Four Tasks of a Robot-assisted Autism Spectrum Disorder Diagnostic Protocol: First Clinical Tests

Frano Petric, Kruno Hrvatinić, Anja Babić,
Luka Malovan, Damjan Miklić and Zdenko Kovačić
LARICS Laboratory
Faculty of Electrical Engineering and Computing
University of Zagreb
Email: larics@fer.hr

Maja Cepanec, Jasmina Stošić and Sanja Šimleša
Faculty of Education and Rehabilitation Sciences
University of Zagreb
Email: mcepanec@globalnet.hr

***Abstract*—Notwithstanding intensive research and many scientific advances, diagnosing autism spectrum disorders remains a slow and tedious process. Due to the absence of any physiological tests, the outcome depends solely on the expertise of the clinician, which takes years to acquire. Complicating the matter further, research has shown that inter-rater reliability can be very low, even among experienced clinicians. As an attempt to facilitate the diagnostic process and make it more objective, this paper proposes a robot-assisted diagnostic protocol. The expected benefit of using a robot is twofold: the robot always performs its actions in a predictable and consistent way, and it can use its sensors to catch aspects of a child's behavior that a human examiner can miss. In this paper, we describe four tasks from the widely accepted ADOS protocol, that have been adapted to make them suitable for the Aldebaran Nao humanoid robot. These tasks include evaluating the child's response to being called by name, symbolic and functional imitation, joint attention and assessing the child's ability to simultaneously communicate on multiple channels. All four tasks have been implemented on the robot's onboard computer and are performed autonomously. As the main contribution of the paper, we present the results of the initial batch of four clinical trials of the proposed robot assisted diagnostic protocol, performed on a population of preschool children. The results of the robot's observations are benchmarked against the findings of experienced clinicians. Emphasis is placed on evaluating robot performance, in order to assess the feasibility of a robot eventually becoming an assistant in the diagnostic process. The obtained results indicate that the use of robots as autism diagnostic assistants is a promising approach, but much work remains to be done before they become useful diagnostic tools.**

I. Introduction

Autism spectrum disorder (ASD) is a developmental disorder characterised by impairment in social interaction, verbal and nonverbal communication and by repetitive behaviours and interests. It has become a commonly diagnosed neurodevelopmental disorder, with increasing prevalence rates, affecting about one in every 100 children [1], [2], and there are no medical markers of autism that could be used in a diagnostic process. Therefore, the diagnosis of ASD is based solely on behavioural observations made by experienced clinicians who rely on:

- Using criteria from Diagnostic and Statistical Manual of Mental Disorders, currently DSM-V [3]
- Testing children using Autism Diagnostic Observation Schedule (ADOS) [4]
- Interview with the caregivers using Autism Diagnostic Interview-Revised (ADI-R) [5]

However, specific behaviors that are included in diagnostic frameworks and the point at which individual differences in behavior constitute clinically relevant abnormalities are largely arbitrary decisions [6]. Studies have shown [7] that the agreement between clinicians on different DSM-IV criteria for autism varies from 0.58 to 0.79. Even when using the *golden standard* instrument in ASD diagnostic procedure (ADOS), the inter-rater reliability is for some ratings as low as 0.38 in modules used with preschool children [4]. The main reason for these discrepancies is that the diagnostic procedure is very complex, due to simultaneous observation, coding and interpretation of many behaviours, as well as administration of various specific tasks. Additionally, the process of learning to observe and code the behavior and the procedure of achieving 80% inter-rater reliability on ADOS might last for a few years. All in all, there is an increasing need for a more objective approach that would help clinicians in gathering multimodal information and coding the social behavior, and modern robotics technologies seem capable of providing the right tools to fill this need. Their potential is evidenced by the rapid growth in the field of Socially assistive robotics [8].

Since autistic children tend to interact with technical devices more than with humans around them, robotics has entered the domain of autism relatively easily. There are many studies on employment of different kinds of robots in the teaching and intervention for children with ASD. Such robotic platforms may take the form of a humanoid robot (such as KASPAR) or the form of a mobile robot (IROMEC) [9]. In both cases, the robot serves as a social mediator, eliciting and enhancing interaction between autistic children and people in their surroundings, mainly their therapists and parents. Interaction with the robot is based on play scenarios and aims towards improving general social skills.

