International Journal of Social Robotics (2018) 10:179–198
https://doi.org/10.1007/s12369-017-0461-4

Human–Robot Facial Expression Reciprocal Interaction Platform: Case Studies on Children with Autism

Ali Ghorbandaei Pour[1] · Alireza Taheri[1] · Minoo Alemi[1,2] · Ali Meghdari[1]

Accepted: 26 December 2017 / Published online: 24 January 2018

Abstract
Reciprocal interaction and facial expression are some of the most interesting topics in the fields of social and cognitive robotics. On the other hand, children with autism show a particular interest toward robots, and facial expression recognition can improve these children's social interaction abilities in real life. In this research, a robotic platform has been developed for reciprocal interaction consisting of two main phases, namely as *Non-structured* and *Structured* interaction modes. In the Non-structured interaction mode, a vision system recognizes the facial expressions of the user through a fuzzy clustering method. The interaction decision-making unit is combined with a fuzzy finite state machine to improve the quality of human–robot interaction by utilizing the results obtained from the facial expression analysis. In the Structured interaction mode, a set of imitation scenarios with eight different posed facial behaviors were designed for the robot. As a pilot study, the effect and acceptability of our platform have been investigated on autistic children between 3 and 7 years old and the preliminary acceptance rate of $\sim$78% is observed in our experimental conditions. The scenarios start with simple facial expressions and get more complicated as they continue. The same vision system and fuzzy clustering method of the Non-structured interaction mode are used for automatic evaluation of a participant's gestures. Lastly, the automatic assessment of imitation quality was compared with the manual video coding results. The Pearson's r on these equivalent grades were computed as $r = 0.89$ which shows a sufficient agreement on the automatic and manual scores.

Keywords Human–robot interaction (HRI) · Reciprocal interaction · Facial expressions · Autism · Fuzzy finite state machine · Imitation

1 Introduction

In the field of social robotics, researchers have been interested in the development of interaction between a humanoid robot and its environment, particularly the possibility of interacting with other robots and/or humans. In human–robot interaction (HRI), understanding a user's emotion is quite valuable. In an HRI system, different decision-making algorithms are used for the emotion synthesis unit, all of which require considering both temporal and emotional states of the user [1]. One of the important factors for measuring the user's moods and behaviors is facial behavior such as facial expression, eye contact, gaze direction regulation, and head orientation [2]. Facial expression is one of the important aspects of nonverbal communication that can be used for autonomous HRI systems [3]. Therefore, facial behavior analyses can be used for real-time human–robot emotional interaction to give robots a better understanding of the user's emotions.

Concurrently, individuals with autism have atypical emotion recognition and mental states, including facial expressions, vocal intonation, and body language, as components of social cognition [4]. In [5], a review of behavioral studies on face processing discusses the nature of the face processing problems for individuals with autism. Baron-Cohen [6,7]

✉ Ali Meghdari
meghdari@sharif.edu
http://meghdari.sharif.edu

Ali Ghorbandaei Pour
ali.ghr@gmail.com

Alireza Taheri
taheri@mech.sharif.edu

Minoo Alemi
alemi@sharif.edu

[1] Social and Cognitive Robotics Laboratory, Center of Excellence in Design, Robotics and Automation (CEDRA), Sharif University of Technology, Tehran, Iran

[2] Faculty of Humanities, West Tehran Branch, Islamic Azad University, Tehran, Iran

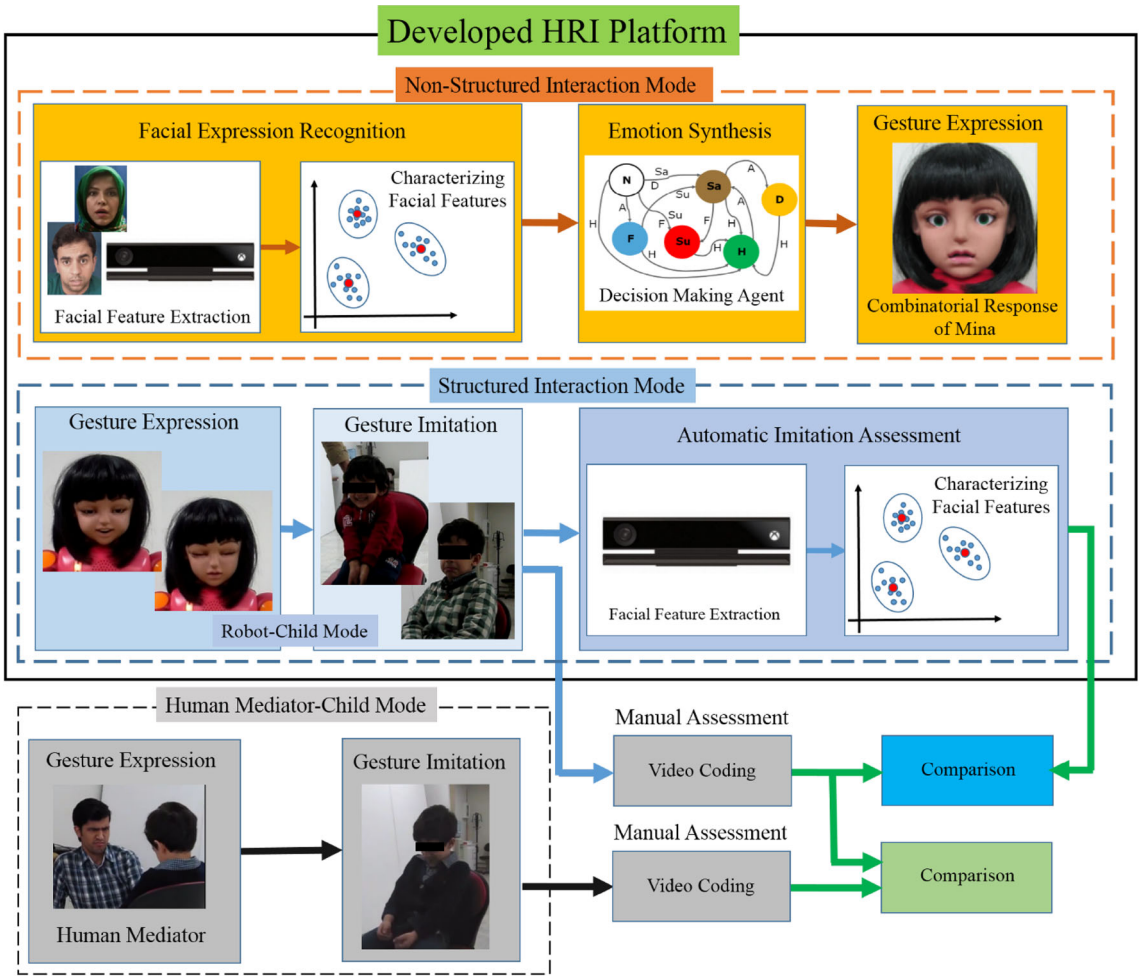

Fig. 1 The architecture of human–robot facial expression reciprocal interaction platform with the evaluation system

