



Personalized Robot Interventions for Autistic Children: An Automated Methodology for Attention Assessment

Fady Alnajjar^{1,2} · Massimiliano Cappuccio^{1,3} · Abdulrahman Renawi¹ · Omar Mubin⁴ · Chu Kiong Loo⁵

Accepted: 25 February 2020 / Published online: 6 March 2020
© Springer Nature B.V. 2020

Abstract

We propose a robot-mediated therapy and assessment system for children with autism spectrum disorder (ASD) of mild to moderate severity and minimal verbal capabilities. The objectives of the robot interaction sessions is to improve the academic capabilities of ASD patients by increasing the length and the quality of their attention. The system uses a NAO robot and an added mobile display to present emotional cues and solicit appropriate emotional responses. The interaction is semi-autonomous with minimal human intervention. Interaction occurs within an adaptive dynamic scenario composed of 13 sections. The scenario allows adaptive customization based on the attention score history of each patient. The attention score is autonomously generated by the system and depends on face attention and joint attention cues and sound responses. The scoring system allows us to prove that the customized interaction system increases the engagement and attention capabilities of ASD patients. After performing a pilot study, involving 6 ASD children, out of a total of 11 considered in the clinical setup, we conducted a long-term study. This study empirically proves that the proposed assessment system represents the attention state of the patient with 82.4% accuracy.

Keywords Autism · Robotics · Assessment · Attention · Therapy

1 Introduction

Social robotics research is concerned with the development of robotic companions. The existing social robots can assist people in simple everyday tasks and provide entertainment through basic forms of verbal and behavioral communication. In addition to serving and amusing, could robots also be used to improve the social abilities of humans, training

their capacity for attention and interaction? This question is related to a much broader issue: is it possible for an individual to learn from robots to establish richer relationships with other humans, fulfilling their inherent empathic potential? These questions are particularly pressing for the individuals who are prevented by clinical factors from establishing typical social relations with others—for example autistic subjects who experience social attention as effortful and potentially upsetting.

As we believe that these questions can be answered positively, we developed and tested a robot specifically designed to train the social attention abilities of children diagnosed with autism spectrum disorder (ASD). Our research was motivated by the awareness that social interaction and communication skills significantly impact human development, learning, and well-being. The lack of these elements can hinder individuals from being successfully integrated into a complex society. Children with ASD are prevented from developing these social skills and the ability to respond to them [1]. ASD symptoms are usually exhibited at early ages. Neurological accounts often highlight that parts of the social brain systems of ASD children are hyporesponsive to social stimuli and gaze cues. This hyposensitivity is the cause of

✉ Fady Alnajjar
fady.alnajjar@uaeu.ac.ae

¹ College of Information Technology, United Arab Emirates University, 15551 Al Ain, United Arab Emirates

² Intelligent Behavior Control Unit, RIKEN Center for Brain Science (CBS), Nagoya, Japan

³ Values in Defense and Security Technology, School of Engineering and Information Technology, University of New South Wales at Australian Defense Force Academy, Canberra, ACT, Australia

⁴ School of Computer, Data and Mathematical Sciences, Western Sydney University, Sydney, Australia

⁵ Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia

their unresponsiveness to the social signals of other individuals and the related difficulties in perceiving the eyes of other people as socially salient, which is why ASD children rarely establish eye contact [2]. Avoiding eye contact leads to impaired social attention [3] and, hence, difficulties in communication and interaction with others, which may result in severe academic and social problems. Owing to the continuous increase in the number of children diagnosed with ASD worldwide, the growing related social costs, and the absence of universally established diagnostic and therapeutic protocols, effective treatment of ASD is considered a public health emergency and an open research question [4].

Traditional ASD assessment focuses on multiple parameters, such as social communication and interactions, as well as restricted and repetitive, inflexible behaviors [5]. Attention (manifested through bodily cues like eye contact and interest in a voice, etc.) is one of the most commonly considered factors in the assessment of ASD through the identification of its symptoms [6]. Traditionally, assessment can only be performed only by licensed experts. It is demanding in terms of time and effort as it requires a complete history of symptoms to be obtained through systematic interviews with parents and patients. Moreover, although early intervention is a key factor for treatment, the younger the patient, the harder it is to accomplish the assessment [5]. Assessment is typically performed at the time of enrollment and repeated during long-term therapy to measure progress. If assessments could be repeated more frequently to trace the severity and development of symptoms [6], then therapists could implement more efficacious diagnostic and therapeutic options, and patients and their parents would be better motivated to withstand long, repetitive therapy sessions. This approach, of course, requires proportional commitments of time and effort by trained experts.

Currently, interventions for children with ASD primarily involve various methods for teaching patients the basic social skills through repetitive behavioral training [7]. During intervention, a therapist may utilize several tools and stimuli, such as toys and game-like activities, depending on the interests of the child and the skills to be tested [8]. One of the most important preconditions for the success of these interventions is the knowledge and availability of the kinds of stimuli that usually draw the attention of the child and increase his or her engagement over time, creating more favorable conditions for teaching social skills. Due to the heterogeneity of ASD symptoms among children, the intervention strategies and stimuli that are valid for one child cannot necessarily be applied to another. Individual customization may make intervention design challenging for therapists. Professional therapists certainly have the training and scientific knowledge necessary to promote the development of social skills in ASD patients. Therapists can adjust their behaviors in real

time to maximize the benefits to the patients. Two aspects of this process, however, remain problematic.

First, like any other human, a therapist is an inherently complex interactant. Their interaction with the patient inevitably involves the social information conveyed through facial expressions, gestures, posture, personality, attitude, etc. These irreducible complexities are not handled with equal ease by all ASD patients, and in most cases, they represent an inevitable barrier, probably due to the difficulty patients have processing high volumes of social information [9].

Second, the adaptability of the therapist to the cognitive and emotional characteristics of patients is limited by their number; the diversity of their cultural and personal backgrounds; and their motivation, language, and clinical condition. Challenges related to cultural and linguistic diversity are prominent in multi-ethnic countries such as the UAE, where expatriates of Bengali, Malayalam, Somali, Tagalog, Tamil, Hindi, and Urdu background (i.e., expatriates whose mother language is neither English nor Arabic, the two official languages of the country) represent more than 88% of the population [10]. As each human therapist inevitably has his or her own linguistic and cultural background, it might prove difficult at times to preserve adaptivity and intervention customization for all patients, due to the inherent difficulty of implementing the same diagnostic approaches and intervention protocols irrespective of the individual differences.

The developments in artificial intelligence and robotic technologies have proven promising for stimulating interactions with ASD children [11, 12] and performing more frequent assessments [13]. Robots fill the gap between conventional human therapy and child toys [14] and can perform endless repetitions without boredom, eliminating the concerns to training intensification. Anthropomorphic robots are designed to reproduce human features, behaviors, and emotions while simplifying their informational complexity, thereby reducing the cognitive and emotional burden and decreasing the possible stress for the patient. Consequently, it is believed that robots can improve the quality and length of engagement during social interaction and increase the possibility of stimulating the patients' social and cognitive abilities [15]. Crucially, robots can also automatically conduct score-like assessments [13, 16]. Robots continuously collect information that is relevant to diagnosis, which is useful for building retrievable databases of patient assessments and interaction histories. Such databases can be used to guide therapists in the personalization of their interactions and assessments. Compared with traditional assessments, the assessment methods augmented through robotic means allow for more comprehensive and articulate tracking of greater numbers of patients with more heterogeneous individual situations and needs [17]. Such personalization can be applied locally, within the same team or clinic, or globally as inte-

grated clinical dataset sharable by professionals in different places through interoperable systems [18].

Current robotics techniques, however, still suffer from various limitations. Most preprogrammed systems have fixed behavior (i.e., they are unable to autonomously perform adaptable closed-loop interactions), are not tailored to the individual needs of patients, and cannot keep track of their recovery progress [19]. For these reasons, semi-autonomous and adaptive robots are greatly needed to recognize the behavioral cues of children and respond accordingly [20]. On the other hand, semi-autonomous adaptive systems, especially complex systems, need high-performance hardware, such as GPUs chips [18], to process real-time data and update interactions. Moreover, such fully autonomous and complex robots and systems are not yet reliable outside controlled research setups [21–23].

In the present study, a simple autonomous assessment system based on attention cues was created and deployed, combined with an enhanced adaptive semi-autonomous interaction system based on patient interests. Both systems were implemented in an interactive autonomous humanoid robot. The function of the robot was to increase the attention and engagement levels of the patients during interactive sessions. Increasing the attentional and interactive capabilities of ASD patients has the potential to enhance their academic functioning by gradually habituating them to socially interactive sessions of increasing length. The proposed approach utilizes simplified hardware with some upgrades to the onboard hardware of the robot to target multiple interaction and attention cues simultaneously. This technique can serve as a useful form of ASD intervention to facilitate adaptive interactions with patients based on their status while involving minimal subjective biases. In this study, we empirically tested the proposed system on a group of ASD children. As shown in our report, the empirical results are promising.