While there are many robotic applications in teaching and intervention, such as [10], [11], diagnostic applications are scarce, although *there exists a need for quantitative, objective measurements of social functioning for diagnosis, for evaluating intervention methods, and for tracking the progress of individuals over time* [12]. This is mainly due to the complexity of the diagnostic process itself, but is also dependent upon efficient processing and reasoning algorithms that are to be implemented on the robot. Work in [13], [14] resulted in the proposal of several quantitative metrics for social response during the diagnostic process, but the robot used had no sensory capabilities and was not able to detect and classify the child's behavior. Quantitative data was collected through passive sensors installed in the examination room.

Although not deployed on robots, there is an obvious trend of involvement of computer scientists who seek to improve the diagnostic process for ASD. Such effort is presented in [15], where language and speech were processed in order to quantify some of the behavioral patterns. Such processing is often called behavioral signal processing (BSP) and it was shown that BSP based only on speech and language can be used to quantify and model autistic behavior to some extent. Since BSP was deployed during a natural conversation of the examiner and the child, it was observed that the behavior of examiners changed depending on perception of the child's behavior, indicating the need for consistent and repetitive stimuli (also called *social presses*) which can be achieved through the use of a robotic examiner.

The main idea behind the work presented in this paper is to enhance the behavior-based diagnostic protocol with tasks which are performed and evaluated by a humanoid robot. The advantage of using a robot is twofold: the robot performs social presses in a completely consistent and repeatable way and it can evaluate the child's reactions using its sensors in a quantitative and unbiased way, thus making the diagnostic procedure more objective. Furthermore, processing of the gathered data can be performed automatically and the robot can provide the clinician with coded information which can be directly fed into standardized ASD classification algorithms, making the diagnostic procedure more efficient. The humanoid robot Nao [16] has been chosen to perform the diagnostic tasks due to its small size and amiable appearance to which children react positively.

As a first step towards a full robot-assisted diagnostic protocol, we have selected four tasks from the "golden standard" ADOS test and adapted them to the capabilities of the Nao robot. These four tasks are:

- Response to name,
- Functional and symbolic imitation,
- Joint attention,
- Simultaneous multi-channel communication assessment.

At this point, the main criterion in task selection and design was the feasibility of execution by the Nao robot. The relevance of these tasks for the diagnostic procedure is the subject of our ongoing research. In this paper we present the tasks from an engineering perspective and discuss the perception and actuation primitives that were implemented in order to enable their successful autonomous execution by the Nao. Finally, we present the results of the first batch of clinical trials, where the robot performed the four ASD diagnostic tasks with three children that have been previously diagnosed with ASD, and one typically developing child.

The paper is organized as follows. Section II describes the four diagnostic tasks performed by the robot in more detail. The implementation of the perception and actuation primitives necessary to perform the tasks is presented in Section III. In Section IV we discuss the methodology and describe the clinical setting used to evaluate the performance of the robot. The clinical trial results are presented in Section V. Section VI provides concluding remarks and an outlook on future work.

II. ROBOT-ASSISTED PROTOCOL TASKS

The proposed robotic protocol consists of four tasks, which are adapted from the Module 2 of the ADOS protocol. These tasks were chosen among several tasks that were suggested by ADOS trained experts, based on the robot's capabilities. Observations from all four tasks are collected independently and analyzed after each evaluation session in order to obtain the overall assessment of robot's performance and eventually the child's status. Tasks are to be performed by Nao, 58 cm tall humanoid robot with 25 degrees of freedom (DOF), each hand having 5 DOFs. It is equipped with two high definition (720p) cameras, two speakers, four microphones and several other sensors, such as ultrasound range sensors, tactile sensors, force sensitive resistors, accelerometers etc. It runs the OpenNAO operating system, based on the Gentoo Linux distribution, on an Intel Atom Z530 processor. This allows for easy graphical programming through the Choregraphe software, and all of the robot's hardware is available through Python and C++ APIs. Depending on the intensity of activity, Nao's autonomy ranges from 60 to 90 minutes. The sensory apparatus and processing power was deemed sufficient to perform the tasks of the proposed robot-assisted protocol. This conclusion is based on the relatively successful work that Nao robots already perform with children with ASD. Authors in [17] report the use of Nao in a robot-mediated therapy task which is very similar to the joint attention task of the proposed protocol. In that research, Nao's capabilities are augmented with a network of infra-red cameras which track the head of the child wearing a hat with infra-red LEDs. The study showed the increased level of child's engagement when the joint attention task was performed by the robot.

