill-learning intervention because individuals with ASD may prefer a robot teacher over a human one [21]. Duquette [31] used a robot behavior as a reward for learning. The robot provided positive reinforcement by raising its arms and saying, 'Happy'. Moreover, the robot could respond to internal stimuli from the child: for example, a pulse (or respiratory frequency) could be used to indicate the affective state or arousal level of the child to enhance the individualized nature of the treatment [64].

D. Our position

Here, we define imitation as: (1) the process by which a learner learns behavioral characteristics from a teacher (associate what he sees with what he does); (2) the actions imitated belong to the motor stock of the learner agent; (3) the teacher imitates the learner to communicate and establish a social link; (4) the quality of imitation is assessed thanks to metrics based on visual and neuronal features. Note that the correspondance problem is learned by our learner (a robot). During the learning phase, the learner performs postures and the teacher must imitate. The learner learns the correspondance between his own body and the partner's body. The same sensory-motor architecture has already been used in previous studies.

In the present study, we show the impact of the participants on the robot learning. This allows extracting social signatures, including those of children with ASD, through a human-robot learning paradigm based on imitation. Here, the robot was used as a tool in clinical investigations to better understand the interactions of children with ASD. Moreover, we highlight that as in the facial expression recognition problem, posture recognition can be autonomously learned using a sensory-motor architecture. This result shows the genericity of the model (learning of different tasks).

III. MATERIALS' & METHODS' OVERVIEW

A. Participants

The protocol was approved by the Pitié-Salpêtrière hospital ethics committee. All the parents received information on the experiment and gave written consent before the participation of their child. Fifteen children participated in the study (Table I). They were followed in the day-care setting for ASD of the Pitié-Salpêtrière hospital. Fifteen typically developing (TD) children were recruited from several schools in the Paris area. The controls met the following inclusion criteria: no verbal communication impairment, no intellectual disability, and no motor, sensory or neurological disorders. The controls were

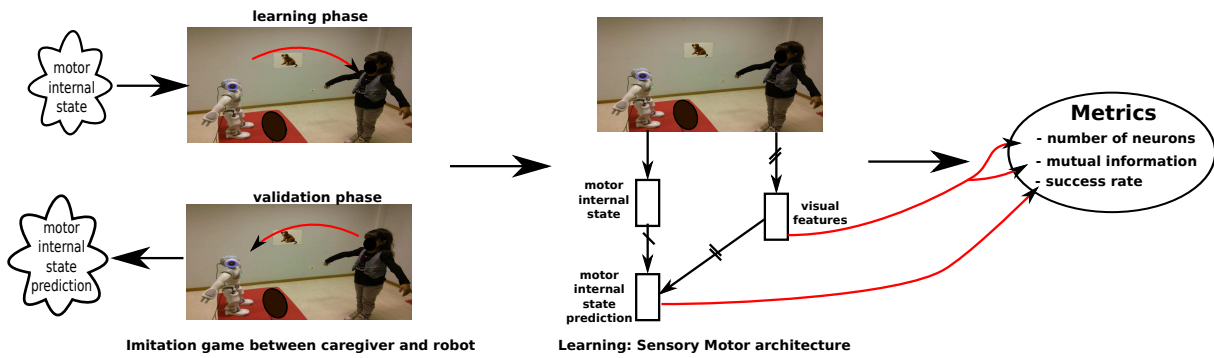


Fig. 2. Overview of experimental protocol: 1) imitation game between participant and robot to learn postures; 2) sensory-motor architecture based on a neural network; and 3) metrics to assess the impact of participants and to allow the extraction of social signatures.

	ASD (N=15)	TD (N=15)
Age, mean (\pm SD), year	9.25 (\pm 1.82)	8.06 (\pm 2.49)
Male - Female	13-5	9-6
ADI-R, current, mean (\pm SD)		
Social impairment score	10.77 (\pm 5.3)	Not relevant
Verbal communication score	7.72 (\pm 4.22)	Not relevant
Non verbal communication score	4.3 (\pm 3.5)	Not relevant
Repetitive interest score	2.5 (\pm 1.88)	Not relevant
Developmental score	3.3 (\pm 1.5)	Not relevant
Total score	31.1 (\pm 5.46)	
ADI-R, 4-5 years, mean (\pm SD)		
Social impairment score	17.33 (\pm 8.47)	Not relevant
Communication verb score	13.75 (\pm 5.72)	Not relevant
Communication non-verb score	8.08 (\pm 4.4)	Not relevant
Repetitive interest score	5.25 (\pm 3.52)	Not relevant
Developmental score	3.83 (\pm 1.47)	Not relevant
Total score	48.25 (\pm 7.34)	
Developmental age	7.47 (\pm 2.9)	8.06 (\pm 2.49)
IQ*	73 (\pm 14)	All controls > 80
GAF score	40.27 (\pm 9.44)	All controls > 90
Imitation score / therapist**	18.0 (\pm 3.46)	19.66 (\pm 1.29)
Imitation score / Nao**	17.27 (\pm 5.24)	19.53 (\pm 1.81)

TABLE I

SOCIO-DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF THE YOUNG PARTICIPANTS. ASD=AUTISM SPECTRUM DISORDER; TD=TYPICALLY DEVELOPING; SD=STANDARD DEVIATION. *ASSESSED WITH THE VINELAND DEVELOPMENTAL SCORE, THE PSYCHOEDUCATIONAL PROFILE-REVISED, THE KAUFMAN ASSESSMENT BATTERY FOR CHILDREN OR THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN. ** FROM MANUAL ANNOTATION OF THE VIDEO RECORDING OF INTERACTION.

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) [50] to assess ASD symptoms, and the Global Assessment Functioning to assess the current severity. The psychiatric assessments and parental interviews were conducted by three child psychiatrist/psychologists who are specialized in autism. The developmental age was assessed using a cognitive assessment. Depending on the childrens abilities and ages, we used either the Wechsler Intelligence scales, the Kaufman-ABC or the Psycho-Educational Profile, third version (PEP-III). Each participant has performed the experiment with the robot only one time.

B. Methods

In this study, we adopted a developmental approach whereby a robot learned through interaction with a human partner. Posture recognition was learned autonomously using a sensory motor architecture through an imitation game between the participant and the robot. Fig. 2 shows an overview of the study. Here, we aimed to investigate how a robot could learn to properly imitate a person’s posture during an interaction composed of two phases. During the learning phase, the robot produced a random posture, and the participant imitated the robot; then, the robot associated what it did with what it saw. The developmental model consisted of a sensory motor architecture based on the neural network developed in [11]. This architecture enables learning without explicit teaching signals that associate a specific posture with the robot’s internal motor state. To test our model, the following experimental protocol was adopted. In the first phase of the interaction (the learning phase), the robot produced a random posture (selected from 4 basic postures and a neutral posture; see Fig. 3) for 2 s. The participant was asked to imitate the robot. This first phase lasted between 1 and 2 min, after which the roles were reversed. The robot then had to imitate the posture of the participant, who led the imitation interaction.

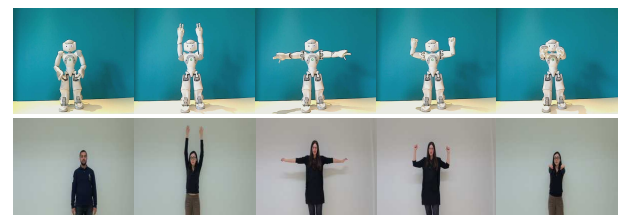


Fig. 3. The robot and the human produce the 4 basic postures and a neutral posture.

During the first phase, the robot learns the task, but also records all the images. Consequently, the database is created to perform offline processing. Each image was annotated with the response of the robot during the online learning. All the images are correctly labeled because the participant mimics the robot’s postures. The participants who interacted with the robot were composed of: 11 adults (corresponding to 2000 images), 15 TD children (corresponding to 3100 images) and 15 children

with ASD (corresponding to 3100 images). The scores are processed by checking the correct annotation. By applying, this methodology on the database, the statistics analysis were computed offline.