believes that a deficit in theory of mind and empathy causes this problem for children with autism. In fact, the ability to recognize and distinguish emotional expressions develops at age 10-weeks in Typically Developing (TD) children [8]; and TD children can interpret others' emotions and read emotions from an adult's eyes at age ∼11 years [9]. However, children with autism might not be able to naturally decode other's emotions and can behave differently from typically developing children during social interactions [10]. Reciprocal imitation games are an appropriate way to teach imitation skills (as a core symptom) to children with Autism Spectrum Disorders (ASD) which could positively affect their social and communication skills, pretended play, and facial gestures [10,11]. Moreover, taking into account the increasing trend of using social robots in education [12], care of children and older adults [13,14], and autism research [15–17], a robot with the ability to show facial expressions may be an appropriate tool for reciprocal imitation games. Due to the fact that progress in fine movement imitation can potentially improve real-life social skills of children with autism [10,15–19], interaction through facial expressions with a robot may be a suitable therapeutic application of such studies. Combining the above statements, human–robot reciprocal interaction through facial expressions is worthy of attention in autism research for possible rehabilitation purposes.

As the main contribution, this paper presents an initial attempt to develop a robotic platform for human–robot reciprocal interactions through facial expressions and investigates its acceptability and performance on a set of Iranian children with autism. This study is built on our previous research on the fundamental design of the first side of the reciprocal HRI platform [20]. A high-level overview of the designed platform, including Non-structured and Structured interaction modes (Fig. 1), is presented in the following.

(a) Phase 1, Non-structured interaction mode: The main goal of this phase is to empower our HRI platform to react emotionally to the participants' facial behaviors in real-time. We presented an automated system with a fuzzy algorithm for recognizing the intensity of the facial

expressions and their relative emotional states. The novelty of this phase is developing a decision-making fuzzy state machine to improve the quality of interaction, which leads to a more realistic real-time reaction from the robot through its facial expression and neck movement. This phase is called the Non-structured interaction mode because the robot adapts itself and reacts to arbitrary facial expressions of the participant.

(b) Phase 2, Structured interaction mode: In the second phase of the developed HRI platform, the user is supposed to react to the robot's facial expressions during robot–child interactions. As a case study, the performance of some children with autism in robot-assisted facial-imitation is measured through manual video coding of the imitation tasks. In order to test the hypothesis: "children with autism perform better in robotic than non-robotic facial imitation actions", the results of the previous section are compared to the scores of the same tasks while imitating a human mediator. A fuzzy-based assessment algorithm is then introduced to do the same scoring for the robot-assisted imitation tasks. Exploring whether there is a correlation between the manual scoring by human-coders (as the reference grades) versus the automatic assessment by the machine (i.e. the robot) is the next contribution of this phase.

This paper follows a scientific approach to study the performance of the robotic reciprocal interactions for each part of the proposed HRI architecture. The findings of such studies would help autism researchers to design and conduct therapeutic intervention scenarios for possible rehabilitation purposes.

2 Related Work

2.1 Facial Expression Recognition for HRI

In recent years, emotion-based human–robot social interaction has attracted considerable attention from academic and research communities [21–25]. Zacharatos et al. [21] explained recent emerging techniques and advances in automatic emotion recognition as well as recent application areas and notation systems. They also described the importance of movement segmentation in automatic emotion recognition. In [22], a set of upper body gestures is used to interact with a humanoid robot in real-time. The robot's reactions were through body movements, facial expressions, and verbal language. Aly and Tapus introduced a multimodal behavior HRI for more naturally emotional interaction. During their experiment, the humanoid robot has engaged and generated an adapted behavior using facial expressions, speech, and head-arm metaphoric gestures [23].

Due to the fact that facial expressions play a key role in emotional interaction, automatic analyzing of facial expressions has been an active research topic during the last 2 decades [26–33]. In the early 2000s, new methods for detecting the anatomic action units (AU) appeared [26]. Recently, the focus has shifted toward analyzing the spontaneous facial expression and studying the dynamics of the expressions [27]. Halder et al. [28] have used an interval and a general type-2 fuzzy set to model the fuzzy face space for emotion recognition purposes. The algorithm adopted in their study has resulted in a classification accuracy of 98.3%. In [29], the concept of a prototype-based model for characterizing facial expressions is introduced. They obtained an expression recognition result of 87% with a person-independent evaluation approach. A novel multiclass Support Vector Machine (SVM) system is used to recognize all basic facial expressions based on geometric deformation of facial features with the accuracy of 99.7% on the Cohn–Kanade database [30]. There have also been several studies concentrating on designing and developing a more intelligent and reliable facial expression analysis for HRI [34–39]. Chakraborty et al. [34] proposed an accurate, simple, and robust scheme for emotion recognition and control based on a fuzzy relational approach. In this study, eye opening, mouth opening, and the length of eyebrow constriction are extracted and mapped onto an emotion space by applying Mamdani-type relational models. In [35], changes of 21 facial distances describe the facial feature deformations, which were classified based on SVM method. Their experimental results showed that the proposed approach has a recognition rate of more than 90% in real-time. In another study, a Dynamic Bayesian Network classifier is used in order to estimate a human emotion for recognition and imitation of facial expressions [37]. The output of their classifier updates a geometric robotic head model. Developing a robotic head with the ability to imitate facial expressions through the robot's eye expression is proposed in [38]. They have used a 3D Constrained Local Model and Hidden Markov Model for localizing the facial features and distinguishing the emotional state of the user, respectively.

The Non-structured interaction mode of this paper includes an automatic facial expression recognition with a fuzzy facial expression response of the robot. As a module of the designed reciprocal architecture, this interaction mode can promote shared attention and social responsiveness in children with autism, among other mentioned social robotic applications. We have used this mode to investigate the acceptability of our robotic platform for children with autism.