A. Response to name

The goal of this task is to evaluate how the child responds when called specifically by name, with the intent to draw the child's attention from something else to the speaker.

The task begins with the child being distracted by playing with a favourite toy or some other object, while the robot is observing the child from a distance of about 1-1.5 meters, positioned in such way that the child has to turn to look at the robot. The robot needs to detect that the child is occupied, and then call the child by name. After the call, the robot detects eye-contact through face detection in order to evaluate if the child transferred the attention from the toy to the robot. If the child responds, the task is completed and the child is rewarded. If there is no response, the procedure is repeated four more times, with 5 seconds between the calls if the child is not responding. If the child does not respond after five iterations, the robot calls the child two more times using a specific phrase, referring to a child's favorite toy, food, activity etc. (i.e. *Hey,* *Luka, here is a car*). Then, the robot waits for a response for five seconds, and the task ends.

B. Joint attention

The goal of this task is to evaluate the child's response to transferring the attention to another object by moving head and pointing. For this, two robots are used, an active one which calls, turns his head and points, and a passive one, which waits for the child to respond. The beginning of the task is similar to the *response to name* task, the child must be occupied, not paying attention to the robots.

The first robot calls the child and turns its head towards the other robot, which awaits the child's response. If there is no response, this procedure is repeated five times. Additionally, the robot reinforces the initiation by performing pointing gestures and using stronger voice commands, as shown in Fig. 1.

Fig. 1. Joint attention - robot calling the child and pointing

If there is no response after 5 calls, the other robot tries to attract the attention by waving its arms, flashing its LED's and making different sounds. After trying to draw the attention twice, the second robot awaits child's response for several seconds, and then the task ends. If at any point the child responds, task is ended by rewarding the child by performing interesting motions such as dancing and playing sounds.

C. Play request

The goal of this task is to instigate the child's vocalizations and eye-contact in coordination with hand and body gestures, in order to assess the child's ability to communicate on multiple channels simultaneously. The role of the robot is to perform an action which the child could find attractive, such as releasing soap bubbles or dancing. By abruptly stopping the behaviour, the robot stimulates child's actions expecting the child to ask for more.

The play request task starts with the robot standing up from the sitting position. The robot then observes the child's behaviour, with focus on vocalization and eye-contact. If the robot detects the child's intention to initiate interaction by either eye contact or vocalization, it repeats the behavior immediately instead of waiting for the predefined amount of time. The procedure is repeated 3 times, regardless of the child's behavior and response.

D. Functional and symbolic imitation

The goal of this task is to evaluate the child's ability to imitate simple actions, both with real objects with real function - functional imitation and with objects that have no obvious function - symbolic imitation. The task begins with the child and the robot on opposite sides of the table, facing each other, as shown in Fig. 2.

Fig. 2. Robot demonstrating an action to the child

The functional and symbolic imitation task consists of 3 imitation subtasks with three objects: a toy frog, cup and wooden cylinder:

1) Robot shows the child the frog, makes a frog-like arm movement which represents the frog jumping while reproducing the sound of a frog
2) Robot shows the child the cup, simulates drinking by raising it to its mouth while playing the drinking sound
3) Robot shows the child the cylinder, simulates aeroplane movement while playing the sound of aeroplane flying

After each demonstration, the robot puts the object back to the table and indicates that it is the child's turn by saying *Now you!* or a similar phrase. Then, the robot observes the behaviour of the child, detects if the appropriate gesture was performed and classifies any vocalization the child made. If the child does not respond correctly, each demonstration is repeated three times before going on to the next demonstration.

During all four tasks, the robot needs to detect eye-contact, detect, analyze and classify the child's vocalization and log the data in a human-readable format to make the clinician's job easier after a session with a child.

III. Visual and auditory perception

In this section we describe the perception primitives necessary to endow the robot with the capabilities necessary for performing the diagnostic tasks. These capabilities are eye contact tracking, vocalization classification, and gesture classification.