The final aim of our protocol was to investigate the impact of the participant on robot learning, which is called social signature. That is, we wanted to show that different learning results (e.g., the parameters of the learning model) were obtained when the interactions were performed with a child with ASD, a TD child or an adult. Metrics were specifically developed to evaluate the impact of the participants. The metrics were used to measure this impact and investigate how the robot learned the task from different participants.

IV. NEURAL NETWORK BASED ON A SENSORY-MOTOR ARCHITECTURE

A. Overview

In this section, we describe a sensory-motor architecture that enables the learning, recognition and imitation of postures. Visual processing enables the extraction of local views (Fig. 4); each local view is then learned by the VF (visual features) group, and the $MISP$ (motor internal state prediction) group learns the association between the visual features and the postures (Fig. 5). This architecture can solve the posture recognition problem if and only if we assume that the robot produces its first posture according to its internal state and that the human partner then imitates the robot's posture, thereby enabling the robot in return to associate these postures with its internal state.

Traditional HRI architectures assume person localization: pos-

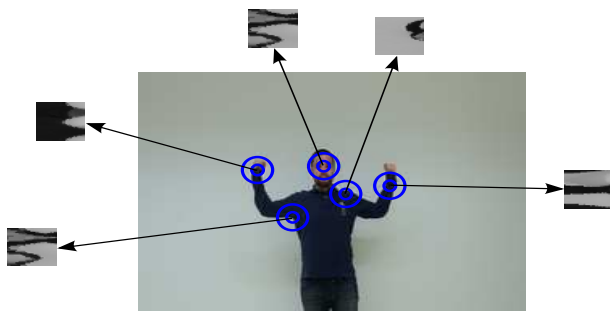


Fig. 4. Visual processing is shown above: this visual system was based on the sequential exploration of the image focus points. A gradient extraction was performed on the input image. A Difference of Gaussian (DOG) convolution provided the focus points. Finally, the local views (shown by arrows) were extracted from around each focus point.

ture recognition is then performed on the normalized image. In this case, the quality of the results is highly dependent on the accuracy of the person localization (i.e., the generalization capability of the neural network can be affected). From an autonomous learning perspective, it is important to avoid an ad hoc framing mechanism. Our solution consisted of eliminating the framing step and directly using all the most activated focus points in the image.

B. Focus points detection

The visual system was based on the sequential exploration of the image focus points. A gradient extraction was performed on the input image. A Difference Of Gaussian (DOG) convolution provided the focus points. Finally, the local views were extracted from around each focus point (Fig. 4). However, there was no constraint on how the local views were selected (i.e., no framing mechanism). This procedure can result in many distractors, such as objects in the background, as well as irrelevant parts of the human body.

C. Visual features

Fig. 5 shows the sensory-motor architecture that enabled the learning, recognition and imitation of postures. The extracted local view around each focus point was learned and recognized by a group of neurons VF (visual features) using a k-means variant that enabled online learning and real-time functions [46] called SAW (Self Adaptive Winner takes all)

$$VF_j = net_j \cdot H_{max(\gamma, \overline{net} + \sigma_{net})}(net_j) \quad (1)$$

$$net_j = 1 - \frac{1}{N} \sum_{i=1}^N |W_{ij} - I_i| \quad (2)$$

VF_j is the activity of neuron j in the group VF . $H_\theta(x)$ is the Heaviside function¹. Here, γ is a vigilance parameter (the threshold of recognition). When the prototype recognition is below γ , then a new neuron is recruited (incremental learning).

\overline{net} is the average of the output, and σ_{net} is the standard deviation. This model enables the recruitment to adapt to the dynamics of the input and to reduce the importance of the choice of γ . Thus, γ can be set to a low value to maintain a minimum recruitment rate. The learning rule allows both one-shot learning and long-term averaging. The modification of the weights (W_{ij}) is computed as follows:

$$\Delta W_{ij} = \delta_j^k (a_j(t) I_i + \epsilon (I_i - W_{ij}) (1 - VF_j)) \quad (3)$$

with $k = ArgMax(a_j)$, $a_j(t) = 1$ only when a new neuron is recruited; otherwise, $a_j(t) = 0$. Here, δ_j^k is the Kronecker symbol², where ϵ is the adaptation rate for performing long-term averaging of the stored prototypes. When a new neuron is recruited, the weights are modified to match the input (the term $a_j(t) I_i$). The other part of the learning rule, $\epsilon (I_i - W_{ij}) (1 - VF_j)$, averages the already learned prototypes (if the neuron was previously recruited). The closer the inputs are to the weights, the less the weights are modified. Conversely, the further the inputs are from the weights, the more the weights are averaged. The quality of the results depends on the ϵ value. If ϵ is chosen to be too small, it only has a small impact.

¹Heaviside function:

$$H_\theta(x) = \begin{cases} 1 & \text{if } \theta < x \\ 0 & \text{otherwise} \end{cases}$$

²Kronecker function:

$$\delta_j^k = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}$$

Conversely, if ε is too large, the previously learned prototypes can be unlearned. This learning rule enables the neurons in the VF group to learn to average the prototypes of postural features (such as an arm).

D. Classification

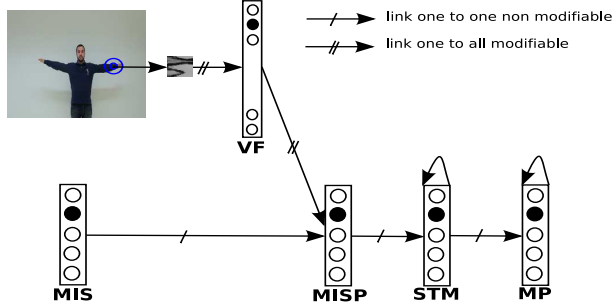


Fig. 5. Global architecture for the recognition and imitation of postures: visual processing enabled sequential local views to be extracted (a circle corresponds to a focus point); the VF group (local view recognition) learned the local views (each group of neurons, MIS , $MISP$, STM and MP , contained 5 neurons that corresponded to 4 postures and a neutral posture).

Distractors can be learned by VF . The figure/ground discrimination can be learned because the local views in the background are not statistically correlated with a given posture. In this interaction, distractors are present for all the postures, and their correlation with an internal motor state tends to zero. Only the local views for the person (Fig. 5) that are correlated with a given robot posture are reinforced. The Widrow and Hoff rule (which is derived from a least mean square (LMS) optimization) can be used to learn the image correctly if sufficient focus points can be found on the person during the period over which one image is explored (Fig. 4). In our network, internal state prediction ISP associates the activity of the visual features VF with the current motor internal state MIS of the robot (a simple conditioning mechanism which uses the least mean square (LMS) rule [85]). The modification of the weights (w_{ij}) is computed as follows:

$$\Delta w_{ij} = \epsilon \cdot VF_i \cdot (MIS_j - MISP_j) \quad (4)$$

The short term memory (STM) is used to sum and filter over a short period (N iterations), and the motor states $MISP_i(t)$ associated with each explored local view are given as follows:

$$STM_i(t+1) = \frac{1}{N} \cdot MISP_i(t+1) + \frac{N-1}{N} \cdot STM_i(t) \quad (5)$$

Here, i is the index of the neurons: for example, $MISP_i$ corresponds to the i^{th} motor state ($0 < i \leq 5$).

A limited amount of time is arbitrarily fixed for the visual exploration of one image to obtain a global frequency of approximately 10 Hz (i.e., 100 ms per image). The system can analyze up to 10 local views for each image (the system usually succeeds in taking 3 to 4 relevant points for a person). This small number of views is sufficient to learn the different postures and enable real-time interactions.