The paper is organized as follows: Sects. 2 and 3 provide thorough descriptions of the proposed assessment and therapy system, including the experimental setup and procedures. Sections 4 and 5 analytically present the study results and discuss their implications in the context of the research objectives. Section 6 concisely restates the study methods and main findings and addresses the scope for future research.

2 Methodology

2.1 Participants

The participants in the study were 11 male patients diagnosed with ASD of mild to moderate severity (Childhood Autism Rating Scale [24]—which is essentially a diagnostic scale; CARS2 of 30–36.5) who were under the age of 16 years, with a mean age of 9.03 (± 2.56) years. Only patients with ver-



Fig. 1 NAO robot with a mobile phone attached to a chest holder to show the Emotions Selector mobile application

bal response capabilities were considered, as the participants were preliminarily asked to understand verbal communication and respond with yes or no at least. All patients were recruited from Al Ain Hospital, Al Ain city, UAE. This study was approved by the Social Sciences Research Ethical Committee of UAE University (Al Ain, UAE) and Al Ain Hospital (Al Ain UAE). The parents of the children were asked to sign parental consent letters.

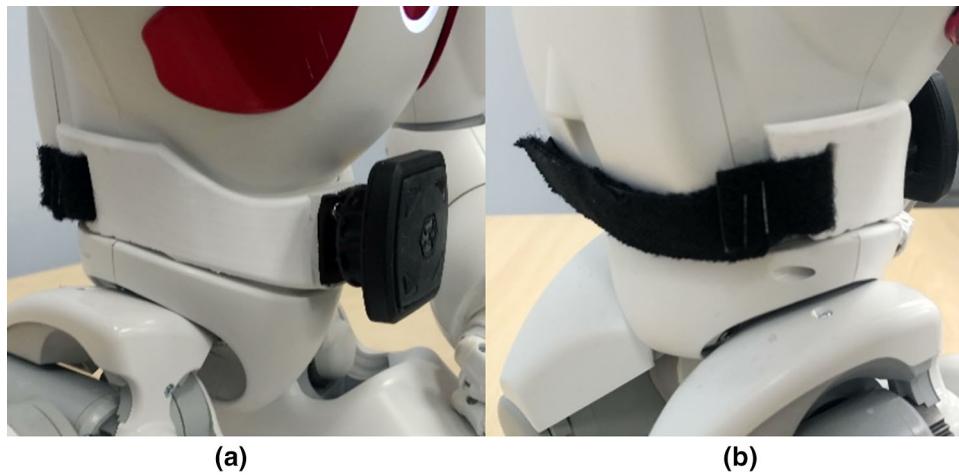
2.2 System Design

2.2.1 NAO Robot with a Chest-Mounted Mobile Phone

The proposed approach to robot intervention in autism therapy and diagnosis uses a specifically designed NAO humanoid robot (Aldebaran/Softbank NAO robot) with an additional display on its chest to show facial features. The modified robot is shown in Fig. 1.

We designed a custom NAO robot chest-mount to firmly hold any object at the lower chest region [25], as shown in Fig. 2. The added weight barely affects the center of mass of the robot body, hence the changes in the moment of inertia are negligible. This method helps maintain the equilibrium capability and mobility of the robot, so that built-in interaction features can be employed to draw attention throughout the diagnosis and therapy processes. A mobile phone is used as the added display. To carry the mobile phone, a customized

Fig. 2 **a** Custom-designed NAO chest holder side view, showing the attached standard mobile holder. **b** NAO chest holder rear view, showing the added Velcro strap for fastening



mobile holder was used to fit the phone on the designed chest mount. This model of chest mount was fully designed and 3D printed in our lab. It is a rigid multipurpose holder with four slots for tightening screws. The holder is fastened from the back side using Velcro straps for greater rigidity. This holder can be used to carry a mobile, camera, laser sensor, or any other helpful item on the chest of the NAO robot for navigation or interaction purposes. We have made the chest-holder model files available on the GrabCad website for research use: <https://grabcad.com/library/nao-robot-chest-holder-chest-mount-1>.

The mobile phone displays emotions using a custom designed mobile application called “Emotions Selector”. Emotions Selector was specially designed with the aid of App Inventor [26]. It accepts control messages to switch between emotion photos as a single-character data from the computer via a TCP socket connection over WiFi. The application is simple and easy to use because it employs non-technical operators. It only asks for the IP address of the computer to establish a connection with the mobile device.

2.2.2 Attention Assessment System

Developing accurate and versatile methods for detecting negative child behaviors is important for minimizing distress and facilitating effective human–robot interactions. This paper presents a novel simple numerical diagnostic assessment method that does not require any external camera or monitoring equipment. It utilizes the naoqi image processing capabilities of the NAO robot and the capabilities of the mobile phone to detect patient attention cues and generate numerical measures. All attention cues are detected and updated in every iteration of the system algorithm, running at average speed of 1 Hz, and each score is updated based on certain criteria as discussed below.

Attention Score The camera of the mobile phone was configured as an IP camera using the IP camera mobile appli-

cation utilized for face detection to produce and accumulate an attention score. Haar classifiers [27] help detect the faces of patients from the image frames received from the mobile camera. If the patient is facing the front of the robot body, where the mobile phone is attached, the face of the patient is detected, therefore it is assumed that he or she is paying attention to the robot. Thus, the assessment system increments the attention score by 5 points. If the face of the patient is not facing the robot, then his or her face is not detected and the assessment system decrements the attention score by 1 point. The operator can change the preset increment and decrement values so that each patient has individual settings. This score value is updated in each iteration of the algorithm continuously. This attention score facilitates comparison of the results from multiple interaction sessions for the same patient. Hence, the increment and decrement parameters are not changed between sessions for the same patient.

Joint Attention Score The joint attention score measures the extent to which the patient’s responses are synchronized with the robot’s motoric actions and requests. For example, it detects whether the patient is looking toward the box that the robot is talking about and pointing towards. For simplicity, the head direction of the patient is detected by identifying his or her right and left ears. The camera detects the left or right ear of the patient depending on which way the face of the patient is turned during the interaction. The detection algorithm uses the same video frames that are employed for face detection, which are retrieved from IP mobile camera. The joint attention score is incremented by 15 points if the head direction of the patient matches the expected direction in response to actions of the robot. Otherwise, if the patient does not respond as expected, the joint attention score can be decremented. The decrement value was set to zero in the following experiments. Joint attention score is updated only when a joint attention request is triggered in a scenario part. Since the attention score is designed to report on the general

attention state of the patient, it is incremented by 10 points and decremented by zero points at joint attention events.

Sound Response A mobile application called WO Mic turns the mobile phone microphone into an IP microphone, and client software on the computer connects the WO Mic as a sound input for a sound response module algorithm. This technique allows the operator to have a microphone monitoring the interaction room without affecting the sound response module functionality. The sound response algorithm starts calculating the delay time at the instant when the NAO robot finishes asking the question and stops at the instant a sound peak is heard, which occurs when the patient starts speaking. The recorded delay time is saved as the sound response time and used in Eq. (1) to calculate the increment (Incr) in the attention score instances as a function of the detected response time (RespTime), preset maximum score for the response time parameter (MaxRespScore), and Response-time-out value:

$$Incr = MaxRespScore - \frac{MaxRespScore}{Response-time-out} \times RespTime. \quad (1)$$

From the equation, the faster the sound response, the bigger the increment. The lowest response time, ideally zero, generates the maximum increment value, which is a parameter that can be preset. Sound response is updated only when sound response request is triggered in a scenario part. The value of MaxRespScore used in the subsequent experiments was 30 points.

Some sound response parameters must be preset to account for the nominal speaking volume of the patient and room noise level. However, sessions with patients would ideally be conducted in a quiet room without interruptions or noise. The sound response parameters are as follows:

- *Threshold-peak-level* the sound peak predefined threshold level, if detected by microphone, initiates sound response module and starts counting the response time in seconds. The module assumes that the patient has started speaking at that instant. This parameter was set to 2000 Hz in most cases, and adjustable by the operator based on the experimental environment and patients voice level.
- *Response-time-out* maximum time response value allowed. If the response time exceeds this value and continues counting, the module resets and ignores the rest. Operators should assign high values for patients with some speech deficiencies. This quantity was set to 2.5 s in our case.
- *Text-is-done-time-out* maximum waiting time for the robot to acknowledge that the patient has finished speaking. The module resets if the robot fails to acknowledge that the patient has finished speaking and this timeout value is

reached. This timeout value check catches the robot timeouts, instants when NAO robot hangs and fails to send text-is-done feedback, and it should be set longer than the speech time of the longest expected text. This parameter was set to 5 s in our case.