A. Eye contact detection

Eye contact is the main source of information for the robot to deduce whether the child is focused on it or something else. To detect eye contact, the robot is first searching for a face in one of the two cameras using the built-in Nao module ALFaceDetection. The face detection algorithm first segments the image to extract areas that have color similar to skin color, then searches for the area with holes that could represent eyes and mouth. Due to the limitations of such face detection algorithm, child has to face the robot directly for the face to be detected, which is used by the robot as the information that the child is looking at it. This approach causes two problems that need to be mentioned:

1) Child could be facing somewhere else but still looking at the robot
2) Child could be facing the robot but looking somewhere else

While the first problem cannot be resolved without using external sensors that are to be mounted either in the room or the child itself, the second problem can be resolved by estimating the eye gaze direction from the information about the face, eyes and irises. This approach is the subject of ongoing research.

B. Vocalization analysis

the vocalization analysis module records sound from NAO's microphones and classifies it in two classes: unarticulated and articulated speech. The module extracts several feature from the audio signal, such as high zero crossing rate ration (HZCRR), low short term energy ration (LSTER) and spectrum flux (SF). During the initial tests on artificial samples, the spectrum flux feature was discarded as redundant for the classification procedure. The classifier is a simple k-nn (k nearest neighbors) classifier, trained on an artificially created database consisting of 104 samples:

- 28 samples of articulated speech
- 18 samples of unarticulated speech
- 35 samples of music of different types
- 14 samples of different human-made sounds (coughing, clapping etc.)
- 9 empty samples containing background noise

Since both HZCRR and LSTER have values in the domain of real numbers, the Euclidean norm is used as a measure of distance between samples.

Classifier performance and sample quality was validated through *leave-one-out cross validation* tests, focusing on accuracy, precision, recall and specificity of each sample. The results of these tests are summarized in Table I.

TABLE I. K-NN SOUND CLASSIFIER VALIDATION

Parameter k	5	7	9
Accuracy	0.8942	0.8942	0.9038
Precision	0.7576	0.7576	0.7500
Recall	0.8929	0.8929	0.9643
Specificity	0.8947	0.8947	0.8816

As can be seen from the Precision row in Table I, and was also observed during clinical trials, the classifier sometimes gives false positives on articulated speech (0.75 precision rate means that of all samples that are classified as articulated speech samples only 75% are in fact articulated speech samples). To mitigate this problem, real data form the sessions with children has been collected which will be used to improve the classifier. Along with HZCRR and LSTER features, which are defined in the time domain, several other features in the frequency domain were recorded (such as audio spectrum centroid (ASC), audio spectrum spread (ASS), audio spectrum flatness (ASF) etc.) for improving the classifier.

C. Intelligent grasping

To successfully perform the functional and symbolic imitation tasks, the robot must be able to detect the test objects' position in three-dimensional space, successfully grasp any of the test objects and pick them up from a table in front of it, and differentiate between actions done by the child while the child holds the object. Due to the lack of a binocular camera or other precise range finding sensor, camera data is combined with prior knowledge about the environment and the object to estimate the position of an object in space. The grasping algorithm makes the following assumptions: the approximate object hue is known, there is only one object of that color on the table, the object is graspable and the table height is known.

An acquired image is segmented by first converting it into the HSV color space, then constructing a one-dimensional Hue histogram, which is median filtered to remove noisy outliers. Assuming an N-modal histogram, the mode closest to a given seed hue value corresponding to the object's color is chosen to construct a hue range. All pixels within this range which also satisfy additional saturation and value criteria are marked as belonging to the object. The image grab point is extracted from the marked pixels by calculating their centroid. Since one of the objects to be grabbed is a cup, the object is also checked for holes. If a hole is found, the image grab point is instead calculated from the part of the object to the left or to the right of the hole, whichever contains the smallest number of pixels. This part is considered to be the cup handle, as can be seen in Figure 3.

Fig. 3. Grab point identification.

Once the object grab point is calculated, the grab hand is determined. For an object without a hole, the grab hand is whichever robot hand is closest to the object. If a hole is detected, the object is grabbed from the direction of its handle. The spatial grab point is determined by constructing a line through the camera focus and the pixel grab point, and intersecting it with several plane models. However, because of monocular vision limitations, it is necessary to supply the height of the table the object is resting on to the algorithm.