The robot posture is controlled via the MP group. The highest activity of the MP_i triggers the i^{th} posture because of a WTA. To increase the robustness of the system, MP also uses short-term memory (thereby attributing more importance to the present than to the past) with $\beta = 1$ and $\alpha < 1$ ($\alpha = 0.8$):

$$MP_i(t+1) = \beta \cdot STM_i(t+1) + \alpha \cdot MP_i(t) \quad (6)$$

However, online learning can involve specific problems. For example, human reaction time to robot postures is not immediate. This time lag can greatly disturb the learning process. If the first images that the robot learns are associated with the human's previous posture, the previous posture is unlearned. A given posture must be presented for a sufficiently long time that the first images have expired. Moreover, this procedure ensures high performance when the correct action is selected more than 50% of the time during the chosen time window. This feature is important for online interactions and can reduce the constraints on the intrinsic recognition quality if the frame rate is sufficiently high.

E. Metrics

For the current analyses, we used four different metrics to assess the impact of different participant groups. The metrics characterized the quality of the imitation and compared the performances of the different groups to extract social signatures. Each participant group was distinguished by its social signature. The metrics emerged from the sensory-motor architecture and were based on visual features, neural networks and clustering, showing that a robot learned and recognized the postures differently for different participant groups.

First, we assessed the behavioral performances by manual annotation and studying the robot success rate for each posture for the different participants during the interaction. For both the children with ASD and the controls, a child-therapist interactive play session was also conducted incorporating an imitation task to assess the ability of each child to interact using imitation. The task was similar to that implemented for the Nao interaction. The task also included 5 postures that the therapist randomly chose for the child to imitate. The child-therapist interactive play session always occurred before the experiment with Nao so that the child received some prior training. All the sessions were video-recorded to annotate imitation using the ANVIL system. Each imitation event was rated a 2 (success), a 1 (intermediate) or a 0 (failure). A total score was produced by simple addition (where the maximum score=10x2=20). An inter-rater reliability study was conducted on a subsample using 10 videos and 3 raters. The kappa value was 0.95, showing perfect agreement between raters.

Second, a metric emerged from the group of neurons VF that recruited new neurons according to the complexity of the visual input. Here, the visual input corresponded to the participant posture. The VF group learned the local views around each focus point. A vigilance mechanism (i.e., γ , the threshold of recognition) was used to recruit a new neuron when the local view had never been learned. More specifically, a new neuron was recruited if the prototype recognition was

below γ . Consequently, this metric was a robust means of evaluating the complexity of the visual input. In our protocol, if a VF recruited many neurons, the visual input of the robot was too complex to analyze. This second metric was the number of neurons needed to learn.

Third, another metric was provided by the clustering of the VF group of neurons. The VF^a was composed of n patterns in k_a clusters. The entropy of this clustering can be expressed as follows:

$$H(VF^a) = - \sum_{i=1}^{k_a} \frac{n_i^a}{n} \log \left(\frac{n_i^a}{n} \right) \quad (7)$$

where n_i^a represents the number of patterns in cluster $C_i^a \in VF^a$.

Here, the normalized mutual information $NMI(VF^a, VF^b)$ was used to measure the similarity between the two partitions VF^a and VF^b :

$$NMI(VF^a, VF^b) = \frac{-2 \sum_{i=1}^{k_a} \sum_{j=1}^{k_b} n_{ij}^{ab} \log \left(\frac{n_{ij}^{ab} \cdot n}{n_i^a n_j^b} \right)}{\sum_{i=1}^{k_a} n_i^a \log \left(\frac{n_i^a}{n} \right) + \sum_{j=1}^{k_b} n_j^b \log \left(\frac{n_j^b}{n} \right)} \quad (8)$$

Note that $0 \leq NMI(VF^a, VF^b) \leq 1$.

The number of shared patterns between the clusters $C_i^a \in VF^a$ and $C_j^b \in VF^b$ is denoted by n_{ij}^{ab} . Two clusters (or partitions) were compared using NMI : for example, one partition corresponded to TD children, and the other partition corresponded to children with ASD. In this case, the NMI measured the similarity between the two groups of children.

Finally, the success rate for each of the postures was used as a metric to compare the different groups. During the learning phase, the participants imitated the robot. During this period, the robot analysed the images, learned the task and recorded all the images (the database is created to perform offline processing). Each image was annotated with the response of the robot during the online learning. The statistical analyses are performed offline where the correct correspondence was verified.

V. RESULTS

A. Behavioral performances

The characteristics of the young participants are given in Table I. Matching for developmental age and sex was adequate. The performance scores from the manual annotation showed that the children from the two groups performed the imitation task well whether the task was conducted with a human partner (i.e., the therapist) or Nao (i.e., all the scores were close to the maximum performance value of 20). However, the children with ASD performed significantly lower than the TD children whether they interacted with the therapist (mean difference=1.7, $p=0.027$) or with Nao (mean difference=2.2, $p=0.048$). The robot learning performance, as assessed by the sensory motor architecture, is described below.

After learning, the associations between the visual features VF and the motor internal state prediction $MISP$ are strong enough to bypass the low-level reflex activity that comes from the motor internal state MIS . In this case, posture recognition

results from the temporal integration of the $MISP$ associated with the different visual features analyzed by the system. Features have a postural value if they are correlated with the robot posture, i.e., the posture features of the human. Fig. 6 shows the temporal activity of the neurons and the posture recognition during a natural interaction: the participant performed the postures, and the robot then imitated the recognized posture. Fig. 7 shows that the interaction between the robot and

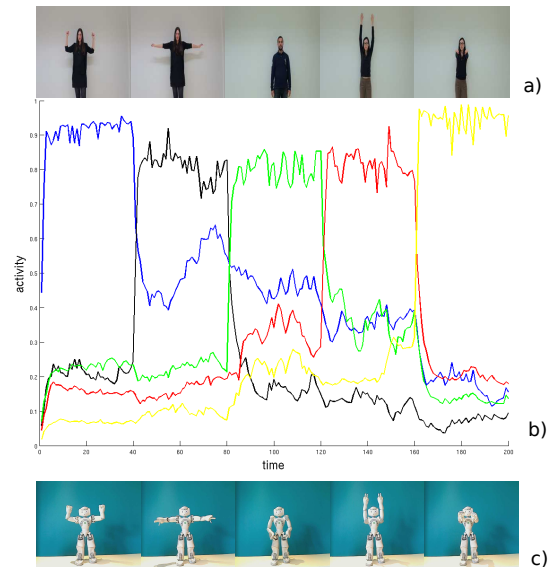


Fig. 6. Results of posture recognition in the context of imitation: a) participant performed a posture; b) temporal activity of the neurons associated with the triggering of different postures when the robot imitated the human (after learning); and c) the robot imitated the participant's postures

the human over a period of 2 min was sufficient for the robot to learn the posture through the imitation game between the human and robot. The results show the success rate for each posture depending on the different participants that interacted with the robot: 11 adults, 15 TD children and 15 children with ASD (three cases were not available for off-line analysis because of technical difficulties during the video registration). Overall, the success rates show the impact of the participant on robot learning. The success rate was 84% when the robot interacted with adults, 69% when the robot interacted with the TD children and 61% when the robot interacted with the children with ASD. The results show that the incremental learning was robust even when the number of participants increased and that the postural positions differed from one person to the next. Performance recognition was also better with adults ($p < 0.05$). Moreover, the measures of 'success' rate show that some postures are not well recognized: the posture (4) is the more complex to recognize while the posture (1) is more easily recognized. However, the aim was not to focus on the performance of the system, per se, but to show the effect of the human partner on the robot learning. In other words, we compared the success rates of each partner group to analyze and understand their impact on the robot learning.