- *Sync-delay* a small delay time to compensate for the delay time of the employed mobile phone microphone, mainly due to WiFi connection delay, to sync it with sound response module timing. This quantity can be found using an external observer microphone connected to an external computer. First, the employed mobile phone, the system computer speaker and the external microphone are placed very close to each other. Second, the system computer speaker is set to echo the sound received from the mobile phone microphone over WiFi, mobile phone position may be slightly changed to avoid a sound loop. Third, the external microphone starts recording and a peak sound is made, a single table knock for instance. Finally, the recorded sound file on the external computer is used to plot the sound recorded and determine the delay time between the sound peak made and its echo from the speaker of the system computer. This parameter was found as 0.25 s in our case.

Emotion Detection The assessment system estimates the emotional state of the patient by detecting facial features. The results of the emotion detection module onboard the NAO robot were used to avoid an additional processing load on the system computer. Retrieving these results was timed in the attention assessment system loop to sync logged emotion estimation results with attention assessment score. This sync method avoids any effect on results' analysis. Emotion detection uses the onboard NAO head camera, and its results were accessed using naoqi Python module. The system supports the detection of five possible emotions (happy, sad, angry, surprised, and neutral) by estimating the probability of each emotion at each instant. The system chooses the emotion with the highest probability, logs it, and represents it in the final results plot. Emotions are represented as color bands, with a different color for each emotion.

2.2.3 Adaptive Dynamic Scenario and Weights

A dynamic interaction scenario is employed, which depends on the real-time scores of the assessment system to tune the interactions based on the previous interaction results. This method helps maintain interaction and assessment sessions by reducing human intervention in the scenario flow. Moreover, it allows the operator to have control, although limited, by interfering with operator-defined robot responses added to the interaction in real time to ensure the meaningfulness and interactivity of the scenario.

Table 1 Scenario section titles

Label	S1	S2	S3	S4	S5	S6	S7
Title	General questions	Playing dialogue	Math questions	Colors learning	Emotions recognition	Food dialogue	Birthday singing
Label	S8	S9	S10	S11	S12	S13	
Title	Driving questions	Repeat words	Who is this person?	See that box	Football playing	Physical exercises	

The scenario is divided into sections, where each section contains interaction dialogues, motions, and plays for a certain topic, such as greetings and getting to know each other, entertainment games and songs, and conversational questions and requests. A typical scenario starts with a greeting phrase such as “Hi” or “Hello.” The greeting is followed by interaction responses and questions, such as the following (see also Table 1):

- (S1) General questions such as “Where have you been?”, “How are you?”, “Do you know my name?”, “How old are you?”
- (S2) Playing responses such as “Can we play together?”.
- (S3) Math related questions such as: “Can you count?”, “How much is 1 plus 1?”.
- (S4) Color teaching questions such as: “What color do you like?”, “What is my color?”, “What is your shirt color?”.
- (S5) Emotions teaching such as: “Are you happy?”, “I feel sad”, “The face on my screen is excited”.
- (S6) Food related responses such as: “Do you like chocolate?”, “I am hungry”, “What did you eat today?”.
- (S7) Birthday singing activity such as: “When is your birthday?”, “Lets sing it together!”.
- (S8) Driving mimicking such as: “Do you drive?”, “I drive very good”, “See how I drive, lets race!”.
- (S9) Learning and repeating words such as: “I want you to repeat after me”, “Say small”.
- (S10) Who is this questions for joint attention initiation such as: “Do you know who this on the right is?”.
- (S11) See that box responses for joint attention initiation such as: “Do you see that box?”, “What is inside it?”, “Can you bring it here?”.
- (S12) Football playing mimicking such as: “Do you play football?”, “I know how to shoot!”.
- (S13) Stimulating physical exercises such as: “Stand up”, “Raise your right arm, like this”, “Where is your right ear?”, “Can you jump?”, “Raise your left hand”.

The scenario provides encouragement phrases as well such as “Great!”, “You are awesome”, “Wow!”, “No, try again”, and “I can’t hear you, talk louder!”. The mobile screen shows an image of an emotion (happy, surprised, excited, sad, or angry) that is relevant to each scenario phrase, presented as the robot’s emotional state.

Each section in the scenario has a numerical weight that is updated whenever the attention score is updated. This value increases or decreases by 0.2 points depending on whether the attention score increases ($\text{attention_trend} = +1$) or decreases ($\text{attention_trend} = -1$) at the time of using the corresponding section, as defined in Eq. (2):

$$W_{\text{new}} = W_{\text{old}} + (0.2 \times \text{attention_trend}) \quad (2)$$

The section weights are stored individually and used in the following session to determine the number of phrases to use from each section, by multiplying the weight times the default (first) number of phrases to use. The weights in the first session are the same for all patients, and they are the default values. The minimum weight is zero, sections are not used, and the maximum value is the size of the section (number of phrases defined in the section). If the weight reaches the section size value, it means that the patient is interested in the section, and the operator is prompted to add more phrases to the section for the subsequent interactions.

The scenario is defined as an Excel file with each section as a separate sheet. Scenarios are easy to edit, and they are updated as soon as a session starts. The scenario sections phrases definition includes the choice to call for sound responses, joint attention requests, and robot movements, in addition to text to be said by the robot. The section titles are shown in Table 1.

2.2.4 Control Interface

We reduced the need for human intervention to make robot-aided sessions more easily replicable in clinical and domestic environments, as reproducibility is essential to make the system usable by therapists and parents without technical supervision.

The interface allows the operator to control the robot as shown in Fig. 3a. It is designed to be operator friendly and simple for operators with no programming experience. Pre-configuration is performed once by setting the parameters in the “config.ini” file. The operator should select the patient name (from the history of children enrolled) or input a new name to be registered and select the language of interaction (English or Arabic in the current version) and which camera to use (an IP camera or a local camera, if a wired camera is

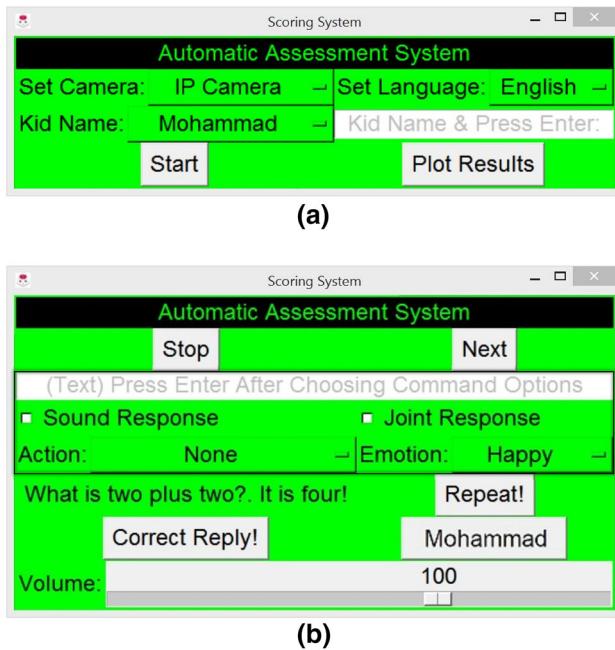
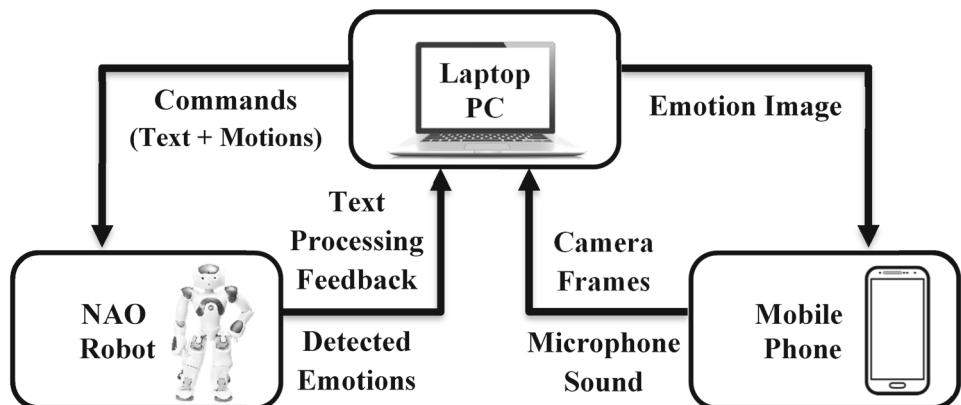


Fig. 3 **a** Scoring system control interface in idle state; no interaction running. The “Plot Results” option plots recent interaction results. **b** Interface during interaction. “Stop” stops the interaction system then saves and plots the results. “Next” goes to the next scenario part. Highlighted box contains operator-defined response definition tools: “(Text)” is to write a text to be said by the robot. All written texts here are saved and can be selected later from a selection lists when clicking into (Text). “Sound Response” is a check box to trigger sound response or not, “Joint Response” is check box to trigger joint response or not. “Action” is to select action to be taken by the robot from a predefined drop-down list of available actions and motions (move hand, point to right, point to left, shoot ball, sing birthday, etc.). “Emotion” is to select emotion photo to be shown by Emotions Selector mobile application from a predefined drop-down list of available emotion photos (Happy, Sad, Neutral, Excited, Crying, Surprised). “Repeat” repeats the last scenario part, which phrase text is shown to the left of “Repeat” button. “Correct Reply!” triggers a random praise phrase. “Mohammad”, or current participant name, triggers calling current participant by his/her name. “Volume:” a slide to simultaneously change the robot speaking volume. (Color figure online)

Fig. 4 Control system block diagram



connected). The system generates a directory with the name of the child in the results folder for the individual data. Then, the operator starts the interaction by pressing “start,” and the interface changes as shown in Fig. 3b.