Once the robot has calculated the spatial grab point, it may proceed to grab the object using the hand and direction determined by the algorithm and perform an action the child is expected to imitate. This is done by sending preprogrammed trajectory references to the robot's joints, which allows the robot to perform an identical motion every time the experiment is attempted. After the action is completed, the object is returned to the position it had been at before the action was started.

D. Gesture recognition

Object tracking starts after the robot has indicated to the child that it is now its turn to repeat the shown action. The aim of tracking is to compare the way the child is moving the object to a general description of the shown action or gesture, and to recognize when the gesture has been successfully mimicked. This requires consistent object tracking the object must not leave the robot's field of vision, it must be detected in each frame and it must not be confused with a part of the background. For this purpose, we use tracking based on a Gaussian mixture model constructed offline from a series of object images. The tracking process is complex, and is fully described in [18]. This algorithm's robustness satisfies the requirement for consistent tracking and keeps the object separate from the background, even if it is of a very similar or identical color. Using the robot's head to track the object ensures that the robot will always see the object. The object tracking module outputs the object's 2D position in the form of head rotation angles at each time step. This data combined with the detection times for each point forms an object trajectory which can be compared to a gesture model at each time step. Comparison is done by separating the trajectory into fixed length segments and calculating their direction angle. Each segment is then categorized into one of eight principal directions: right, up-right, up, up-left and so on, based on the previous segment's categorization and the segment's angle. The directions each claim an angle range of $\frac{\pi}{4}$, with an additional overlap implementing a hysteresis switching principle to avoid bounces between categorization into adjacent directions. The gesture model consists of an ordered list of principal directions. A trajectory is considered a match when a continuous string of segments have directions identical to the model, disregarding adjacent duplicate directions. For example, a *drinking* movement consists of an up-down motion and has a gesture model of $\{up, down\}$. A trajectory would be considered a match if its directions were categorized as left, up, up, up, down, down, right but not if they were categorized as $\{left, up, up, left, down, down\}$. A sample result of trajectory identification for the *frog* gesture is given in Figure 4.

Segments between red points are negative matches, while segments between orange points represent positive matches. Each green point represents a change in direction. The model matched here can be represented as $\{up, up - left, left, down - left, down\}$. Gestures that the robot recognizes during the imitation task and their representations are summarized in table II.

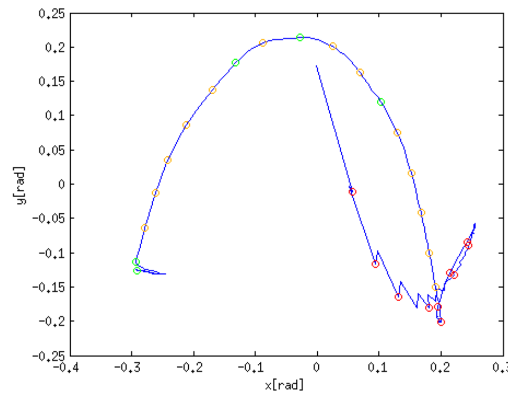

Fig. 4. Observed frog-like trajectory

TABLE II. GESTURE REPRESENTATION

Gesture	Representation	Alternative
Frog	up-left, left, down-left	up-right, right, down-right
Drink	up-down	N/A
Airplane	left, right	N/A

IV. METHODOLOGY

First clinical tests of the robot-assisted ASD diagnostic protocol were performed through sessions with four children, three of them being already diagnosed with ASD and one typically developing child, as shown in Table III.

TABLE III. PARTICIPANTS.

ID	Age	Condition
ASD001	7y 8m	ASD
ASD002	5y 9m	ASD
ASD003	5y 3m	ASD
CON001	6y 4m	typically developed

Sessions with children were carried out in the examination room of the Centre for Rehabilitation, Faculty of Education and Rehabilitation Sciences, University of Zagreb. The room is equipped with a sound and video recording system, along with a one-way mirror for observers. Data obtained through video recording is used to validate the robot's performance (see Section V).

Although the protocol consists of four independent tasks, the implementation and setup in the examination room resulted in the following order of tasks:

1) Response to name;
2) Joint attention;
3) Play request;
4) Functional and symbolic imitation.