2.2 Robotic Expressive Faces in Autism Research

The motivation behind the increasing trend of using social robots in different aspects of autism research falls with

the reports presented by scientists during the last decade; contrary to their inability to interact with their typically developing peers, children with autism have shown that they can interact with intelligent robots and technological social agents in a natural way (cited or confirmed in [15,16,19,40–43]). Scassellati et al. [16] mentioned that social robots are less complex than human and animate social beings, and more interactive tools than inanimate toys; the former could be a source of distress or confusion, and the later cannot usually elicit novel social behaviors for children with ASD. In the past, researchers have reported utilizing computer avatars [18,40] and social robots with low or high facial complexities [16,41–44] in teaching different facial expressions to children with ASD. As an exploratory study, Duquette et al. [41] involved two children with low-functioning autism in robot-assisted and two other children with ASD in human-mediated reciprocal interactions and imitative games. They reported that the two participants paired with the Tito robot showed increased shared attention during the body and facial imitation games. Moreover, these participants showed better imitation performance in facial expression tasks than the two other subjects in the human mediator–child mode. Salvador et al. [42] used a Zeno (R50) robot to investigate the comparison of the emotion expression recognition of 11 children with autism and 11 typically developing peers. In this study, the authors programmed their robot to show 13 emotions and the children were asked to guess what emotion the robot displayed. While the authors did not observe a significant difference in the emotion prediction between the two groups, they reported a specific deficit in identifying the fear emotion by the ASD group. Moreover, they confirmed the potential of using a robot with facial expressions to improve the social skills of children with autism. Hopkins et al. [18] investigated the effect of their computer-based training program, FaceSay, on the emotion and facial recognition skill development of their subjects with autism. They reported the improvement of their low-functioning participants with autism in emotion recognition, and high-functioning subjects in emotion and facial recognition after involving them in interactive realistic avatar games.

In order to have a reciprocal facial imitation training HRI for children with autism, the Structured interaction mode has been embedded in the developed architecture. Through this interaction mode, we will be able to investigate the participants' facial imitation performance in the robot–child mode.

3 Research Methodology

3.1 Robotic System

The Microsoft Kinect Sensor for Xbox One [45] is used for the machine vision application. The Microsoft Kinect Sensor is a physical device with depth sensing technology, a built-in color camera, and an infrared emitter. With the help of version 2 from the Kinect for Windows Software Development Kit (SDK), it is possible to access a list of high detail face points to extract facial features.

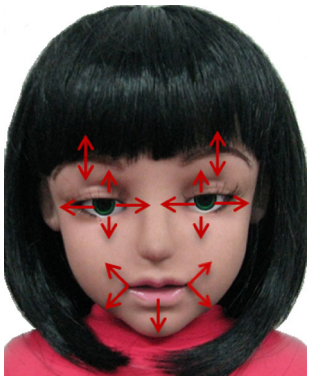

Fig. 2 Eight DOF in Mina's face

The humanoid robot used in this study is the R50-Alice by Hanson RoboKind Company [46] designed specifically for human–robot social interaction. The robot was named "Mina" (a Persian name) since it was going to be used for interaction scenarios with Iranian children. Mina has 11 degrees of freedom (DOF) in her head, 3 DOF in her neck, and 8 DOF for generating facial expressions (Fig. 2). In this research, Mina does not have any verbal communication with the user.

3.2 Non-structured Interaction Mode

3.2.1 Facial Expression Recognition

In this part of the study, an algorithm has been developed to use the Kinect sensor for facial expression recognition. Three main steps must be taken to automatically recognize facial expressions, as follows:

(1) Facial feature extraction
(2) Gathering a database
(3) Classification

In our study, first, the facial features are extracted with the help of the Kinect SDK; second, the facial expression database is collected; and thirdly, we used the Fuzzy C-Means (FCM) method for classification.

Facial expression feature extraction algorithms mainly use feature points or convert pixel intensity. Studies on facial feature modeling in the literature [35] are based on the following items:

(1) Facial point displacements.
(2) Facial point coordinates.

Table 1 List of the extracted facial features

Facial features	Description
F1	Distance between inner corner of the left eye and inner left eyebrow
F2	Distance between outer corner of the left eye and outer left eyebrow
F3	Distance between center of the left eyebrow and the line crossing inner and outer corners of left eye
F4	Distance between inner corner of the right eye and inner right eyebrow
F5	Distance between outer corner of the right eye and outer right eyebrow
F6	Distance between center of the right eyebrow and the line crossing inner and outer corners of right eye
F7	Distance between inner left eyebrow and inner right eyebrow
F8	Distance between left corner and right corner of the mouth
F9	Distance between middle of the upper lip and middle of the lower lip
F10	Distance between left corner of the mouth and inner corner of the left eye
F11	Distance between left corner of the mouth and left cheekbone
F12	Distance between left corner of the mouth and left end of the lower jaw
F13	Distance between right corner of the mouth and inner corner of the right eye
F14	Distance between right corner of the mouth and right cheekbone
F15	Distance between right corner of the mouth and right end of the lower jaw
F16	Distance between inner corner of the left eye and bottom left of the nose
F17	Distance between inner corner of the right eye and bottom right of the nose
F18	Distance between bottom of the nose and the line crossing from inner corner of the left and the right eye

(3) Distances between points based on the deformation of the facial contour.
(4) Deformation of the shape from the neutral state regardless of the contraction of facial muscles.
(5) Modeling of muscle contraction using the variation of muscle distances.

Since the Kinect has been used as the vision system for feature extraction, a technique based on the distance between facial points is used. Details of assigning the facial points and feature extraction mechanism are available in our previous works on facial expression HRI [20,39]. Eighteen facial features were chosen in such a way to represent the action units of the Facial Action Coding System (FACS) [47]. These features are listed in Table 1.

The data is recorded from the Kinect sensor and each feature is updated at the rate of 30 frames per second. The recorded data contains noise and cannot be used for the recognition phase. A moving average filter with a period of 5 previous data points is applied to each feature to reduce the effect of noise on the extracted features. A second issue is that the user's facial features should be scale invariant; for this reason and to avoid the effect of user's distance from the Kinect sensor, two methods are applied to these features. In the first method, features F1 to F17 are divided by the value of the participant's F18. The resulting features are named as features represented in "Mode (A)". In the second method, each feature (from F1 to F17) is divided by its neutral value (i.e. the value of the feature when the face is in the neutral state) and these results are called features represented in "Mode (B)". Both of these feature types are scale invariant, but they have different characteristics. For Mode (A) representation, the neutral state of the user's face is not needed which can be taken as a potential advantage; and features in Mode (B) are normalized which can lead to a more robust classification. These two feature types resulted in different recognition rates which are discussed in the Results section. A suitable set of data is needed to train and then evaluate the efficacy of an automated facial expression recognition system that can detect emotional state. Since there was no available detailed database for high detail facial points recorded by the Kinect sensor, an added value of this work was to gather a database for facial expressions using the Kinect data.

The training dataset for the Non-structured interaction mode consists of facial features of 4000 samples gathered from different poses of 6 main facial expressions from 12 different young adults (500 samples for each facial emotional state). These main emotional states are happiness, sadness, anger, surprise, disgust, and fear. The subjects of this database are 6 males and 6 females with a mean age of 24.8 years and standard deviation of 4.5 years. The procedure of recording facial expressions and choosing the samples was done under the supervision of a psychologist. The feature vectors were selected from a range of feature sequences captured from

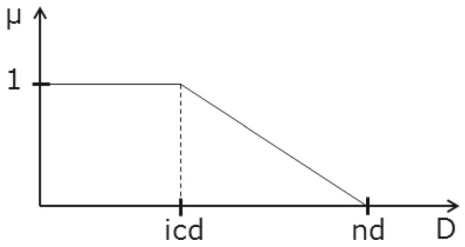

Fig. 3 Membership function for the clusters versus the distance (D) from the center of the cluster

our subjects while they were asked to watch six 15- to 45-s selected Iranian videos in order to elicit more realistic emotional states in response to the videos.