During the interaction, the operator can go to the next scenario part, repeat the last part, call the patient by his or her name, change the speaking volume, trigger a praise phrase, and define and trigger an operator-defined response. The operator-defined responses are saved as a personal individualized scenario.

With the systems described above working together, the control system block diagram can be represented as depicted in Fig. 4. The system parts (laptop PC, NAO robot, and mobile phone) communicate via a WiFi connection. The software is explained in the interaction and assessment system flowchart in Fig. 5. The number of phrases to use from each section is called ‘section load’. Initially, all section loads are set to 3 phrases. Later, after each session, each section load is multiplied by the corresponding updated section weight to update the number of phrases to be used.

3 Experimental Setup

Experiments were conducted in Al Ain Hospital rehabilitation center in collaboration with autism therapists. The operator, who is not a therapist, sat in a separate control room monitoring the robot-child interaction setup remotely and controlling the assessment system. The robot was placed in the therapy and interaction room standing on a table so that it was at the same level as the patient sitting on a chair, as shown in Fig. 6a. The setup was composed of the robot, patient, therapist to intervene in potentially harmful situations, and parent (if available), as shown in Fig. 6b. A camera recorded the interaction sessions, so that the parents could watch the sessions if they were not available at the session time.

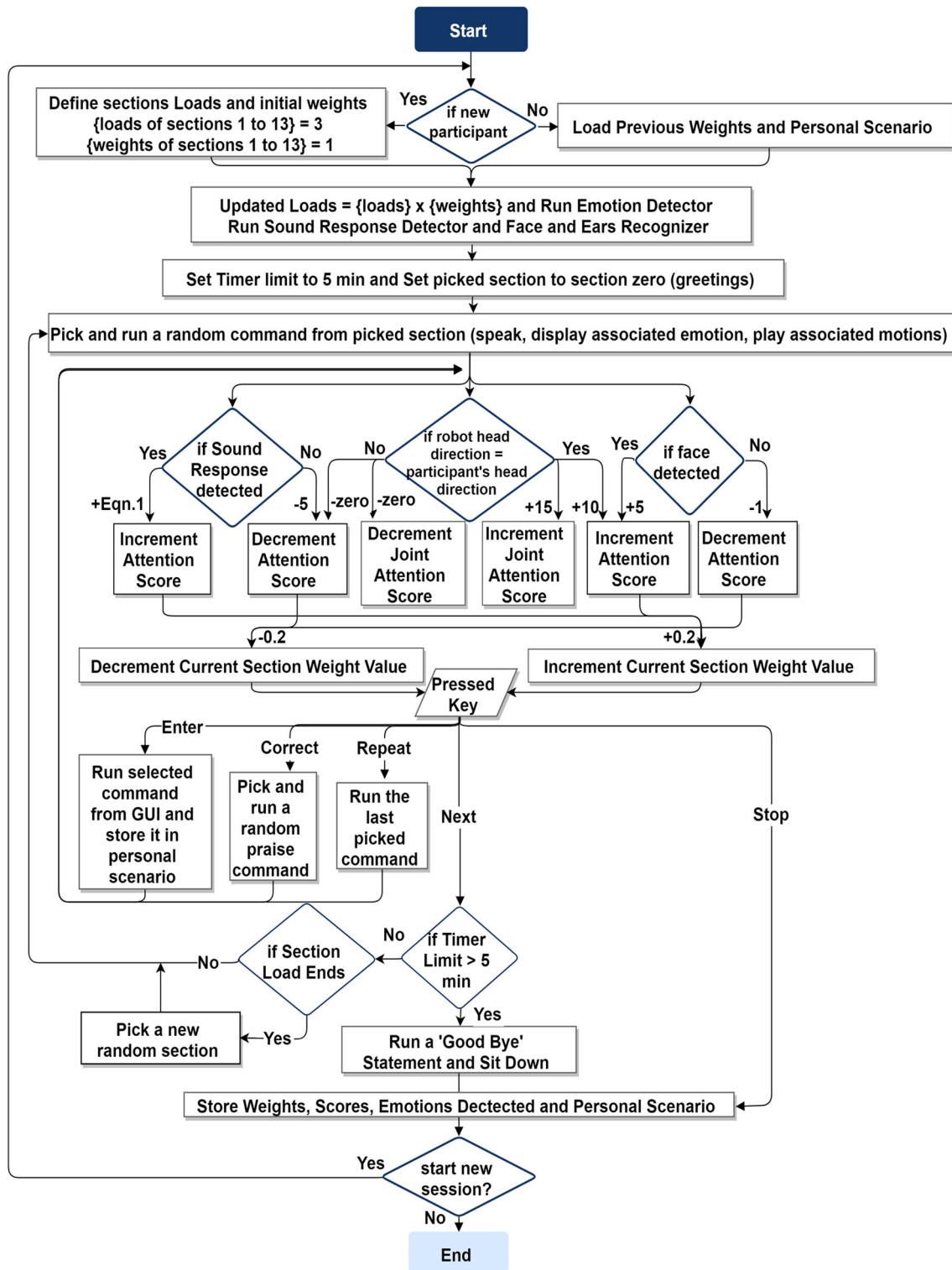


Fig. 5 Interaction and assessment system flow chart

3.1 Experimental Protocols

The robot was introduced to the child as a new friend. The experiment started by asking the child to sit on a chair facing

the NAO robot and talk with it. The therapist was asked to fill out the therapist assessment sheet during the experiment if possible, or later at the time of watching the video. At the end of the session, the parent was asked to fill the parent

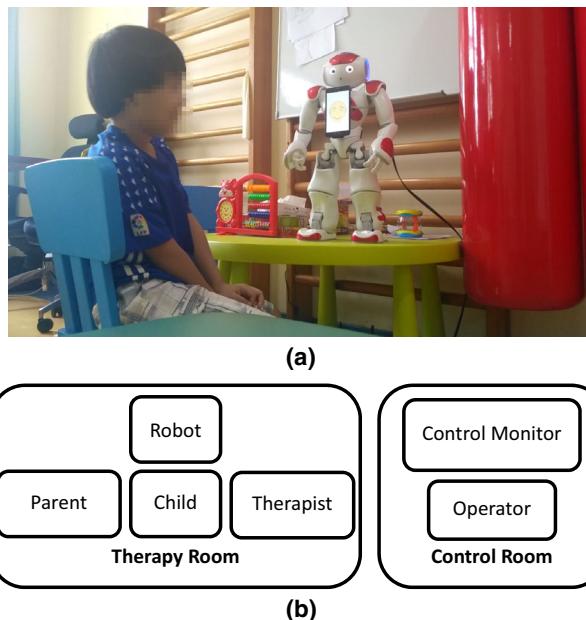


Fig. 6 **a** Therapy and interaction room setup. **b** Schematic diagram of experimental setup

feedback form. The therapist assessment sheet was an attention scale from 0 to 100. The therapist rated the attention of the child every 30 s and marked the value on the assessment sheet; subsequently, the values on this sheet were compared to the assessment system scores. The parent feedback form had two scale questions: “How do you rate your child interaction with the robot today?” and “How do you rate your child interaction at home?,” as well as a third Yes or No question: “Do you think robot therapy is helpful?” and a blank space with the heading “Any Comments.” The parent feedback was important to rate the interactions of the children with the robot compared to their interactions with their families, and to detect whether interacting with the robot influenced their interactions at home. The robot interacted with the child for 5 min (specified time limit), after which the session ended. At

the end, the robot thanked the child, encouraged him or her to come again for subsequent sessions, and said “bye bye.”