During all four tasks, the robot tracks and logs the observations of the child's face and vocalizations along with social presses that it performs. Additionally, during the imitation task, the robot tracks the object after the demonstration and evaluates the trajectory that the child performed online to detect whether the imitation was successful. Logs are generated in a time efficient way to preserve processing time for higher priority tasks, and human-readable logs with coding are generated after the session.

Along with the child and the robot, several other people were present in the examination room: one or both parents/caregivers for the child to be comfortable, a roboticist that knows to operate the robot to keep both the child and the robot safe if unexpected errors occur, and an expert clinician, as shown in figure 5.

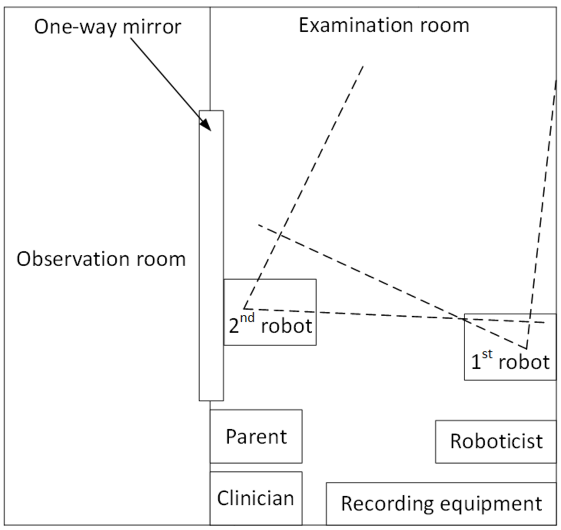

Fig. 5. Layout of the examination room.

The role of the clinician is to observe the interaction between the child and the robot, log the same data that the robot logs and code the behaviour of the child.

A. Coding the behaviour of the child

During the sessions, expert clinicians proficient in ASD diagnostics observe the interactions between the child and the robot and code the behaviour as summarized in table IV.

TABLE IV. TASK CODING.

Task	Behaviour	Code
Response to name	Child responds after 1 or 2 calls	0
	Child responds after 3, 4 or 5 calls	1
	Child did not respond to call, but reacted to phrase or vocalized	2
	Child did not react at all	3
Joint attention	Child reacted to first three calls	0
	Child reacted after 4 or 5 calls	1
	Child reacted after the second robot activates	2
	Child did not react at all	3
Play request	Simultaneous eye contact and other behaviour (gesture, vocalization)	0
	Only other behaviour without eye contact	1
	Child touched robot	2
	Child did not react at all	3
Imitation	Child imitated all three behaviours	0
	Child imitated only functional tasks	1
	Child grabbed the object but performed no imitation	2
	Child did not react at all	3

As can be seen, the coding scheme is very similar to that of the ADOS protocol. Along with codes for each task, clinicians log other behaviors that emerged during the interaction, such as vocalizations, gestures and touching of the robot.

After the session, the robot generates codes based on the log from the session, which is then compared to the observations of the clinicians. Additionally, sessions are recorded and analyzed afterwards to obtain an objective assessment of the interaction, validating both observations of the robot and the clinicians.

V. Results of first clinical tests

As already mentioned, sessions were carried out in the controlled clinical setting of the Centre for Rehabilitation, Faculty of Education and Rehabilitation Sciences, University of Zagreb. Parents and caregivers of all children signed a consent form, allowing the sessions to be filmed and results obtained to be used in further research and dissemination. Results of the session with the first child are presented in Table V.

TABLE V. SESSION RESULTS FOR CHILD ASD001.

Task	Successful		Vocalization		Speech		Code	
	H	R	H	R	H	R	H	R
Response to name	✓	✓	✓	✓	✓	×	**0**	**1**
Joint attention	×	×	×	✓	✓	✓	**3**	**3**
Play request	✓	✓	✓	×	✓	✓	**0**	**1**
Imitation	✓	✓	×	✓	✓	×	**1**	**1**

Table V compares the performance of the robot with observations made by the clinician, which are additionally confirmed through video analysis. It can be seen that the robot produced code 1 for the first task by detecting child's response after fourth iteration, while it missed the eye contact which occurred after second iteration (which results in code 0, see table IV). Regarding the sound analysis module, it's performance is really poor, which is mainly caused by not having proper training data. Additionally, some of the information is lost due to the robot not being able to listen to sound while it is playing sound or moving and not being able to discriminate between child's vocalizations and vocalizations of others in the room. In the second task the robot again picked up some vocalizations that the clinician did not hear from the child, but correctly coded the task as the child did not respond. During the play request task, the robot did not detect the simultaneous emergence of several behaviours, therefore it ended up coding the task differently. The robot's performance of functional and symbolic imitation resulted in correctly grabbing, demonstrating and observing the gestures of the child.