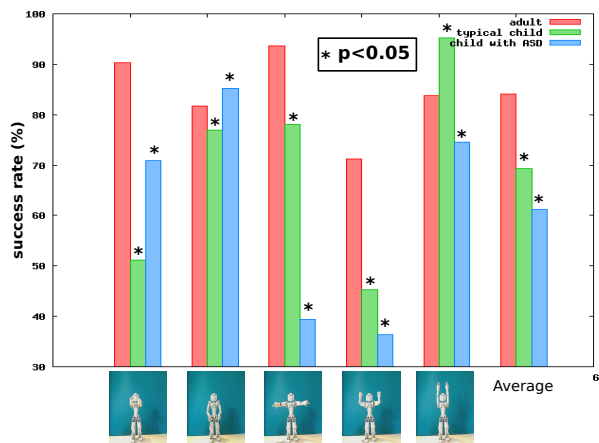


Fig. 7. The success rate is shown for each posture for different participants (adults, TD children, and children with ASD) obtained during natural interaction with the robot. During the learning phase (which only lasted 2 minutes), humans imitated the robot, and then the robot imitated them. Different participants interacted with the robot: 11 adults (corresponding to 2000 images), 15 TD children (corresponding to 3100 images) and 15 children with ASD (corresponding to 3100 images). Each image was annotated with the response of the robot during the online interaction, enabling the statistical analyses to be performed offline.

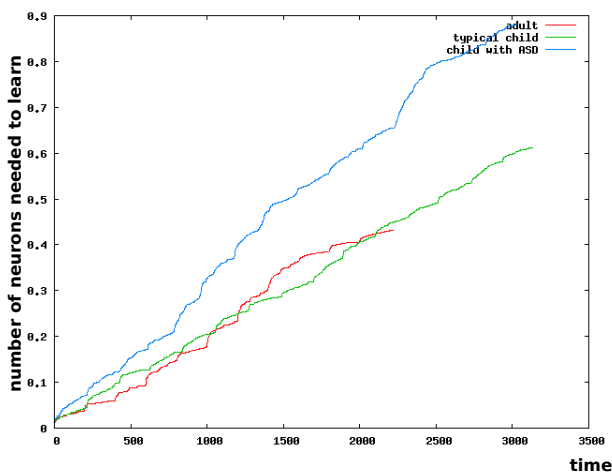


Fig. 8. The number of neurons needed to learn with different participants (adults, TD children, and children with ASD) interacting with the robot: 11 adults (corresponding to 2000 images), 15 TD children (corresponding to 3100 images) and 15 children with ASD (corresponding to 3100 images)

B. Learning according to the number of recruited neurons in adults, TD children and children with ASD

As explained above, we used a new metric to assess the impact of participants on robot learning. Fig. 8 shows the number of neurons needed to learn with different participants. To compare the three groups composed of adults, TD children and children with ASD, we used the Chow test [18] and a significance threshold of $p < 0.05$. The results show that the *VF* group recruited more neurons when the robot interacted with children with ASD than when the robot interacted with adults or a TD child ($p < 0.05$). Consequently, the robot appeared to recognize the postures of adults and TD children more easily than those of children with ASD.

This result can be attributed to the higher complexity of the

visual input (i.e., the participant posture) when children with ASD imitated the robot. Despite successful performance during the imitation task, as given by the annotation score, more variability was observed in the postures of the children with ASD. Fig. 7-8 confirm the impact of the participants on the robot learning and show the different developmental trajectories of the robot. The number of neurons required by the robot and the postures recognition varied with the characteristics of the participants. Consequently, the robot saw features that the therapist did not see. All these results show that the robot precisely analyzed the human's postures unlike the therapist annotating the video recording (subjective criterion, annotation time). The robot had difficulties to learn from a demonstrator who has more movement variability. However, the result is interesting because the robot discovers this complexity without the intervention of the human engineer. This property emerges from the sensory-motor architecture based on neural network, through which the robot can provide a sophisticated metric.

Finally, we also assessed learning using normalized mutual information for paired participants. Table II shows the normalized mutual information used to measure the agreement between two clusters belonging to two groups. The NMI measured the similarity between the two groups. The $NMI_{TDvs.Adult}$ between the TD children and the adults was 0.57. The $NMI_{ASDvs.Adult}$ between the children with ASD and the adults was 0.6. The $NMI_{TDvs.ASD}$ between the TD children and the children with ASD was 0.62. The results reinforce previous results (Fig. 7), enabling us to draw the following conclusions: (1) age affects learning because the NMI for both groups of children differed substantially from that for the adult group, though the same number of neurons were needed to learn with the TD children as for the adults; and (2) a pathological effect also contributes to learning because the NMI between the TD children and the children with ASD was far from 1.

	TD/ASD	TD/adult	ASD/adult
NMI	0.622	0.570	0.603

TABLE II
NMI (NORMALIZED MUTUAL INFORMATION), WHICH MEASURES THE AGREEMENT BETWEEN TWO CLUSTERS BELONGING TO TWO POPULATIONS AND SHOWS THE SIMILARITY BETWEEN THE TWO POPULATIONS

C. Does first partnership (ASD vs. TD) during the learning phase influence robot learning?

Two cases were tested to assess this issue. In the first case, the robot interacted and learned first with the TD children, followed by interacting and learning with the children with ASD. In the second case, the robot interacted and learned first with the children with ASD, followed by interacting and learning with the TD children. Fig. 9 shows the number of neurons needed to learn during the learning phase. In both cases, the curve varies significantly when the robot started to interact with the second group ($p < 0.05$). The results show that the same final solution was obtained for both cases but that

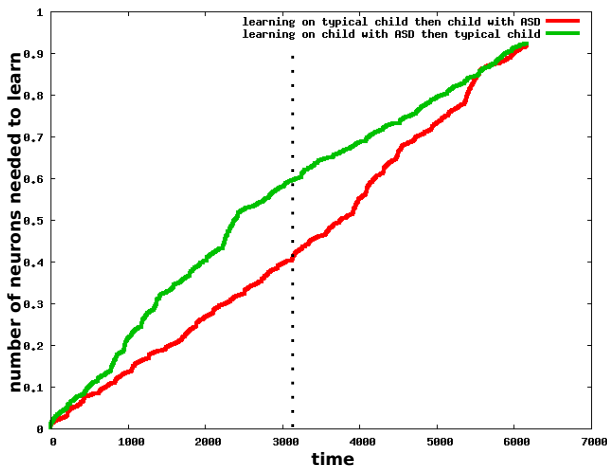


Fig. 9. Number of neurons needed to learn during the learning phase, which were obtained by testing two conditions: 1) the robot interacted with TD children, followed by children with ASD and 2) the robot interacted with children with ASD, followed by TD children. The change was made after 3100 images).

different intermediate steps occurred. Consequently, the results reflect the plasticity of the model, which depended on the inputs that had to be processed. The robot learned the task, but different steps were required to converge towards the solution. Fig. 10 shows the success rate of the postures recognition for

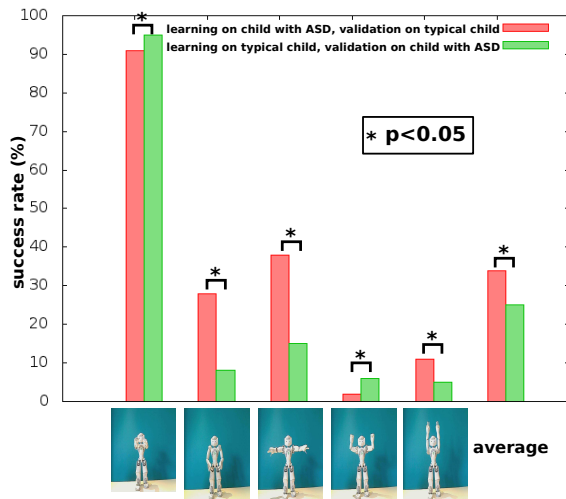


Fig. 10. The success rate for each posture was tested for two cases: 1) the robot learned with the children with ASD, and the validation was performed with the TD children, and 2) the robot learned by interacting with the TD children, and the validation was performed with the children with ASD.

both cases. The results show that better robot performances were obtained for the first case. Generalization was easier when learning was performed with the children with ASD and validation was performed with the TD children ($p < 0.05$). In summary, to answer the original question, it appears that (1) robot learning depended on the first partnership (ASD vs. TD) and the intermediate steps required for learning, instead of the similar number of neurons recruited at the end of the two conditions (Fig. 9); and (2) better postures recognition

was obtained when the system was first exposed to the most complex behavior (i.e., corresponding to the children with ASD $p < 0.05$).