3.2 Real-Time Score Visualization

The accumulated attention, joint attention, and sound response scores were presented in real time in a separate single compact plot, as shown in Fig. 7, to convey the level of interaction between the patient and the robot to the operator. The plot shows the sound response request instants and joint attention requests instants as well. Sound response request is generated when scenario phrase has sound response option. Joint attention request is generated when scenario phrase has joint attention option. The algorithm saved a copy of the detailed results at the end of each session in a.pickle file for processing. The scoring system GUI allowed the operator to plot the detailed results immediately after each interaction session.

4 Results

4.1 First Impression

Patients were included in a “one trial” experiment to assess their first impressions of the robot. Since adaptation depends on previous interaction history, all patients had no history at the first session, which means that the system had no adaptation capabilities yet. Sample score results generated after the session of one patient are shown in Figs. 8, 9 and 10.

The normalized attention scores in Fig. 8 show that the attention level increases over time, with minor drops representing gaze distractions. In addition, the joint attention level increases two times, representing two successful joint attention events, where the robot talks about and points towards an object to either the right or left and the patient is expected to follow. Moreover, the emotion color map shows several detected happy facial emotions, mainly at the beginning of

Fig. 7 Sample real-time plot showing the attention score (red line), joint attention score (blue line), sound response requests (yellow bars), i.e., time instants at which a sound response was expected, and joint attention requests (green bars), i.e., time instants at which a joint attention cue was expected. The white numbers above the red line represent accepted sound responses in units of seconds, which are less than the response-time-out parameter values. (Color figure online)

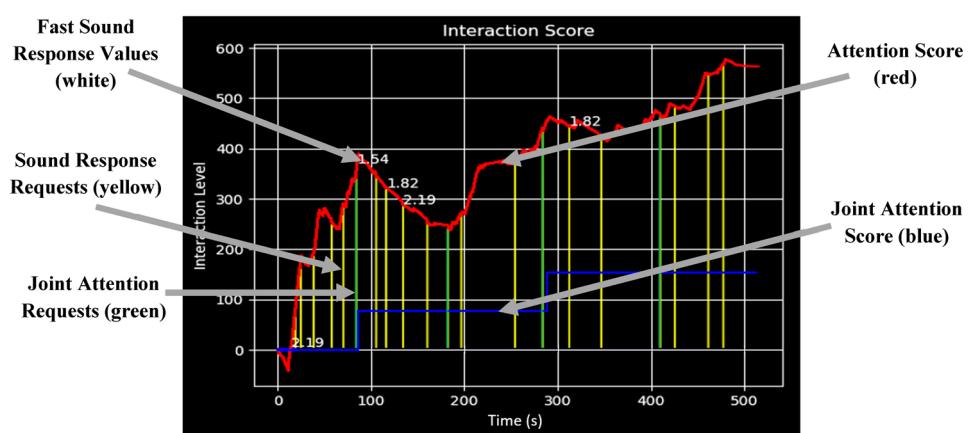


Fig. 8 Normalized attention (red) and joint-attention (blue) scores with emotion color bands. Scores were normalized by scaling data to have the values between 0 and 100, by dividing values by the maximum value and multiplying by 100. The attention scores and predicted emotions could be mapped for each time instant. An emotion result of “None” represents an instant at which emotion prediction is not possible. **a** Highlights a case of attention score drop with sad emotions, then joint attention score increase with happy emotions. **b** Highlights a case of attention score increases with surprise emotions. (Color figure online)

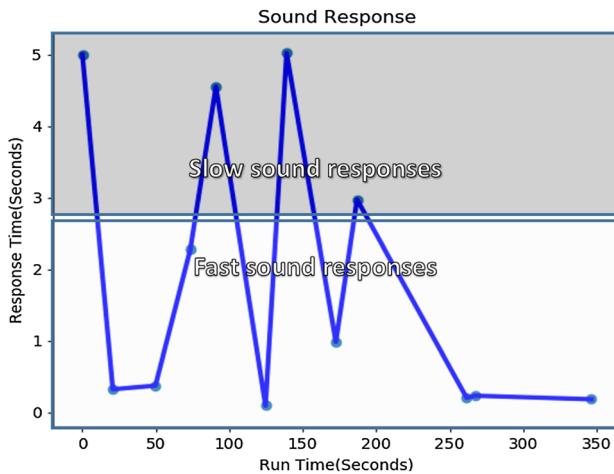
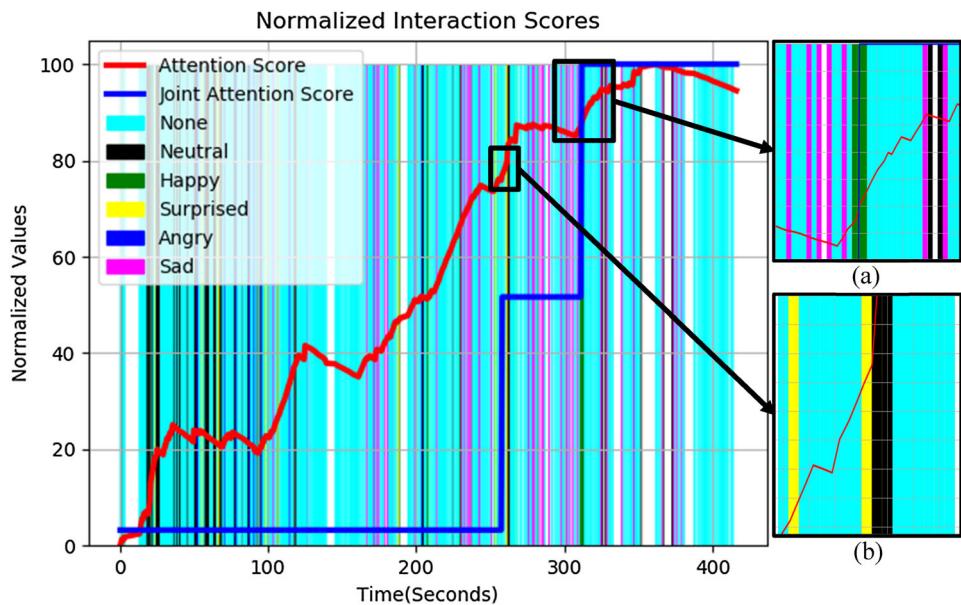


Fig. 9 All recorded sound response values. Only fast responses (equal to or less than the Response-time-out value) are plotted in real time. Response-time-out and several other parameters explained in Sect. 2 can be adjusted to suit the needs of each patient for better cue detection

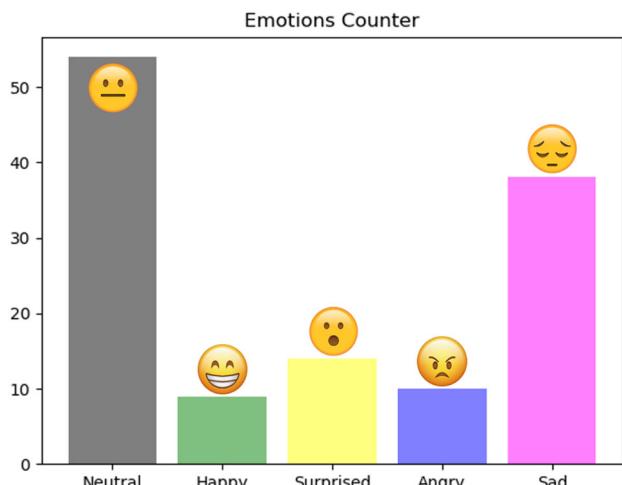


Fig. 10 Emotion counts, showing how many times each emotion was detected

interaction and at the instants of joint attention, as shown in Fig. 8a. Some sad and angry emotions were detected, where the patient could have been experiencing some discomfort. Thus, these emotions mainly occurred after drops as shown in Fig. 8a, where the patient was not looking at the robot and hence no emotion data were obtained, or during a disturbed gaze. Some surprise was detected during steadily increasing attention phases as shown in Fig. 8b, when the robot regained the attention of the patient using an attractive and surprising action.

The sound response values in Fig. 9 are divided into slow and fast sound responses based on the Response-time-out value. This patient had more fast sound responses than slow

sound responses (only four slow sound responses), which means that the patient was very responsive.

The emotion counter in Fig. 10 shows that the neutral emotion is the most dominant and that sadness in discomfort instants lasted for some time. It also shows that happiness was detected 10 times and surprise 15 times, corresponding to excitement instants.

The average attention score of all patient trials was found and is plotted in Fig. 11. The data show that most of the patients had positive first impressions and that their attention levels increased during the interaction. However, some patients did not follow this pattern, which explains the large standard deviations.