Table VI summarizes the results obtained during the session with the second child.

TABLE VI. SESSION RESULTS FOR CHILD ASD002

Task	Successful		Vocalization		Speech		Code	
	H	R	H	R	H	R	H	R
Response to name	✓	✓	×	✓	×	×	**0**	**1**
Joint attention	×	×	×	×	×	×	**3**	**3**
Play request	×	×	×	×	×	×	**3**	**3**
Imitation	✓	✓	×	×	×	×	**0**	**0**

Similar to the response to name task of the first session, the robot did not detect the first eye contact which occurred after the second call, but this time the main reason was that the child was out of the robot's field of view, indicating the need for implementing some kind of child tracking. When the child moved in front of the robot after the third call, the robot successfully detected eye contact and ended the session with code 1. Joint attention task was performed and coded in the same way that the human examiner coded it, since the child did not respond to any calls. Since the robot performs somewhat aggressive moves during the play request task, the child got scared and did not react to the robot's initiation at all. The imitation task was successfully performed by the robot, but the child did not react to the robot's instructions to imitate. However, after the parent explained to the child what needs to be done, the child successfully imitated all gestures, which the robot correctly recognized. Although such performance of the task would not be valid for the diagnostic process, it can be used to validate the performance of the robot.

The third child was a typically developing one from the control group, and the results of that session are in Table VII.

TABLE VII. SESSION RESULTS FOR CHILD CON001

Task	Successful		Vocalization		Speech		Code	
	H	R	H	R	H	R	H	R
Response to name	×	×	×	✓	×	×	**3**	**2**
Joint attention	Not performed due to child losing interest							
Play request	×	×	×	✓	×	✓	**3**	**1**
Imitation	×	×	×	✓	×	×	**3**	**3**

The robot coded the first task with code 2, since the child did not respond but the robot detected vocalizations. Additionally, the child showed no interest in the robot, causing the decision to skip the joint attention task since the play request should be more fun and interesting. Although the robot performed it's initiations for the interaction, the child did not react. The robot coded the task with code 1 because it detected the vocalization of other people in the room who tried to persuade the child to play with the robot. Similarly, the child did not react to the robot demonstrations in the functional and symbolic imitation task.

The fourth and final session of first batch of clinical tests was carried out with a child already diagnosed with ASD, and results are shown in Table VIII.

TABLE VIII. SESSION RESULTS FOR CHILD ASD003

Task	Successful		Vocalization		Speech		Code	
	H	R	H	R	H	R	H	R
Response to name	×	×	×	✓	✓	×	**2**	**2**
Joint attention	✓	✓	×	✓	✓	✓	**2**	**2**
Play request	✓	✓	×	✓	✓	✓	**0**	**1**
Imitation	Not performed due to robot malfunction							

Both the robot and human examiner coded the first session with code 2, following the occurrence of child's vocalization (note that the child did not respond and that the sound was not correctly classified). the joint attention task was successfully completed by both robots and the child, resulting in code 2, meaning that the child responded when the second robot activated to attract attention. Throughout the play request task, the child was out of the field of view so the robot could only register partial information, therefore coding the task differently. During this task, it was observed that the child was touching the robot rather than trying to vocalize or perform gestures, indicating a need to develop some kind of distance tracking since distance can also be used as an indicator of attention. In this session, imitation was not performed due to robot malfunction.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have introduced the concept of a robot-assisted diagnostic protocol for Autism Spectrum Disorders (ASD). We described four tasks, inspired by the state of the art ADOS diagnostic procedure, which have been adapted for performance by the Nao humanoid robot. We presented the technical solutions that were implemented in order to endow Nao with the capabilities to perform these tasks. The described work represents the first step towards the development of a robotic assistant for ASD diagnostics, with the goal of making the diagnostic procedure shorter and more reliable. We presented the results of the first batch of clinical trials, which included three children previously diagnosed with ASD, and one typically developing control child, all of preschool age. The observations made by Nao during the diagnostic procedure were benchmarked against the observations of experienced clinicians. Technical difficulties notwithstanding, the observations made by the robot matched the clinicians' observations in most tests, and the overall results have been deemed promising by clinicians.