The final stage of facial expression recognition is to apply a classifier (or clustering approach) to discriminate the facial features to one of the existing facial expressions (basic emotional states). In this study, a fuzzy clustering method called the Fuzzy C-Means method is used. Fuzzy C-Means (FCM) is a clustering algorithm introduced for classification of numeric data [48,49]. This method was chosen for two main reasons. First, we needed a classifier with fuzzy output in order to find the membership values related to the input vector of features captured by the Kinect. As the next step, these membership values were used to produce Mina's response based on a fuzzy finite state machine (proposed in Sect. 4.1.2). Second, FCM is a clustering technique and since the clustering algorithms are considered as unsupervised learning techniques, knowing the correct label for the training dataset is not necessary. This trait gives the recognition algorithm the ability to increase its dataset by adding the input samples to the database during interaction with new users.

In this part of the study, the first step is finding the center of the clusters (each of the six basic emotions) for the samples in our database. Each center of a cluster is a vector with 17 elements (normalized distance of each cluster from the neutral state). Then, the inside cluster deviation is calculated for each of these emotional states. Finally, the membership value of a given facial expression (normalized facial feature vector) for each of the emotional states is defined according to the membership function presented in Fig. 3. The horizontal axis represents the distance of the given facial expression from the center of the cluster, and the vertical axis is the membership value output corresponding to that cluster. "icd" indicates the inside cluster deviation and "nd" stands for the center of the cluster's distance from the neutral state.

3.2.2 Human–Robot Emotional Interaction

The first step toward a more realistic response from Mina in the Non-structured interaction mode is tracking the user by moving the head of the robot. For this purpose, Neck Yaw and Neck Pitch (angles for rotating in the azimuth and elevation planes, respectively) are able to change such that Mina is always facing her user as she responds to the user's emotional state. These angles are calculated according to the position of the user's head. From the data output of the Kinect, the head position is available in the Kinect body coordinate system. Therefore, the position of the Kinect sensor and Mina's head center should be stationary. The Kinect sensor is attached to a physical fixture to maintain the relative distance and orientation between the two coordinates (see Fig. 4). To calculate the proper neck angles, this position needs to be transferred to the robot's head coordinate system. Figure 5 illustrates the head position in both coordinate systems and proper rotating angles in the robot's head coordinate system. The Kinect Body Coordinate System (xyz) and Robot's Head Coordinate System (XYZ) are shown with brown and black colors, respectively.

Fig. 4 Mina sits on a chair with the Kinect attached to it

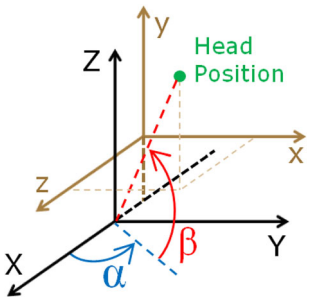

Fig. 5 Mina's Neck Yaw (α) and Neck Pitch (β) in the reference coordinates

The next step is to implement the decision-making agent. A finite state machine was used to generate an emotional reaction to indicate the emotional state of the robot. The user's emotional state was the input to the state machine (see Fig. 6).

The output of each state was a set of facial expressions produced by Mina, declaring her reaction to the user's emotional

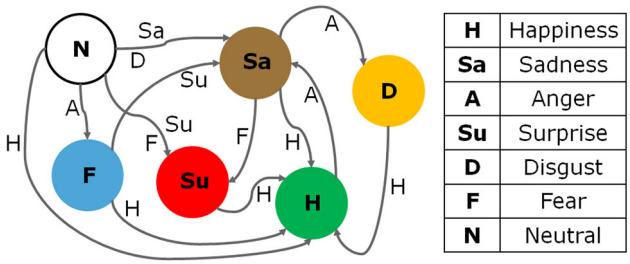

Fig. 6 The finite state machine diagram for the Structured interaction mode

Fig. 7 Mina winks at the user

state. This output is set as a vector containing the actuation level for each degree of freedom in Mina's face. Since the designed HRI system is going to be used for interaction with children with autism, the interaction scenarios were designed through consultation with a clinical child psychologist. The scenarios were finalized during the session with the psychologist, and since we decided that Mina should not become angry during the interactions with the children, there is no "Anger" state considered in the state machine diagram. The transition between states is according to the user's detected emotional state.

The designed finite state machine starts in the "Neutral State" and according to the current emotional state of the user, Mina will change her emotional state by using the different outputs assigned to each state. The output for each emotional state is produced based on FACS action units and the robot's facial expression for each state is confirmed based on FACS by the psychologists in our team. Mina, as mentioned before, has eight DOF in her face for expressing facial gestures. Therefore, the outputs for the states are vectors with eight elements, each with values between zero and one. Table 2 lists the actuation values for each state.

In order to make the interactions more appealing to the users, in some selected states, a timer is also considered that counts the time Mina stays in those states. Such additional extra actions can potentially help to emphasize the meaning of each emotional state which is also used in [23]. If Mina stays in these certain states for a specified amount of time, extra motions will be triggered. For example, if Mina stays in the "Happiness" state for 10 s, she will then wink at the user (see Fig. 7). Table 3 shows the extra motions and their relevant trigger time. For states with two or more possible extra actions, one of the actions is chosen randomly.

Table 2 Mina's face actuation values for each state; the actuators' values of the robot can be set between 0 and 1

State	Output vector	FACS AUs
Neutral	Jaw = 0	–
	Other actuation values are 0.5	
Happiness	Right smile = 1, left smile = 1	Lip corner puller, lips part
	Jaw = 0.4	Cheek raiser
	Other actuation values are 0.5	
Sadness	Right smile = 0, left smile = 0	Lip corner depressor, Cheek raiser
	Jaw = 0, Brows = 0.2	Inner brow raiser, Chin raiser
	Eyelids = 0.3	Brow lower
	Other actuation values are 0.5	
Surprise	Jaw = 1, Brows = 1	Inner brow raiser, Upper lid raiser
	Eyelids = 1	Outer brow raiser, jaw drop
	Other actuation values are 0.5	
Disgust	Brows = 0, Eyelids = 0.7	Brow lower, Nose wrinkle
	Other actuation values are 0.5	Upper lip raiser, Lips part
Fear	Right smile = 0.3, Left smile = 0.3	Inner brow raiser, brow lower
	Jaw = 0.2, Brows = 0.8	Outer brow raiser, lips part
	Eyelids = 0.7	Upper lid raiser, lip stretcher
	Other actuation values are 0.5	