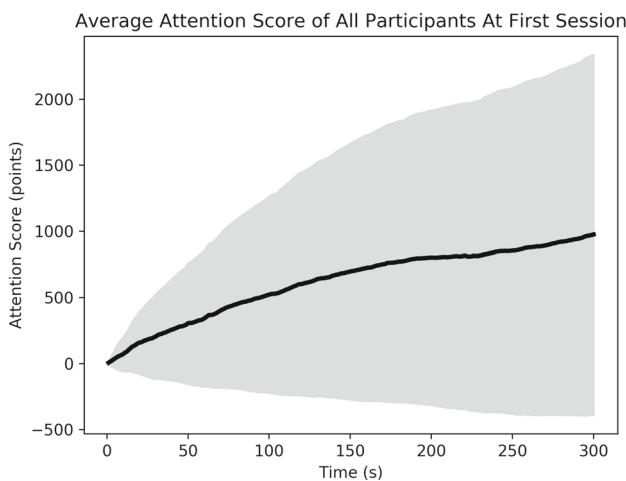


Fig. 11 Average attention scores of all patients in the first session

Fig. 12 Session attendance by week

	Patient No. 1	Patient No. 2	Patient No. 3	Patient No. 4	Patient No. 5	Patient No. 6	
	X	X	X	X	X	X	Week 1
	X	X	X	X	X	X	Week 2
	X	X	X	X	X	X	Week 3
	X	X	X	X	X	X	Week 4
							Week 5
							Week 6
							Week 7

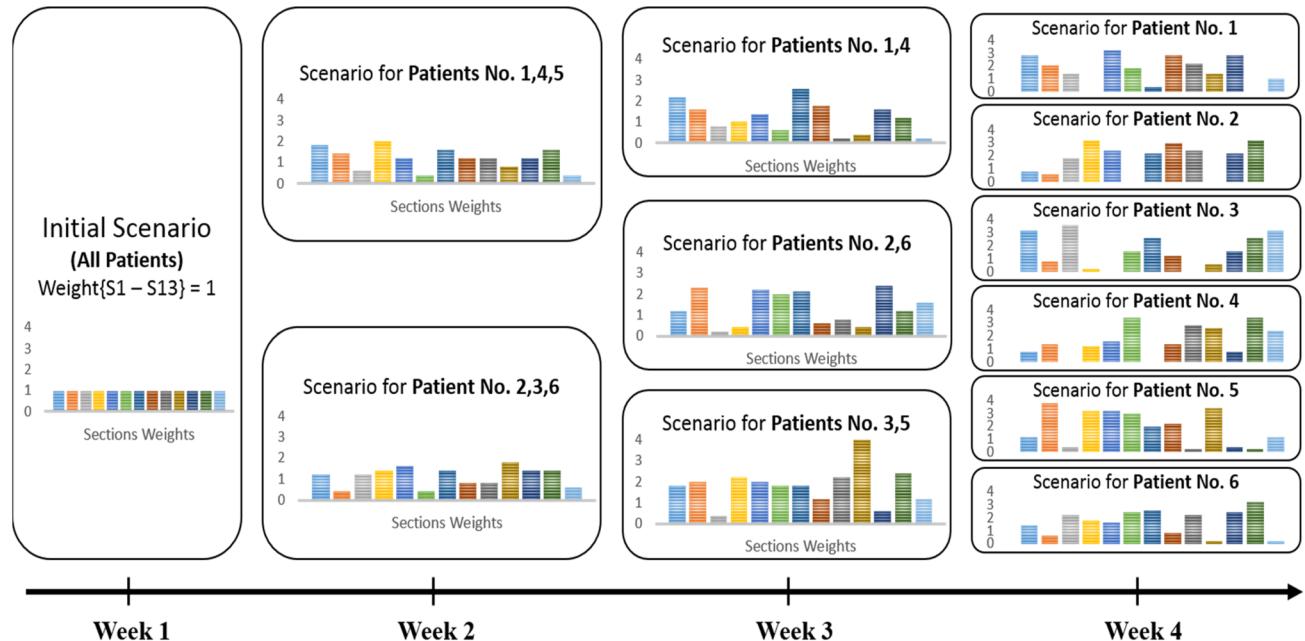


Fig. 13 Scenario flow and sections weight diagram. The weights change in each session based on the attention scores of the patients. Any “zero weight” section ceases to appear. The weight could not be zero from the first two sessions. S1–S13: section 1 to section 13

4.2 Interaction Progress

Six out the 11 patients were able to continue in a long-term progress study. Further experiments involving these six patients were performed during the following 7 weeks to examine their interaction progress and the variations in the obtained weights of the scenario sections.

The session attendance of the patients over the 7-week period is illustrated in Fig. 12. Some patients could not come to all the sessions. Although some patients had more sessions than others, each patient had at least four sessions.

The weight changes of the scenario sections over the first 4 weeks are shown in Fig. 13. The scenario was the same for all patients in the first session, so they have the same section weights. In the second week, the weights took two different sets of values. It took up to the fourth week for six different

sets of weights to emerge; thus, each patient had different personal weight values at this point.

The results of the empirical sessions for all patients are shown in Fig. 14, which compares the progress in terms of attention scores, therapist assessment, and parent feedback. The results show that all six participants portrayed a trend of increasing attention scores. However, the therapist and system assessment trends are similar for most of the patients.

5 Discussion

The average attention score in the first impression trials generally increases for 9 out of 11 patients from the beginning of interaction until the end of the session. The other two patients show decrease at some points. This finding demonstrates that, on average, patients had positive first impressions and could complete the 5 min of interaction. Since not all scenario sections were familiar to all patients and scenario personalization requires at least one previous interaction history to be active, attention discontinuities occurred at the times of unfamiliar scenario sections. Consequently, attention results varied among patients, and the average attention score had a large standard deviation.

The later sessions had increasing (patients 1, 2, and 3) or fluctuating (patients 4, 5, and 6) attention levels with some drops. These drops have several possible causes. The primary possible cause is hyperactivity, which the therapist explained as being due to the patient not taking energy-draining sessions for a long time or not taking the prescribed medicine. Another possible cause is the patient's interest in small scenario sections (i.e., those with few defined statements), which decreases gazing time. In this case, the system prompts the operator to resolve the issue and define more statements for the subsequent sessions in the sections of patient interest.

The long-term interaction progress results show that robot intervention in autism therapy is highly beneficial for autistic patients. They enjoy the treatment sessions and give enough attention to learn or enhance a skill in every session. Noticeable changes in patient behaviors and skills took a few weeks, until the patients became familiar with the robot voice and moves. Breaking the ice took one or two sessions on average, until the patients interacted with the robot freely and excitement reached a steady level.

The autistic patients built strong bonds with the robot as a friend who encouraged them to withstand the therapy time. One of the patients once brought some friends to meet the robot as a new friend. We observed that the therapists at the hospital went further by encouraging their patients to do energy-draining physical therapy exercises by offering a session with the robot as a reward. The therapists reported that the patients asked about the robot every day and looked forward to the day of the robot session.

The parent of patient 3 reported that the interaction of the patient at home increased noticeably after taking sessions with the robot. The parent commented:

He has improved verbally since he has been interacting with the robot... There has been an impressive progress with his communication and interaction at home. He has a long way to go. But each step has been a joyful success for us. Thank you.

Moreover, most parents reported that their children mimicked and repeated many of the robot responses at home, and some of the children asked their parents at home the same questions that the robot had asked them during the session. Five out of six parents believed that their children performed as well as at home or even better in some sessions (patients 2 and 5).

The therapist assessment shows an increasing trend, with all patients demonstrating an increased attention level over time. Four out of six drops in session attention progress levels detected by the therapist were detected by the assessment system as well. Overall, six mismatches, highlighted with red circles in Fig. 14, in the attention score were found between the system and therapist assessments out of 34 sessions (82.4% accuracy).

Emotion predictions are of great assistance in understanding the emotional states of children in different scenario sections. Mapping the predicted emotions using the attention score and associated scenario section enables understanding of the facial responses of patients to specific topics and could facilitate the detection of difficulties in emotional responsiveness and other similar autism characteristics.

The objective of this study was to develop a simple robot-mediated assessment and interaction technique to prepare such systems for long-term presence in autism rehabilitation centers. This system can be used by therapists and parents after short training. The proposed system is easily replicable due to its simplicity and ease of use, and can cope with a large number of patients simultaneously. In addition, it may play a complementary and coadjutant role in the therapy of the patients who cannot continue traditional treatment sessions or need sessions with a frequency higher than rehabilitation centers can offer due to a lack of therapists.