D. Does learning with more groups of partners (2 in this experiment) during the learning phase affect posture recognition performance?

Given the influence of participants on learning, we investigated whether increasing the number of participants during the learning phase (in our case, two different groups of participants) affected posture recognition performance. During the learning phase, two groups of participants imitated the robot, and another group performed postures in front of the robot during the validation phase. The results are summarized in Fig. 11 and Fig. 12. The results show that the robot recognized the postures performed by a new participant. The neural network could generalize to participants who were not present during the learning phase. Thus, the robot could generalize posture recognition, even when learning was different in terms of the participants interacting with the robot. Several cross experiments were performed to investigate this impact under various conditions.

Fig. 11 shows the success rate for each posture, which was tested for two cases: learning was performed with both the adults and the TD children, and the validation was performed with the children with ASD. In the second case, learning was performed with both the adults and the children with ASD, and the validation was performed with the TD children. The results for both cases show that the robot's success rate improved when it learned with the adults and the group of children with ASD (i.e., the posture recognition performance improved $p < 0.05$). Fig. 12 also shows the effect of participant age on the

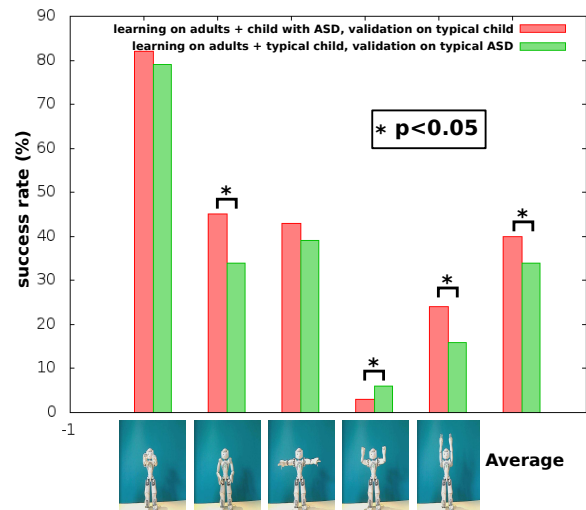


Fig. 11. The success rate for each posture was tested for two cases: 1) the robot learned by interacting with adults and TD children, and validation was performed with children with ASD and 2) the robot learned by interacting with adults and children with ASD, and the validation was performed with TD children.

success rate. The learning phase was conducted for two cases. In the first case, the robot learned by interacting with all the

children (i.e., the children with ASD and the TD children), and the validation was performed with the adults. In the second case, the robot learned by interacting with the adults, and the validation was performed by interacting with the TD children and the children with ASD. The robot's success rate improved

important element of learning a task and communicating with a partner (see below).

Second, the results showed the impact of a participant on both robot recognition and learning. The robot recognized and learned the postures of adults more easily than those of TD children or children with ASD. Table I shows the socio-demographic and clinical characteristics of the young participants. A therapist manually annotated the video recording of the interaction. High imitation scores for both groups of children were obtained for their interaction with the robot. In parallel, Fig. 7 also shows the success rate of the robot (i.e., for posture recognition), for which the postures of the adults were better known than those of the children, as previously mentioned. All these results show that the robot precisely analyzed the visual scene (i.e., a participant performing the posture in front of the robot) unlike the therapist annotating the video recording. From the therapist's perspective, the behavior of children with ASD was similar to that of TD children; however, the behaviors of the two groups of children were different for the robot (see Fig. 7 for the success rate, and Fig. 8 for the number of neurons needed to learn). Fig. 8 shows the number of neurons needed to learn for the different groups of participants. The results show that more neurons were recruited when the robot interacted with children with ASD than when the robot interacted with TD children. This result may be explained by the higher variability in the postures of the children with ASD than in those of the TD children. Here, the complexity was assessed in terms of the number of neurons needed to learn. Learning this task has a "neural cost" or a "cognitive cost" for the robot, i.e., the robot needs more or less neurons. It is important not to misinterpret the similarity in the imitation abilities of the children described by the therapist, which is a reflection of the time scales involved. The current robot visual scene analysis examined movement at the level of the video frame (15 images per second). None of the children with ASD exhibited behavioral motor dysfunction; thus, for the therapist, the children with ASD performed as well as the TD children did. However, the robot was able to detect subtle instabilities, i.e., in the spatial and temporal micro-stability and in learning.

Third, some results show that the first partnership, i.e., the children with ASD or the TD children, during the learning phase affected robot learning. Robot learning depended on the first group of participants because the intermediate steps in learning the task were different, and recognition was enhanced when the robot learned by interacting with the partner that exhibited the most complex behavior (i.e., the children with ASD). However, the robot had difficulties in generalizing to "unknown" participants. This result may be explained by the robot extracting some type of social signature. For example (see Fig. 10), the robot learned visual cues when it interacted with children with ASD that were generalized to TD children. However, the reverse situation was more difficult. In this study, metrics were used to evaluate the behavior of different participants interacting with the robot. The metrics were used to assess the quality and complexity of the interaction to evaluate how the robot reacted to different groups of participants. The results showed that robot learning depended on

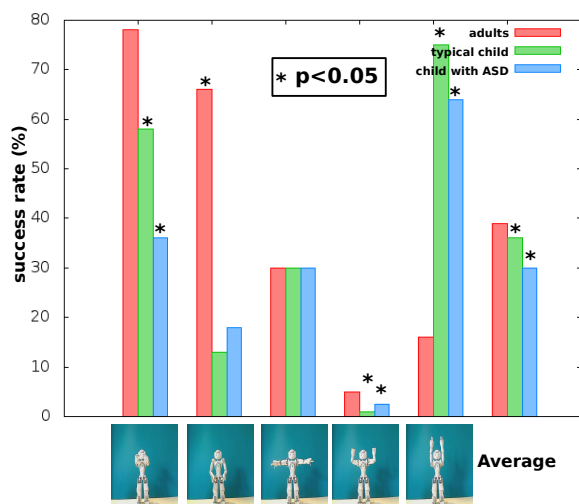


Fig. 12. The success rate for each posture was tested for three cases: 1) the robot learned by interacting with children with ASD and TD children, and the validation was performed with the adults; 2) the robot learned by interacting with adults, and the validation was performed with TD children; and 3) the robot learned by interacting with adults, and the validation was performed with children with ASD.

when it learned by interacting with all the children, and the generalization occurred via interaction with the adults ($p < 0.05$). Moreover, when the robot learned by interacting with the adults, and the robot had to reproduce the postures either with the children with ASD or the TD children, the success rate showed that it was easier to recognize the postures of the TD children than those of the children with ASD.

VI. DISCUSSION & CONCLUSION

In this paper, we investigated posture learning through imitation between a human and a robot. Our specific aim was to assess the influence of participants on robot learning. First, the results showed that the robot could learn a task autonomously by interacting with groups of participants (adults, TD children and children with ASD). The robot was able to learn, recognize and imitate many specific postures autonomously through an imitation game (i.e., one person copies another person). A period of 2 min was sufficient for the robot to learn the postures, and Fig. 7 shows the success rate for each posture for different participant groups interacting with the robot. Our approach enabled autonomous learning through interaction when a robot performed a gesture that is then imitated by a human partner. A sensory-motor architecture based on neural network enabled the task by associating what the robot did with what it saw. This new experiment confirmed previous results using the same architecture to learn other tasks such as facial expression recognition [11] or joint attention [12]. We emphasize that, during early development, imitation is an

the participants. The robot extracted social signatures for each participant group via the sensory-motor architecture.