Table 3 Mina's extra actions for selected states

State	Extra motion	Time needed to start (s)
Happiness	One side smile	5
	Closing mouth smile	7
	Wink at the user	10
Sadness	Looking down with head bow	10
Disgust	Turning head to the sides	10
Fear	Opening mouth	5
	Covering the eyes with hands	10

Since the algorithm used for emotional state recognition in the previous section has a fuzzy output, a more realistic reaction can be generated by realizing the membership values of the user's facial expression for each emotional state. Also, the state machine can be implemented with a number of if-then rules as follows:

- `if (state == 'N' && entry == 'H') then {state = 'H'}`
- `if (state == 'N' && entry == 'A') then {state = 'F'}`
- `if (state == 'N' && entry == 'D') then {state = 'Sa'}`
- `if (state == 'N' && entry == 'F') then {state = 'Su'}`
- `if (state == 'H' && entry == 'A') then {state = 'Sa'}`
- `if (state == 'F' && entry == 'Su') then {state = 'Sa'}`
- `if (state == 'Sa' && entry == 'F') then {state = 'Su'}`
- `if (state == 'Su' && entry == 'H') then {state = 'H'}`
- ...

These rules are taken as the rule base of our fuzzy inference system. A fuzzy inference system is a method that interprets the membership values in the input vectors; and based on defined rules, assigns values to the output vector [50]. Then, for the system entry and each state, a membership value is considered (all state membership values are zero at the beginning). A new level of emotional reaction is generated by assigning the minimum of the system entry and the current state's membership values to the next coming state. Finally, computing weighted average between the outputs of each state produces Mina's facial response. States with membership values of more than 0.5 are considered for calculating the weighted average. For extra motions, the system chooses the action related to the state with the highest membership value, which must also be more than 0.85.

3.3 Structured Interaction Mode

Another objective of this research was the design and development of robot-based protocols to investigate facial behavior responses of children with autism through imitation actions. The hypothesis of this section is: "performance of children with autism in facial imitation tasks is better in robotic than non-robotic" which is investigated in this study. In addition, we would like to find a preliminary answer for the following research question: "Can the proposed fuzzy algorithm be used for automatic assessment of facial imitation?" In this phase, we designed and executed a set of fine movement facial imitation scenarios for children with autism and an initial attempt was made to develop our HRI system for automatic assessment by the Kinect data (see Fig. 1).

3.3.1 Participants with ASD

Fourteen Iranian children with autism, including 10 boys and 4 girls, 8 with high-functioning and 6 with low-functioning autism, voluntarily enrolled in this study to participate in the facial imitation actions for one 15-min session. The mean age of the participants was 5.4 years and the standard deviation was 1.5 years. The children with autism and their parents participated voluntarily in this study and they signed a consent form for moral obligations.

3.3.2 Facial Expressions Imitation Actions

Facial expressions are categorized as fine movements. The ability to concentrate and recognize fine movement is an important factor that children with ASD need to improve [10,16,18]. In order to explore the facial imitation quality of the participants, a set of easy-to-difficult hierarchical facial gestures should be designed. Since the basic facial expressions of FACS are not necessarily a set of easy-to-difficult actions, eight different facial behaviors (i.e. actions no. 1–8) were designed with the help of two clinical child psychologists. Table 4 presents more detailed information on each of these facial gestures. The actions start with simple facial expressions and get more complicated as the participant goes on. The number of involved facial components in each action has been shown in Table 4. For each participant, the actions were held in two different modes: robot–child mode and human mediator–child mode (Fig. 1); therefore, each subject has to imitate a total of 16 facial expressions. Moreover, to begin, half of the participants were randomly involved in the robot–child mode and then switched to the human mediator–child mode. Then, the other half proceeded in reverse order (counterbalance condition). For the robot's expressions, each

Table 4 Facial gestures' imitation actions for both robot–child and human mediator–child Modes

Action no.	Description	No. of involved facial components
1	Smiling with an open mouth	1
2	Closing and opening the eyes two times	1
3	Opening and closing the mouth without any other facial behavior	1
4	Smiling with a closed mouth	1
5	Corners of lips first to left and then to right	2
6	Surprise (mouth open and circular, eyes wide open, eyebrows up)	3
7	Sadness (eyes half closed, eyebrows and corners of lips down)	3
8	Disgust (mouth half open, eyebrows half up, corners of lips down)	3

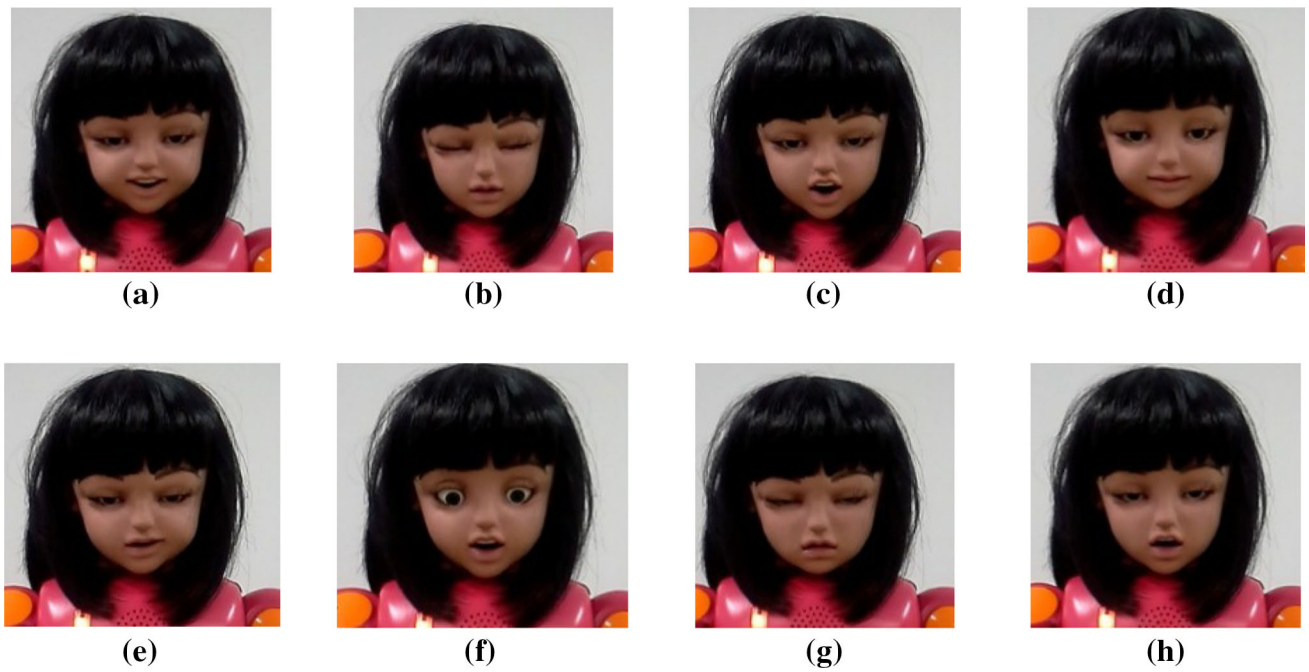

Fig. 8 Mina's facial gesture in each of the scenarios (actions no. 1–8)

facial gesture has been animated and produced in Workshop (RoboKind interface software) [46].