In our future work, we plan to pursue two main goals. On one hand, we will be working on improving the technical aspects of the task implementations by improving the sound classifier, using face analysis to improve the eye gaze direction estimation, and fusing the information provided by other sensors available on the robot, such as touch sensors and ultrasound proximity sensors. On the other hand, we will perform further extensive clinical trials in order to assess and quantify the relevance of the proposed tasks for improving the ASD diagnostic procedure, and expand the number and scope of the robot -performed tasks accordingly.

REFERENCES

[1] M.D. Kogan, S.J. Blumberg, L.A. Schieve, C.A. Boyle, J.M. Perrin, R.M. Ghandour, G.K. Singh, B.B. Strickland, E. Travathan, and P.C. van Dyck. Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007. *Pediatrics*, 154(5):1395–1403, 2009.

[2] K. Williams, S. MacDermott, G. Ridley, E.J. Glasson, and J.A. Wray. The prevalence of autism in Australia. Can it be established for existing data? *Journal of Paediatrics and Child Health*, 44(9):504–510, 2008.

[3] *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association, 4th edition, 2002.

[4] C. Lord, M. Rutter, P.C. Dilavore, and S. Risi. *Autism Diagnostic Observation Schedule*. Western Psychological Services, 2002.

[5] M. Rutter, A. LeCouteur, and C. Lord. *The Autism Diagnostic interview, Revised (ADI-R)*. Western Psychological Services, 2003.

[6] C.F. Norbury and A. Sparks. Difference or disorder? Cultural issues in understanding neurodevelopmental disorders. *Developmental Psychology*, 49(1):45–58, 2013.

[7] A. Klin, J. Lang, V. Chicchetti, and F.R. Volkmar. Interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: results of the DSM-IV autism field trial. *Journal of Autism and Developmental Disorders*, 30(2):163–167, 2000.

[8] D. Feil-Seifer and M.J. Mataric. Socially Assistive Robotics. *Robotics Automation Magazine, IEEE*, 18(1):24–31, 2011.

[9] I. Iacono, H. Lehmann, P. Marti, B. Robins, and K. Dautenhahn. Robots as social mediators for children with Autism - A preliminary analysis comparing two different robotic platforms. In *Development and Learning (ICDL), 2011 IEEE International Conference on*, volume 2, pages 1–6, 2011.

[10] Changchun Liu, K. Conn, N. Sarkar, and W. Stone. Online affect detection and robot behavior adaptation for intervention of children with autism. *Robotics, IEEE Transactions on*, 24(4):883–896, 2008.

[11] G. Pioggia, R. Igliozzi, 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.

[12] B. Scassellati, Henny A., and M. Mataric. Robots for use in autism research. *Annual Review of Biomedical Engineering*, 14(1):275–294, 2012.

[13] B. Scassellati. How social robots will help us to diagnose, treat, and understand autism. In *Robotics Research*, volume 28 of *Springer Tracts in Advanced Robotics*, pages 552–563. Springer Berlin Heidelberg, 2007.

[14] B. Scassellati. Quantitative metrics of social response for autism diagnosis. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 585–590, 2005.

[15] S. Narayanan and P.G. Georgiou. Behavioral Signal Processing: Deriving human behavioral informatics from speech and language. *Proceedings of the IEEE*, 101(5):1203–1233, 2013.

[16] Aldebaran Robotics. *Nao Software Documentation*, v1.14.5 edition, 2013. https://community.aldebaran-robotics.com/doc.

[17] E.T. Bekele, U. Lahiri, A.R. Swanson, J.A. Crittendon, Z.E. Warren, and N. Sarkar. A step towards developing adaptive robot-mediated intervention architecture (aria) for children with autism. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 21(2):289–299, March 2013.

[18] Antonis A. Argyros and Manolis I.A. Lourakis. Real-time tracking of multiple skin-colored objects with a possibly moving camera. In Tom Pajdla and Ji Matas, editors, *Computer Vision - ECCV 2004*, volume 3023 of *Lecture Notes in Computer Science*, pages 368–379. Springer Berlin Heidelberg, 2004.