The current study has several implications for developmental robotics. A sensory-motor architecture based on a neural network adapts to the environment (or participants) via the learning mechanism experienced during imitation and the statistical meaning of co-occurrences (i.e., the robot associates what it sees with what it does). Several parallels can be drawn between recent studies in developmental psychology and neuroscience. First, the importance of using statistical properties from the early environment was shown for human language development in the late 1990s in the seminal works of Saffran and Kuhl [48], [72]. Second, a key advantage of humans compared to other primates from an evolutionary perspective is that human cumulative culture is based on learning that requires a set of critical psychological processes, including teaching by verbal instruction, imitation and prosocial tendencies that are present in humans but are absent or impoverished in primates [26]. The setting and architecture in this study used imitation via an interactive game to produce changes in neural networks and learning. Third, learning is based on plastic changes in neural assemblies, as reflected by the modulation of brain responses during tasks and during development, which has led to the cortical recycling hypothesis [27]. Plasticity and neural commitment to specific tasks can even occur prenatally. In our study, it is important to note that this architecture, made of computational neural networks, has "similar" properties to those of real neurons during learning. For example, in child psychology, auditory learning implicates the formation and strengthening of neural long-term memory traces, which improve discrimination skills, particularly those required for forming the prerequisites for speech perception and understanding. Behavioral observations show that newborns react differently to unfamiliar sounds versus familiar sounds that they were exposed to as fetuses [62]. Our results, specifically the metrics we developed, show the plasticity of the sensory-motor architecture. The steps involved in learning a task differed from one group of participant to another. The robot adapted to participants to learn postures and converge toward a solution (Fig. 10).

Our robotics results may be controversial in the field of children with ASD. Recent literature has shown that robots produce a high degree of motivation and engagement in children with learning disabilities, especially autistic children, including those who are unlikely or unwilling to interact socially with human educators and therapists [74]. Some authors (e.g., [69]) have presented anecdotal results of introducing robots into experiments with ASD individuals. These authors have questioned why the best means of integrating robots into therapy sessions has not been investigated. These authors have consequently remained highly critical toward the results obtained in the field of robotics and ASD. Several open questions must be addressed to improve the research quality of this field: what are the best roles for robots in therapy? how can robots best be integrated into interventions? In this paper, we reversed the paradigm by asking the question of how the robot learning will react to different participants. This paradigm change was based on numerous studies of developmental

psychopathologies, which showed that mental disorders in mothers can negatively affect the quality of early interaction with their infants [33], [38] and that early developmental conditions in infants can impact the early interactive behavior of participants [19], [83]. This concept was used in a robotics experiment to evaluate the impact of participants on robot learning. To characterize this impact, we used metrics to evaluate how the robot learned postures recognition for different participant groups. The results confirmed the legitimacy of this concept by showing the impact of different groups of participants (children with ASD, TD children and adults) on robot learning.

In future studies, we want to develop a model to categorize a set of motor commands (i.e., motor commands for different articulations, such as the shoulder, elbow or knee) rather than for a single category per posture. Accordingly, the robot would perform a significant number of rich postures. A postural state could be obtained by categorizing the motor states (after several reproductions of the different gestures). These future models will improve the interactions between a robot and a human. Moreover, to improve the human robot communication, we want to build an architecture that integrates several capabilities (e.g., posture, emotion, and prosody) [28].

VII. ACKNOWLEDGMENTS

The authors would like to thank P. Gaussier for interesting discussions. This work was supported by the European Commission (FP7: MICHELANGELO under grant agreement no. 288241) and the endowment fund "Entreprendre pour aider". We would like also to thank MICHELANGELO study group: S. Bonfiglio, K. Maharatna, E. Tamburini, A. Giuliano and M. Donnelly.

REFERENCES

- [1] J. Aleotti and S. Caselli. Robust trajectory learning and approximation for robot programming by demonstration. *Robotics and Autonomous Systems*, 54(5):409–413, 2006.
- [2] P. Andry, P. Gaussier, S. Moga, J.P. Banquet, and J. Nadel. Learning and communication in imitation: An autonomous robot perspective. *IEEE transactions on Systems, Man and Cybernetics, Part A*, 31(5):431–444, 2001.
- [3] P. Andry, P. Gaussier, and B. Nadel, J. and. Hirsbrunner. Learning invariant sensory-motor behaviors: A developmental approach of imitation mechanisms. *Adaptive behavior*, 12(2), 2004.
- [4] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida. Cognitive developmental robotics: A survey. *Autonomous Mental Development, IEEE Transactions on*, 1(1):12–34, 2009.
- [5] P. Bakker and Y. Kuniyoshi. Robot see, robot do: An overview of robot imitation. In *AISB96 Workshop on Learning in Robots and Animals*, pages 3–11, 1996.
- [6] J.M. Baldwin. *Development and evolution*. JSTOR, 1902.
- [7] L. Berthouze, P. Bakker, and Y. Kuniyoshi. Learning of oculo-motor control: a prelude to robotic imitation. In *Intelligent Robots and Systems' 96, IROS 96, Proceedings of the 1996 IEEE/RSJ International Conference on*, volume 1, pages 376–381. IEEE, 1996.
- [8] A. Billard and G. Hayes. Learning to communicate through imitation in autonomous robots. In *Artificial Neural Networks/ICANN'97*, pages 763–768. Springer, 1997.
- [9] G. Bird, J. Leighton, C. Press, and C. Heyes. Intact automatic imitation of human and robot actions in autism spectrum disorders. *Proceedings of the Royal Society B: Biological Sciences*, 274(1628):3027–3031, 2007.
- [10] S. Boucenna, P. Gaussier, and P. Andry. What should be taught first: the emotional expression or the face? *epirob*, 2008.