The main elements of the experimental setup included the Mina robot, Kinect sensor, chairs, and two cameras for filming sessions. A child, one of his/her parents, and the human mediator were present in the room. Also, a robot operator was present in an observation room. The participants were asked to sit on a chair about one meter from the Mina robot or human mediator and the Kinect system. The session was handled by the human mediator which included two main parts: (1) the Non-structured interaction mode, and (2) the Structured interaction mode. At the beginning of each part, the instructions and expectations of that section were described by the human mediator. First, the Non-structured interaction mode was performed for $\sim$ 5 min for each of these children to observe their reaction and find out their acceptance of the robot when they first notice the robot's facial responses. For the participants who accepted to interact with the robot (for at least 3 min), without taking breaks, the Structured mode was held as the next part. The Structured mode took $\sim$10 min. During the imitation actions, the children were asked to pay attention to Mina's or the human mediator's face and imitate the gestures with as much detail as possible. Figure. 8a–h, show Mina's facial gestures.

3.3.3 Manual and Automatic Imitation Assessment

In order to have a valid imitation assessment, we need a reliable reference. Therefore, two tasks were done. The action session was filmed from two angles (front and back). First, two specialists watched the films and rated the imitation quality of the participants on a Likert scale of 0, 1, 2, 3, or 4, for each of the mentioned eight scenarios in both robot–child and human mediator-child modes. In order to compare the

performance of the participants in robotic and non-robotic situations, a two-way ANOVA analysis was applied on the imitation scores of the children. As mentioned, the hypothesis of investigating the Structured interaction mode is that the imitation scores of the children with autism are better in the robot–child mode than the human mediator-child mode. Moreover, it is expected that actions with different difficulty levels would affect the subjects' performance. Therefore, we have considered "Action Number" and "Imitation Mode" as the two main study's independent factors in our ANOVA statistical test.

Alternatively, because of the differences between the facial gestures of Mina in the Non-structured and Structured interaction modes, a new dataset has to be provided. It should be noted that for the new dataset, the subjects needed to be trained to carry out the gestures. Therefore, in the second step, three available psychologists, who were qualified to manually assess the actions, were requested to play the actions and imitate Mina's gestures three times while their facial feature data was recorded with the Kinect and considered as the database for the Structured mode. The extracted data from this part was the reference for our automatic assessment algorithm. On the other hand, typically developing peers might also be used as subjects for capturing the training dataset. However, their performance should be confirmed by the evaluators through a manual assessment process in advance which is one of the limitations of the current study.

Since children with ASD have atypical eye gaze and concentration, the main problem with recording Kinect data was the unwanted gestures and head movements they made during the actions. To remove these parts of the data, we considered the parameter "Engaged" from "Face properties" of the Face Tracking Kinect SDK. Whenever the detection result from "Engaged" was "Yes", we used the extracted facial features for the assessment (Fig. 9). The algorithm used in this part is the same FCM introduced and used in the Non-structured interaction mode. Then, the membership value (between 0 and 1) of the captured gesture performed by the child for each scenario is considered for automatic assessment of his or her performance. The only difference from the Non Structured mode is in the database used for the assessment. In order to be able to compare the results from the automatic assessment with the manual assessment (i.e. the video coding), it was decided to rate the imitation's quality through mapping the related membership value of the captured gesture in a 0 to 4 interval.

4 Results and Discussion

The results of performing the Non-structured and Structured interaction modes will be presented and discussed in the next subsections.

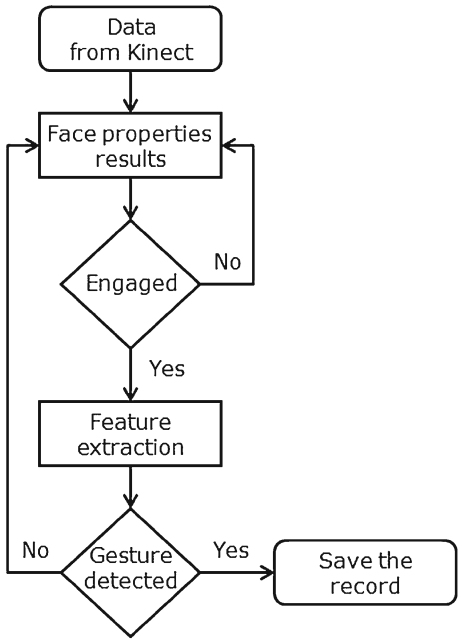

Fig. 9 Automatic imitation assessment flow chart

4.1 Non-structured Interaction Mode

4.1.1 Feature Extraction and Emotion Recognition

Figure 10 shows some of the facial feature evolution (after normalizing and noise filtering) captured from one of the subjects of the Non-structured mode database. The x-axis of each plot is frame number and the y-axis illustrates the normalized value of the facial feature. These selected features illustrate three facial expressions: happiness, anger, and surprise, respectively. Features in all of these video sequences begin in a neutral state. As it can be seen, face features are defined in such a way to be noticeably different for each facial expression, which leads to easier and more accurate classification.

A test dataset, containing 700 samples (100 samples for each emotional state and 100 samples of neutral face) from a new group of people (3 males and 2 females with a mean age of 21.2 years and standard deviation of 1.5 years), is used to validate the classification process for the Non-structured interaction mode. The highest membership value indicates the emotional state of each sample. A sample is considered neutral if all of the corresponding membership values are less than 0.65. Tables 5 and 6 presents the results from this test dataset for both feature types. Each row indicates the detection results for each set of samples with the same emotional state.