Because of the heterogeneity of ASD disorders, one predefined intervention scenario cannot possibly address the needs of all patients. The proposed adaptive dynamic scenario and weighting technique allows customized interactions for each child. Such adaptive techniques have proven effective in sustaining social engagement during long-term children–robot interactions [28]. The employed techniques maximize engagement, which is one of the strongest predictors of successful learning [29, 30], using a ludic mobile robot to stimulate social skills in ASD children. Moreover,

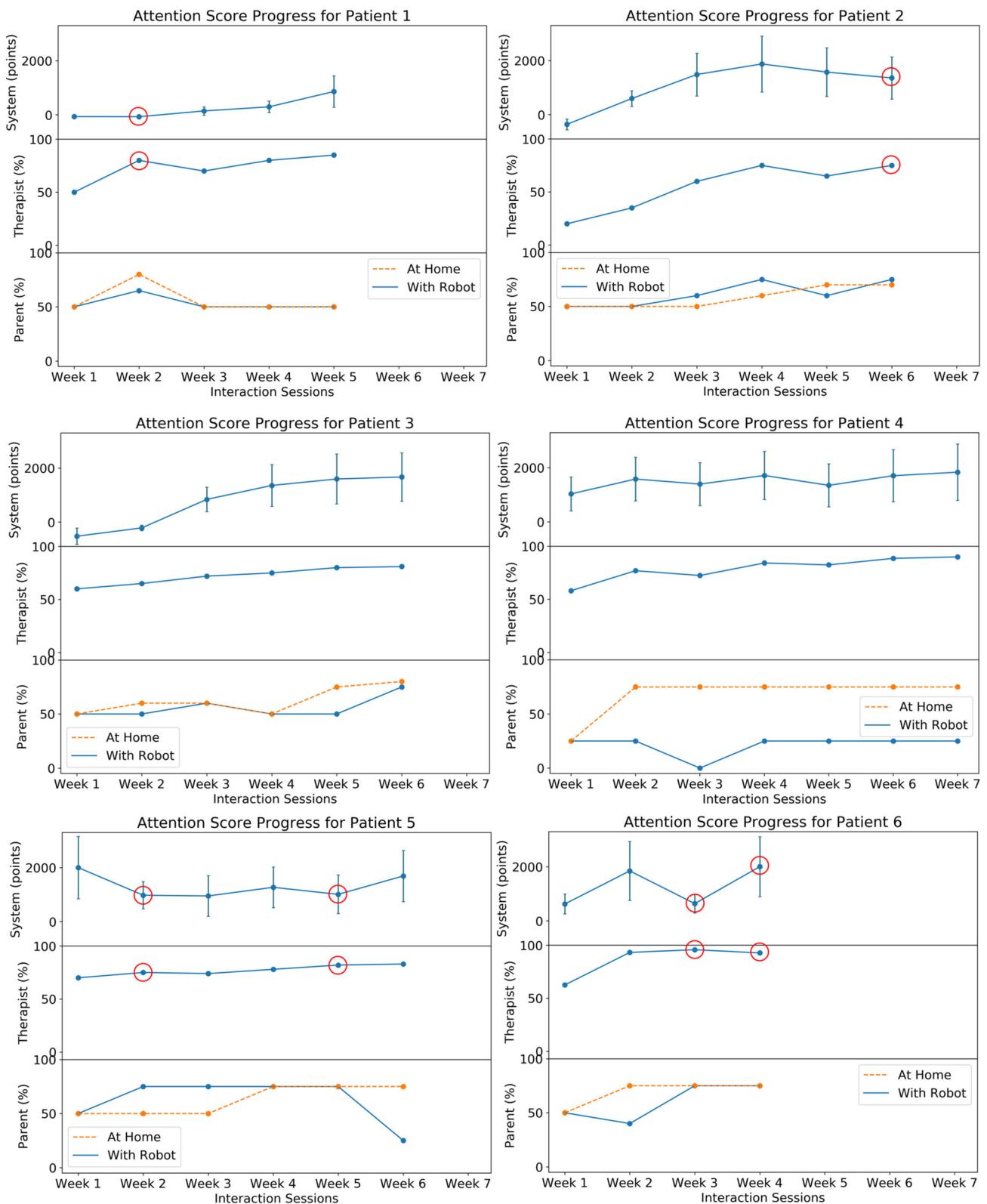


Fig. 14 Session results over several sessions for all patients, showing the average system attention scores (upper plots), the average therapist attention assessments (middle plots), and parent feedback when asked

about the interaction level after each session with the robot and at home (lower plots). Red circles show the mismatches between therapist and system results. (Color figure online)

the proposed system reduces the therapist's subjectivity in assessment and allows for early intervention.

The current study has several limitations that need to be highlighted with a view to address them through future developments and integrations. First, only one-on-one interaction is possible, to allow the child to focus on the robot only and to allow the system to capture the child's attention cues. Moreover, the study included a relatively small number of patients, and more results may be revealed when it is applied to more patients. Also, the proposed interaction system is only applicable to patients with moderate severity and who have at least minimal verbal response capabilities. Some children may become distracted by the mobile phone display on the robot, since they are used to playing games with such devices, which may lead to drops in attention score. This issue occurred only at the beginning of the early sessions where the children were exploring the robot features and it has been partly addressed by fixating the display in a way that it cannot be moved by the children.

6 Conclusions

An adaptive robot intervention system for ASD assessment and therapy was designed for clinical use and tested empirically on six ASD patients in an autism rehabilitation center. The results demonstrate that the proposed assessment system can accurately represent the attention levels of patients with mild to moderate ASD and simple verbal communication skills, matching over 80% of therapist assessments. The proposed adaptive, dynamic interaction system yielded remarkable improvements in the attention levels of most of the patients in long-term therapy.

Based on these outcomes, our hypothesis is that a properly designed robot intervention system can increase the attention levels of ASD children insofar as it enhances their engagement and, in so doing, helps them improve their communicative and social skills. Moreover, the same system can facilitate the assessments of autism symptoms providing therapists with a useful set of reliable and objective quantitative methods. The proposed system is so flexible, robust, user-friendly, and easily customizable that we infer it could be utilized without effort by parents in domestic environments. Not only does not the system require any previous technical experience, but—thanks to the scalability of robotic intervention—it enables the efficient treatment of a large number of patients, increasing the frequency of the sessions that can be administered to the children while implementing exactly replicable protocols.

Some authors maintain that exposure to digital technology can aggravate the social symptoms of autism in children, worsening their deficits and possibly increasing the chances of developing obsessive compulsive behaviors [27]. The out-

comes of our research mitigate these worries. Our research provides robust anecdotal evidence that proper design and supervised application of robots in autism therapy has the potential to make ASD subjects feel spontaneously engaged in basic forms of social interaction and, at least apparently, significantly less anxious than during the typical interactions with other humans. We speculate that one of the contributing factors in achieving this positive result is the fact that robots allow effective forms of pseudo-social interactive engagement while decreasing the complexity of the social context and hence reducing the excessive emotional and cognitive burden that autistic children typically have to process. Our hypothesis is that this technology offers ASD children an opportunity to familiarize with social interaction in a context that they find conducive and reassuring with a pace that they can comfortably control by means of tasks that they can repeat at will without hurry. That is why we believe our robot-based interventions have the potential to support the development of social skills in autistic subjects and can teach them new ones, consolidating their ability to interact with other humans.

Future developments of the proposed system would have to aim to increase the assessment accuracy and further enhance the patient's engagement with the robots broadening the set of available types of interaction and increasing the degrees of freedom that define such interactions. Multiple cameras could be employed where a fixed observation setup is possible (considering the specificities of the clinical setting) to preserve the robot's mobility while broadening their interactive capabilities. Furthermore, a large set of patients would be desirable to test more extensively all the functionalities of the system and tune them for performance improvement. Finally, it would be useful to test whether a virtual avatar displayed on a tablet could reduce the costs involved in the use of embodied mobile robots, simplifying the use of the interactive system in the home setting: however, we anticipate that the quality of the interaction with a virtual avatar might be inferior in terms of both quality and length to the interaction established by the children with a physically embodied robot.

Funding This study was funded by a 31R188-Research AUA- ZCHS -1-2018, Zayed Health Center.

Compliance with Ethical Standards

Conflict of interest The second and fourth authors are coeditors of the special issue to which this paper is submitted, however do not anticipate to be involved in the review process for this paper. The other authors declare that they have no conflict of interest.