- [11] S. Boucenna, P. Gaussier, P. Andry, and L. Hafemeister. Imitation as a communication tool for online facial expression learning and recognition. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 5323–5328. IEEE, 2010.
- [12] S. Boucenna, P. Gaussier, and L. Hafemeister. Development of joint attention and social referencing. In *Development and Learning (ICDL), 2011 IEEE International Conference on*, volume 2, pages 1–6. IEEE, 2011.
- [13] S. Boucenna, P. Gaussier, L. Hafemeister, and K. Bard. Autonomous development of social referencing skills. In *From Animals to Animats 11*, volume 6226 of *Lecture Notes in Computer Science*, pages 628–638, 2010.
- [14] S. Calinon, F. Guenter, and A. Billard. On learning, representing, and generalizing a task in a humanoid robot. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 37(2):286–298, 2007.
- [15] B. Calvo-Merino, D. Glaser, J. Grèzes, R. Passingham, and P. Haggard. Action observation and acquired motor skills: an fmri study with expert dancers. *Cerebral cortex*, 15(8):1243–1249, 2005.
- [16] C.L. Campbell, R.A. Peters, R.E. Bodenheimer, W.J. Bluethmann, E. Huber, and R.O. Ambrose. Superpositioning of behaviors learned through teleoperation. *Robotics, IEEE Transactions on*, 22(1):79–91, 2006.
- [17] M. Carpenter, K. Nagell, M. Tomasello, G. Butterworth, and C. Moore. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development*, 1998.
- [18] G.C. Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, pages 591–605, 1960.
- [19] D. Cohen, R. Cassel, C. Saint-Georges, A. Mahdhaoui, M.C. Laznik, F. Apicella, P. Muratori, S. Maestro, F. Muratori, and M. Chetouani. Do parentese prosody and fathers’ commitment facilitate social interaction in infants who later develop autism? *PLoS ONE*, 8(5):e61402, 2013.
- [20] J.F. Cohn and E.Z. Tronick. Mother-infant face-to-face interaction: Influence is bidirectional and unrelated to periodic cycles in either partner’s behavior. *Developmental psychology*, 24(3):386–392, 1988.
- [21] K. Dautenhahn. Roles and functions of robots in human society: implications from research in autism therapy. *Robotica*, 21(4):443–452, 2003.
- [22] K. Dautenhahn and I. Werry. Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, 12(1):1–35, 2004.
- [23] G. Dawson, S. Rogers, J. Munson, M. Smith, J. Winter, J. Greenson, A. Donaldson, and J. Varley. Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model. *Pediatrics*, 125(1):17–23, 2009.
- [24] A. de Rengervé, S. Boucenna, P. Andry, and P. Gaussier. Emergent imitative behavior on a robotic arm based on visuo-motor associative memories. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 1754–1759. IEEE, 2010.
- [25] A. de Rengervé, J. Hirel, P. Andry, M. Quoy, and P. Gaussier. On-line learning and planning in a pick-and-place task demonstrated through body manipulation. In *IEEE International Conference on Development and Learning (ICDL) and on Epigenetic Robotics (Epirob)*, volume 2, pages 1–6, Frankfurt am Main, Germany, August 2011. IEEE.
- [26] L.G. Dean, R.L. Kendal, S.J. Schapiro, B. Thierry, and K.N. Laland. Identification of the social and cognitive processes underlying human cumulative culture. *Science*, 335(6072):1114–1118, 2012.
- [27] S. Dehaene and L. Cohen. Cultural recycling of cortical maps. *Neuron*, 56(2):384–398, 2007.
- [28] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Vieux, and D. Cohen. Interpersonal synchrony: A survey of evaluation methods across disciplines. *Affective Computing, IEEE Transactions on*, 3(3):349–365, 2012.
- [29] J.J. Diehl, L.M. Schmitt, M. Villano, and C.R. Crowell. The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in autism spectrum disorders*, 6(1):249–262, 2012.
- [30] R. Dillmann. Teaching and learning of robot tasks via observation of human performance. *Robotics and Autonomous Systems*, 47(2):109–116, 2004.
- [31] A. Duquette, F. Michaud, and H. Mercier. Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots*, 24(2):147–157, 2008.
- [32] D. Feil-Seifer and M. Matarić. Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders. In *Experimental robotics*, pages 201–210. Springer, 2009.
- [33] T. Field. Infants of depressed mothers. *Development and psychopathology*, 4(1):49–66, 1992.
- [34] D. François, S. Powell, and K. Dautenhahn. A long-term study of children with autism playing with a robotic pet: Taking inspirations from non-directive play therapy to encourage children’s proactivity and initiative-taking. *Interaction Studies*, 10(3):324–373, 2009.
- [35] I. Fujimoto, T. Matsumoto, P.R.S. De Silva, Masakazu Kobayashi, and Masatake Higashi. Mimicking and evaluating human motion to improve the imitation skill of children with autism through a robot. *International Journal of Social Robotics*, 3(4):349–357, 2011.
- [36] P. Gaussier, K. Prepin, and J. Nadel. Toward a cognitive system algebra: Application to facial expression learning and imitation. In *Embodied Artificial Intelligence, F. Iida, R. Pfeifer, L. Steels and Y. Kuniyoshi (Eds.) published by LNCS/LNAI series of Springer*, pages 243–258, 2004.
- [37] G. Gergely and J. Watson. Early socio-emotional development: contingency perception and the social-biofeedback model. In P. Rochat, (Ed.), *Early Social Cognition: Understanding Others in the First Months of Life*, pages 101–136, 1999.
- [38] S.H. Goodman and H.E. Brumley. Schizophrenic and depressed mothers: Relational deficits in parenting. *Developmental Psychology*, 26(1):31, 1990.
- [39] F. Guenter, M. Hersch, S. Calinon, and A. Billard. Reinforcement learning for imitating constrained reaching movements. *Advanced Robotics*, 21(13):1521–1544, 2007.
- [40] S. Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.
- [41] G.M. Hayes and J. Demiris. *A robot controller using learning by imitation*. University of Edinburgh, Department of Artificial Intelligence, 1994.
- [42] Micha Hersch, Florent Guenter, Sylvain Calinon, and Aude Billard. Dynamical system modulation for robot learning via kinesthetic demonstrations. *Robotics, IEEE Transactions on*, 24(6):1463–1467, 2008.
- [43] A.J. Ijspeert, J. Nakanishi, and S. Schaal. Movement imitation with nonlinear dynamical systems in humanoid robots. In *Robotics and Automation, 2002. Proceedings. ICRA’02. IEEE International Conference on*, volume 2, pages 1398–1403. IEEE, 2002.
- [44] S.S. Jones. Imitation in infancy the development of mimicry. *Psychological Science*, 18(7):593–599, 2007.
- [45] S.S. Jones. The development of imitation in infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528):2325, 2009.
- [46] T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24:881–892, 2002.
- [47] K. Kaye and A. Fogel. The temporal structure of face-to-face communication between mothers and infants. *Developmental Psychology*, 16(5):454–464, 1980.
- [48] P.K. Kuhl, J.E. Andruski, I.A. Chistovich, L.A. Chistovich, E.V. Kozhevnikova, V.L. Ryskina, E.I. Stolyarova, U. Sundberg, and F. Lacerda. Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277(5326):684–686, 1997.
- [49] Y. Kuniyoshi. The science of imitation-towards physically and socially grounded intelligence. In *Special Issue TR-94001, Real World Computing Project Joint Symposium, Tsukuba-shi, Ibaraki-ken*. Citeseer, 1994.
- [50] C. Lord, M. Rutter, and A. Le Couteur. 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(5):659–685, 1994.
- [51] M. Lungarella, G. Metta, R. Pfeifer, and G. Sandini. Developmental robotics: a survey. *Connection Science*, 15(4):151–190, 2003.
- [52] A. Maurer, M. Hersch, and A. Billard. Extended hopfield network for sequence learning: Application to gesture recognition. In *Artificial Neural Networks: Biological Inspirations-ICANN 2005*, pages 493–498. Springer, 2005.
- [53] A.N. Meltzoff and M.K. Moore. Imitation of facial and manual gestures by human neonates. *Science*, 198:75–78, 1977.
- [54] L. Morency. Modeling human communication dynamics [social sciences]. *Signal Processing Magazine, IEEE*, 27(5):112–116, 2010.
- [55] L. Murray. The impact of postnatal depression on infant development. *Journal of Child Psychology and Psychiatry*, 33(3):543–561, 1992.
- [56] L. Murray and P.J. Cooper. *Postpartum depression and child development*. Guilford Press, 1997.
- [57] J. Nadel. *Imitation et communication entre jeunes enfants*, volume 13. Presses Universitaires de France-PUF, 1986.

[58] J. Nadel, M. Simon, P. Canet, R. Soussignan, P. Blancard, L. Canamero, and P. Gaussier. Human responses to an expressive robot. In *Procs of the Sixth International Workshop on Epigenetic Robotics*. Lund University, 2006.

[59] C.L. Nehaniv and K. Dautenhahn. The correspondence problem. *Imitation in animals and artifacts*, page 41, 2002.

[60] M. Pardowitz and R. Dillmann. Towards life-long learning in household robots: The piagetian approach. In *Development and Learning, 2007. IC DL 2007. IEEE 6th International Conference on*, pages 88–93. IEEE, 2007.

[61] M. Pardowitz, S. Knoop, R. Dillmann, and R. Zollner. Incremental learning of tasks from user demonstrations, past experiences, and vocal comments. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 37(2):322–332, 2007.

[62] E. Partanen, T. Kujala, R. Naatanen, A. Liitola, A. Sambeth, and M. Huotilainen. Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences*, 2013.

[63] J. Piaget. *La formation du symbole chez l'enfant: imitation, jeu et rêve, image et représentation*. Delachaux & Niestlé Paris, 1945.

[64] R.W. Picard. Emotion research by the people, for the people. *Emotion Review*, 2(3):250–254, 2010.

[65] A.C. Pierno, M. Mari, D. Lusher, and U. Castiello. Robotic movement elicits visuomotor priming in children with autism. *Neuropsychologia*, 46(2):448–454, 2008.

[66] G. Pioggia, R. Iglizzi, M. Ferro, A. Ahluwalia, F. Muratori, and D. De Rossi. An android for enhancing social skills and emotion recognition in people with autism. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 13(4):507–515, 2005.