Based on the confusion matrices presented in Tables 5 and 6, the average detection rates for features presented in Mode (A) and in Mode (B) are 77.0% (standard devia-

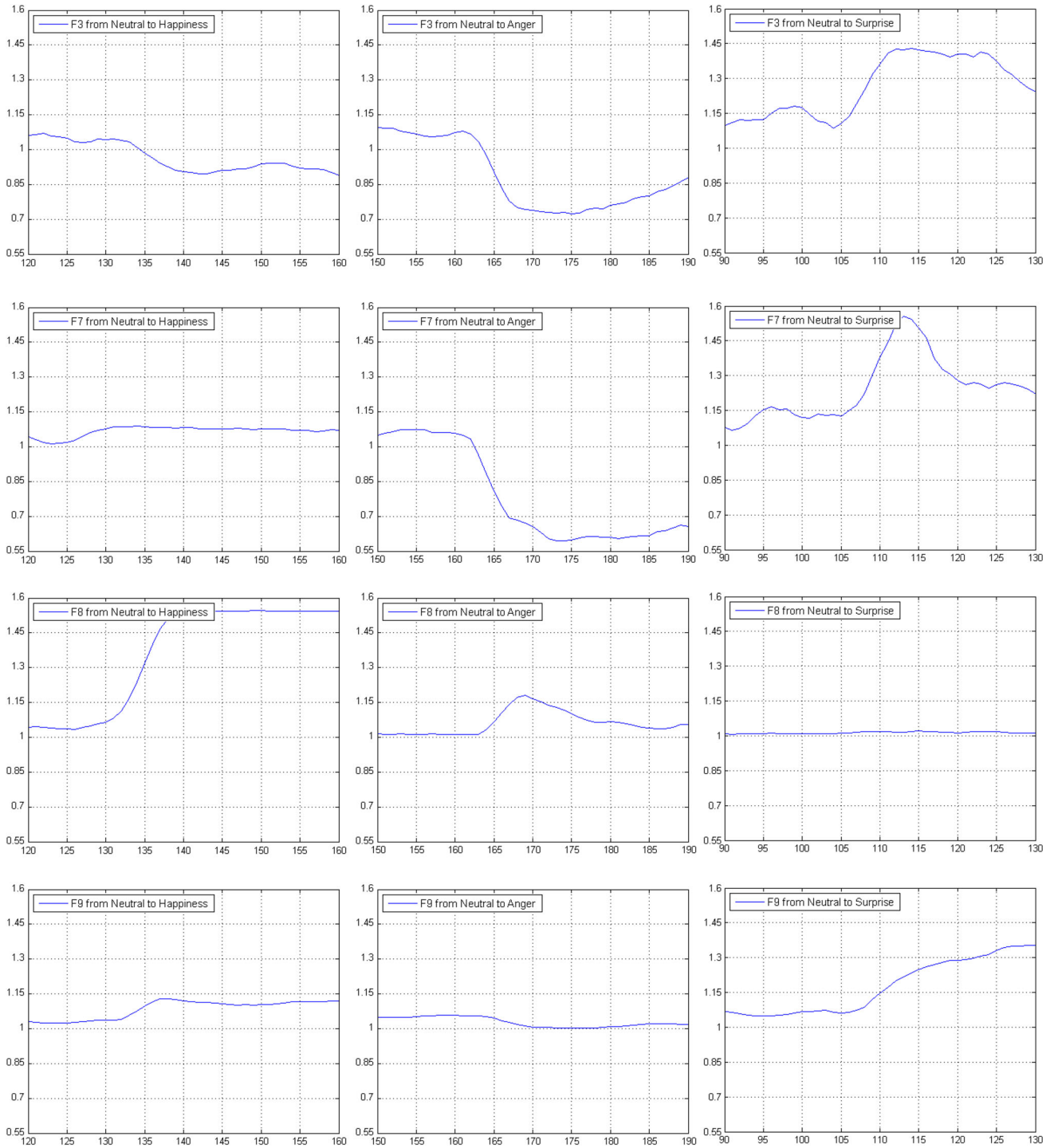

Fig. 10 Some facial features variation from video sequences for happiness, anger, and surprise (the x- and y-axis are frame number and the normalized value of the feature, respectively)

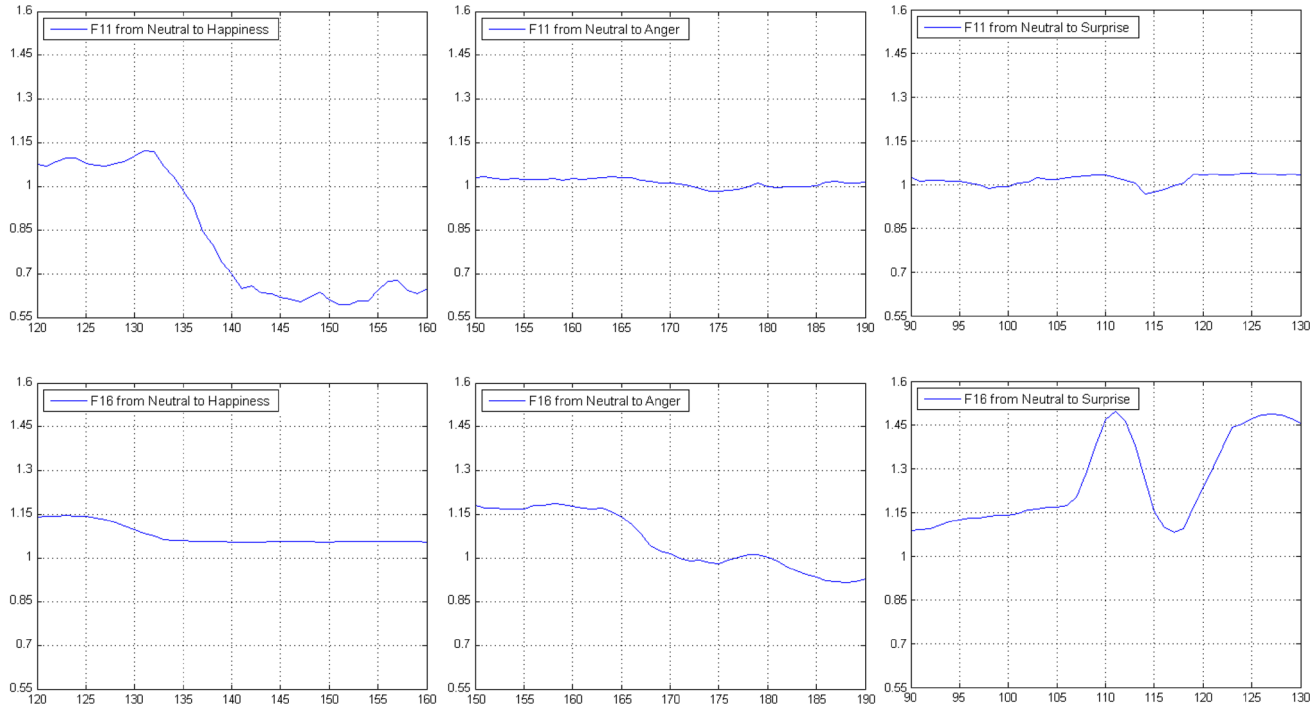

Fig. 10 continued

Table 5 Emotion recognition rate for the test dataset (features in Mode (A) format)

	Happiness (%)	Sadness (%)	Anger (%)	Surprise (%)	Disgust (%)	Fear (%)	Neutral (%)
Happiness	87	0	0	4	2	0	7
Sadness	0	74	8	0	0	10	8
Anger	0	6	76	0	11	3	4
Surprise	1	0	0	95	0	1	3
Disgust	2	2	9	1	71	3	12
Fear	0	8	0	19	2	62	9
Neutral	2	11	3	0	1	9	74

Table 6 Emotion recognition rate for the test dataset (features in Mode (B) format)

	Happiness (%)	Sadness (%)	Anger (%)	Surprise (%)	Disgust (%)	Fear (%)	Neutral (%)
Happiness	98	0	0	1	0	0	1
Sadness	0	91	4	0	0	3	2
Anger	0	2	94	0	4	0	0
Surprise	0	0	0	99	0	1	0
Disgust	0	0	6	0	94	0	0
Fear	0	4	0	8	0	83	5

tion: 10.8%) and 93.2% (standard deviation: 5.8%), respectively. The currently used facial features have resulted in the highest recognition rate of "Surprise" and the lowest rate for "Fear". The rate of emotion recognition directly depends on the feature extraction process. Generally speaking, based on [28–30,34–38], there is no cut agreement on the highest and lowest detection rate of basic emotional states.