References

- Rotheram-Fuller E, Kasari C, Chamberlain B, Locke J (2010) Social involvement of children with autism spectrum disorders in elementary school classrooms. *J Child Psychol Psychiatry* 51:1227–1234
- Moriuchi JM, Klin A, Jones W (2016) Mechanisms of diminished attention to eyes in autism. *Am J Psychiatry* 174:26–35
- Freeth M, Foulsham T, Kingstone A (2019) What affects social attention? Social presence, eye contact and autistic traits. *PLoS One* 8(1):e53296. <https://doi.org/10.1371/journal.pone.0053296>
- Roddy A, O'Neill C (2019) The economic costs and its predictors for childhood autism spectrum disorders in Ireland: how is the burden distributed? *Autism* 23(5):1106–1118
- Huerta M, Bishop SL, Duncan A, Hus V, Lord C (2012) Application of DSM-5 criteria for autism spectrum disorder to three samples of children with DSM-IV diagnoses of pervasive developmental disorders. *Am J Psychiatry* 169:1056–1064
- Volkmar F, Siegel M, Woodbury-Smith M, King B, McCracken J, State M (2014) Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *J Am Acad Child Adolesc Psychiatry* 53:237–257. <https://doi.org/10.1016/j.jaac.2013.10.013>
- Llaneza DC, DeLuke SV, Batista M, Crawley JN, Christodulu KV, Frye CA (2010) Communication, interventions, and scientific advances in autism: a commentary. *Physiol Behav* 100:268–276
- Paul R (2008) Interventions to improve communication in autism. *Child Adolesc Psychiatr Clin N Am* 17:835–856
- Costa S (2014) Robots as tools to help children with ASD to identify emotions. *Autism* 4(1):1–2
- De Bel-Air F (2015) Demography, migration, and the labour market in the UAE. *Gulf Labour Mark Migr* 7:3–22
- Aresti-Bartolome N, Garcia-Zapirain B (2014) Technologies as support tools for persons with autistic spectrum disorder: a systematic review. *Int J Environ Res Public Health* 11:7767–7802
- Pennisi P, Tonacci A, Tartarisco G, Billeci L, Ruta L, Gangemi S, Pioggia G (2016) Autism and social robotics: a systematic review. *Autism Res J* 9:165–183
- Alahbabi M, Almazroei F, Almarzoqi M, Almeheri A, Alkabi M, Al Nuaimi A, Cappuccio ML, Alnajjar F (2017) Avatar based interaction therapy: a potential therapeutic approach for children with Autism. In: IEEE international conference on mechatronics and automation, pp 480–484
- Zheng Z, Zhang L, Bekele E, Swanson A, Crittendon JA, Warren Z, Sarkar N (2013) Impact of robot-mediated interaction system on joint attention skills for children with autism. In: IEEE international conference on rehabilitation robotics, pp 1–8
- Robins B, Dautenhahn K, Dubowski J (2006) Does appearance matter in the interaction of children with autism with a humanoid robot? *Interact Stud* 7:479–512
- Zhang K, Liu X, Chen J, Liu L, Xu R, Li D (2017) Assessment of children with autism based on computer games compared with PEP scale. In: 2017 international conference of educational innovation through technology (EITT), pp 106–110
- Rudovic O, Lee J, Dai M, Schuller B, Picard RW (2018) Personalized machine learning for robot perception of affect and engagement in autism therapy. *Sci Robot* 3:eaa06760
- Ferrer EC, Rudovic O, Hardjono T, Pentland A (2018) Robochain: a secure data-sharing framework for human–robot interaction. *arXiv preprint arXiv:180204480*
- Anzalone SM, Tilmont E, Boucenna S, Xavier J (2014) How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3D+ time) environment during a joint attention induction task with a robot. *Res Autism Spectr Disord* 8:814–826
- Kim ES, Paul R, Shic F, Scassellati B (2012) Bridging the research gap: making HRI useful to individuals with autism. *J Hum Robot Interact* 1:26–54
- Bensalem S, Gallien M, Ingrand F, Kahloul I, Thanh-Hung N (2009) Designing autonomous robots. *IEEE Robot Autom Mag* 16:67–77
- Billard A, Robins B, Nadel J, Dautenhahn K (2006) Building Robota, a mini-humanoid robot for the rehabilitation of children with autism. *Assist Technol* 19:37–49
- Kozima H, Nakagawa C, Yasuda Y (2005) Interactive robots for communication-care: a case-study in autism therapy. In: IEEE international workshop on robot and human interactive communication, pp 341–346
- Schopler E, Van Bourgondien M, Wellman J, Love S (2010) Childhood autism rating scale—second edition (CARS2): manual. Western Psychological Services, Los Angeles
- Alnajjar FS, Renawi AM, Cappuccio M, Mubain O (2019) A low-cost autonomous attention assessment system for robot intervention with autistic children. In: 2019 IEEE global engineering education conference (EDUCON)
- Pokress SC, Veiga JJD (2013) MIT app inventor: enabling personal mobile computing. *arXiv preprint arXiv:13102830*
- Wilson PI, Fernandez J (2006) Facial feature detection using Haar classifiers. *J Comput Sci Coll* 21:127–133
- Ahmad MI, Mubin O, Orlando J (2017) Adaptive social robot for sustaining social engagement during long-term children–robot interaction. *Int J Hum Comput Interact* 33:943–962
- Marcos-Pablos S, González-Pablos E, Martín-Lorenzo C, Flores LA, Gómez-García-Bermejo J, Zalama E (2016) Virtual avatar for emotion recognition in patients with schizophrenia: a pilot study. *Front Hum Neurosci* 10:421
- Powell S (1996) The use of computers in teaching people with autism. In: Autism on the agenda: papers from a National Autistic Society conference, London, pp 128–132

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Dr. Fady Alnajjar is Assistant Professor of Computer Science and Software Engineering at the College of IT of UAE University. Dr. Alnajjar is also the director of AI and Robotics Lab at the same university. He is also a visiting research scientist at the Intelligent Behavior Control Unit, BSI, RIKEN. He received a M.S. degree from the Department of Human and Artificial Intelligence System at the University Of Fukui, Japan (2007), and a Dr. Eng. degree in System Design Engineering from the same university in 2010. Since then he had interest on brain modeling in aim to understand higher-order cognitive mechanisms. He also started to study motor learning and adaptation from the sensory and muscle synergies perspective in order to suggest practical robotics applications for neural-rehabilitations.

Dr. Massimiliano Cappuccio is Deputy Director of the Values in Defence and Security Technology group at University of New South Wales Canberra. The group investigate the normative frameworks that regulate human–robot teams and autonomous weapon systems in the military and the law enforcement contexts. Max also holds a secondary affiliation with UAE University, the national university of the United Arab Emirates, where he collaborates with the AI & Robotics Lab, developing a line of interdisciplinary research in neuro-cognitive robotics, robotics in education, and child–robot interaction. In this context, Max has used humanoid robots to conduct empirical studies and

test neuro-cognitive models of choking effect and social interactions of autistic children. Max's research has three main thematic foci: human-robot interaction and social cognition (human-robot teams, normative frameworks for social coordination); technology ethics (including the ethics of AI and virtue ethics applied to human-agent interaction); human performance and embodied cognition (skill acquisition, performance under pressure). This research is interdisciplinary in nature as it integrates phenomenological analyses, empirical experimentation, and synthetic modelling. Max's early work, conducted in Italy, France, Netherlands, and the US, has focused on mirror neurons theory, joint attention and deictic pointing, the frame problem of artificial intelligence, the theoretical foundations of computationalism. Max is the organiser of international academic events like the annual Joint UAE Symposium on Social Robotics in Abu Dhabi and the TEMPER workshop on human performance in Canberra.

Eng. Abdulrahman Renawi has M.S. in mechatronics from the American university of sharjah. Currently working as a research assistant at the AI and Robotics lab in UAE University. His main research interest is to utilize AI techniques to design automatic assessment system for attention record level of autistic children in classroom.

Dr. Omar Mubin is a Senior Lecturer in Human Computer Interaction at the School of Computer, Data and Mathematical Sciences at Western Sydney University, Australia. Dr Mubin's primary research interests are human robot interaction and human-agent interaction. Specifically, he studies social robotics and their applications and consequently interaction with humans in educational, public spaces and information dissemination scenarios. Dr Mubin is also the general chair for the Human Agent Interaction (HAI2020) conference. His current h-index according to Google Scholar is 16 with a more than 1400 citations. Dr Mubin is involved in teaching and supervising (undergrad and post-grad) students in the broader area of Human Computer Interaction, Mobile Computing and Health Informatics.

Professor. Dr. Chu Kiong Loo completed his Bachelor of Mechanical Engineering degree with honours at the University of Malaya in 1996. He then received his PhD degree at the University of Science Malaysia in 2004, specialising in neurorobotics and its applications in Connected Healthcare. Currently he is a full Professor at the Department of Artificial Intelligence, Faculty of Computer Science & Information Technology, University of Malaya. His main research interests are neuroscience inspired machine intelligence. He has led numerous research projects as Principal Investigators of several international funding projects including Joint Research Program between UAE University and Institutions of Asian Universities Alliance (AUA). From these research activities, he has published more than 200 peer-reviewed journal and conference papers on neurorobotics and machine intelligence. Prof. Loo is also the recipient of JSPS fellowship award from Japan and Georg Forster Fellowship award from Humboldt Foundation, Germany. He has been visiting professor of University of Perfectual Osaka University, Japan and King Mongkut's Institute of Technology Ladkrabang, Thailand.