[67] G. Pioggia, R. Iglizzi, M.L. Sica, M. Ferro, F. Muratori, A. Ahluwalia, and D. De Rossi. Exploring emotional and imitational android-based interactions in autistic spectrum disorders. *Journal of CyberTherapy & Rehabilitation*, 1(1):49–61, 2008.

[68] P. Ravindra, S. De Silva, K. Tadano, A. Saito, S.G. Lambacher, and M. Higashi. Therapeutic-assisted robot for children with autism. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 3561–3567. IEEE, 2009.

[69] D.J. Ricks and M.B. Colton. Trends and considerations in robot-assisted autism therapy. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 4354–4359. IEEE, 2010.

[70] B. Robins, K. Dautenhahn, and J. Dubowski. Does appearance matter in the interaction of children with autism with a humanoid robot? *Interaction Studies*, 7(3):509–542, 2006.

[71] B. Robins, K. Dautenhahn, R. Te Boekhorst, and A. Billard. Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society*, 4(2):105–120, 2005.

[72] J.R. Saffran, R.N. Aslin, and E.L. Newport. Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–1928, 1996.

[73] C. Saint-Georges, A. Mahdhaoui, M. Chetouani, M.C. Laznik, F. Apicella, P. Muratori, S. Maestro, F. Muratori, and D. Cohen. Do parents recognize autistic deviant behavior long before diagnosis? taking into account interaction using computational methods. *PLOS ONE*, 6(7):e22393, 2011.

[74] B. Scassellati. How social robots will help us to diagnose, treat, and understand autism. In *Robotics research*, pages 552–563. Springer, 2007.

[75] B. Scassellati, H. Admoni, and M. Mataric. Robots for use in autism research. *Annual Review of Biomedical Engineering*, 14:275–294, 2012.

[76] S. Schaal. Is Imitation Learning the Route to Humanoid Robots? *Trends in Cognitive Sciences*, 3(6):233–242, 1999.

[77] J.J. Steil, F. Röthling, R. Haschke, and H. Ritter. Situated robot learning for multi-modal instruction and imitation of grasping. *Robotics and Autonomous Systems*, 47(2):129–141, 2004.

[78] P. Stribling, J. Rae, and P. Dickerson. Using conversation analysis to explore the recurrence of a topic in the talk of a boy with an autism spectrum disorder. *Clinical linguistics & phonetics*, 23(8):555–582, 2009.

[79] A. Tapus, M. Mataric, and B. Scassellati. Socially assistive robotics. *IEEE Robotics and Automation Magazine*, 14(1):35, 2007.

[80] E.L. Thorndike. Animal intelligence: An experimental study of the associative processes in animals. *Psychological Monographs: General and Applied*, 2(4):i–109, 1898.

[81] M. Tomasello and M.J. Farrar. Joint attention and early language. *Child development*, pages 1454–1463, 1986.

[82] H. Wallon. *De l'acte à la pensée*. Flammarion, 1942.

[83] M.W. Wan, J. Green, M. Elsabbagh, M. Johnson, T. Charman, and F. Plummer. Quality of interaction between at-risk infants and caregiver

at 12–15 months is associated with 3-year autism outcome. *Journal of Child Psychology and Psychiatry*, 2012.

[84] A. Whiten and R. Ham. On the nature and evolution of imitation in the animal kingdom: reappraisal of a century of research. *Advances in the Study of Behavior*, 21:239–283, 1992.

[85] B. Widrow and M.E. Hoff. Adaptive switching circuits. In *IRE WESCON*, pages 96–104, New York, 1960. Convention Record.



typically developing children and children with autism spectrum disorder) on robot learning.

Sofiane Boucenna is postdoctorant at Pierre et Marie Curie University in the Institut des Systèmes Intelligents et de Robotique lab (ISIR). He obtained its PhD at the Cergy Pontoise University in France in 2011, where he worked with the Neurocybernetic team of the Image and Signal processing Lab (ETIS). His research interests are focused on the modelling of cognitive mechanisms and the development of interaction capabilities such as imitation, emotion and social referencing. Currently, he attempts to assess the effect of the type of partners (adults,



Salvatore Anzalone received his B.S. in Computer Engineering (2004) and M.S. Degree in Intelligent Systems (2006) from the University of Palermo. From the same University he received his Ph.D. degree in Computer Engineering (2010), guided by Prof. A. Chella and Prof. R. Sorbello. He cooperated with Prof. H. Ishiguro at the Osaka University during one year of his Ph.D. (2009) and during his post-doc year at University of Padova (2010), guided by Prof. E. Pagello and Prof. E. Menegatti. He received a grant from the Japan Society for the Promotion of Science and from April 2011 to April 2012 was a research fellow at the Intelligent Robotics Laboratory directed by Prof. H. Ishiguro at Osaka University (2011–2012). From May 2012 he is a Post-Doc researcher at University Pierre et Marie Curie in Paris, France, working with Prof. M. Chetouani at the Institute for Intelligent Systems and Robotics (ISIR-UPMC). Last experiences at ISIR-UPMC in Paris involve the use of personal robots for autistic children in the FP7 European Project Michelangelo.



Elodie Tilmont clinical psychologist graduated from the University of Amiens in France (2009). Work with people with autism spectrum disorder. Referent of a day hospital for autistic children since 2010 in the department of child psychiatry Prof. Cohen in the Pitié-Salpêtrière hospital. Since 2012 working with Prof. Mr. Chetouani at the Institute for Intelligent Systems and Robotics (ISIR-UPMC) within the framework of a European Project Michelangelo (new technology of communication for people with autism).



Prof. Mohamed Chetouani Mohamed Chetouani is the head of the IMI2S (Interaction, Multimodal Integration and Social Signal) research group at the Institute for Intelligent Systems and Robotics (CNRS UMR 7222), University Pierre and Marie Curie-Paris 6. He received the M.S. degree in Robotics and Intelligent Systems from the UPMC, Paris, 2001. He received the PhD degree in Speech Signal Processing from the same university in 2004. In 2005, he was an invited Visiting Research Fellow at the Department of Computer Science and

Mathematics of the University of Stirling (UK). Prof. Chetouani was also an invited researcher at the Signal Processing Group of Escola Universitaria Politecnica de Mataro, Barcelona (Spain). He is currently a Full Professor in Signal Processing, Pattern Recognition and Machine Learning at the UPMC. His research activities, carried out at the Institute for Intelligent Systems and Robotics, cover the areas of social signal processing and personal robotics through non-linear signal processing, feature extraction, pattern classification and machine learning. He is the head of the interdisciplinary research group IMI2S (Interaction, Multimodal Integration and Social Signal) gathering researchers from social signal processing, social robotics, psychopathology and neuroscience. This group develops models and methods for the analysis, recognition and prediction of social signals, behaviors with a life-span perspective with a particular attention to disorders (autism, Alzheimer). He has published numerous research papers including some in high impact journals (Plos One, Biology Letters, Pattern Recognition, IEEE Transactions on Audio, Speech and Language Processing). He is also the co-chairman of the French Working Group on Human-Robots/Systems Interaction (GDR Robotique CNRS) and a Deputy Coordinator of the Topic Group on Natural Interaction with Social Robots (euRobotics).



Prof. David Cohen received a M.S. in neurosciences from the UPMC and the Ecole Normale Supérieure in 1987, and a M.D. from Necker School of Medicine in 1992. He specialized in child and adolescent psychiatry and certified in 1993. His first field of research was severe mood disorders in adolescent, topic of his PhD in neurosciences (2002). He is Professor at the UPMC and head of the department of Child and Adolescent Psychiatry at La Salpêtrière hospital in Paris. His group runs research programs in the field of autism and other pervasive

developmental disorder, severe mood disorder in adolescent and childhood onset schizophrenia and catatonia. He is a member of the International Association of Child and Adolescent Psychiatry and Allied Disciplines, the European College of Neuro-Psychopharmacology, the European Society of Child and Adolescent Psychiatry, and the International Society of Adolescent Psychiatry