It should be noted that based on the feature definition in Mode (A), the neutral state of the user's facial data is not needed in the feature preparation process and all the facial points distances are scaled by the same ratio (i.e. the nose

Fig. 11 Combinatorial facial expressions generated by Mina

length of the user); therefore, the scaled features are still dependent to the geometry of the user's face. In contrast, for feature represented in Mode (B), the user's facial features are prepared in proportion to the neutral state of his/her face; therefore, the nature of these normalized features would be less dependent on the face geometry. Accordingly, it makes sense that the recognition rate for features in Mode (A) is lower than features in Mode (B) format.

4.1.2 Robot's Emotional Reactions

Using the fuzzy finite state machine for generating proper facial expressions caused more interacting modes and a variable output level. Also, the change rate of the emotional state of Mina is dependent on the intensity of the user's facial expression. Figure 11 shows some of Mina's facial reactions to her user's emotional state which are combinations of the pre-defined basic emotional gestures.

4.1.3 Neck Movements of Mina

To improve the quality of human–robot interactions, Mina's head turns to face the user during the interaction. A smooth path with third order polynomial trajectory is also considered for each of the neck's angles of turning. If the position of the user's head moves while neck angles are moving toward their previous goal position, a new path will be generated according to the current neck angles values and the new destination angles (Fig. 12). Also, the new trajectory is considered to have the same velocity as the previous trajectory at the time the neck angle path changes its trajectory. This helps to have a smooth transition between trajectories.

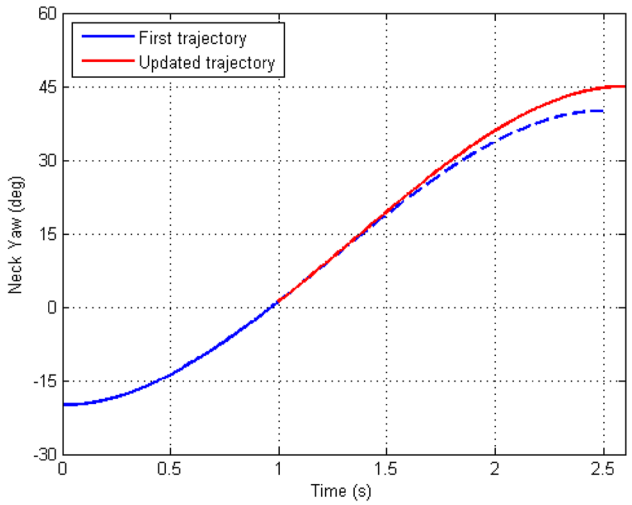

Fig. 12 Neck Yaw trajectory change in reaction to the change of user's head position

4.2 Structured Interaction Mode

4.2.1 Robot's Acceptability for the Participants with ASD

Before starting the imitation actions and in order to familiarize and draw the participants' attention to the robot, each of the children was involved in the Non-structured interaction mode for about 5 min; and Mina responded to his/her facial expressions. The introduction of the robot and the instructions of this part were presented to the child and his/her parent by the human-mediator. 9 out of 14 autistic participants showed a reasonable interest in interacting with the robot in the Non-structured mode (i.e. pursued interacting with Mina for at least 3 min). Two of them felt unpleasant at the beginning of the session, especially when Mina turned its face toward them; both of them stayed close to their mothers and avoid getting closer to the robot. But after encouragement from both their parents and the human mediator, they found the environment safe and gradually started to interact with Mina. 3 out of the 14 showed no interest in meeting the robots and refused to participate in the session. These three children had low-functioning autism and were less than 5 years old and two of them were female. One of the uncooperative female subjects did not get close to see Mina and insisted to her parent to take her back home. The other female subject sat on the floor while screaming continuously as soon as entered the room (Fig. 13a). For the uncooperative male subject, we understood that he could not tolerate the sound of Mina's cooling fan and after turning the robot off, he accepted to stay in the room. Overall, 11 out of the 14 children were interested in participating in the next part of the session (i.e. the imitation section). As for participants' attendance, in our experimental conditions, 78.6% of our volunteers took part in the imitation actions which could be a preliminary estimate

(a) (b)

Fig. 13 Mina and participants during the robot–child mode of phase 2: **a** an uncooperative participant, and **b** a cooperative subject

of the acceptability of the designed HRI platform. Alternatively, we have to consider that some of the children with ASD might not have a tendency to enter the room or interact with the robot. Although the subjects were involved in facial expression interaction with a robot for the first time and they voluntarily took part in the program, our prediction of the acceptability should be taken with caution before generalizing the obtained result to the autism population. In fact, the small number of the participants, different ages and perception of instructions, and the autism severe heterogeneity of the subjects might directly affect the robot's acceptability rate. Moreover, it is not possible to figure out whether the families had certain information about their children's willingness to interact with robots before attending our sessions. Participating in the Non-structured mode was our criteria for exploring the acceptability of the HRI architecture for children with ASD and the interaction data from this mode did not affect the performance analysis of the children in the Structured mode.

As the most important qualitative happenings of the Non-structured interaction mode, we observed some joint attention situations between some participants and their parents during the imitation session (for example when the robot winked at the subject). Moreover, some spontaneous initiation of verbal communications with the robot were observed from some of the children during the sessions. The mentioned observations could be referred to as social referencing behaviors which the autism clinicians would like to observe during treatment sessions. According to the one-session nature of this program as well as not conducting skill-based pre-Tests, we cannot claim whether the children had ever performed these behaviors before. The Mean Length of Utterance (MLU) of the participants were less than their typically developing peers. However, according to the video records and regarding the verbal communication, the following situations have been observed for some of the participants: (a) after pointing to the face of the robot by the mother, she asked her child: "What is Mina doing?", and the child answered: "head, . . . head, . . . head!" (the 3-year-old child assumed that the mother asked what part of the robot's body is.), (b) after the introduction of Mina, one of the participants asked the human-mediator: "What is your name?", (c) the human-mediator asked the child whether he wanted to start the session and he told Mina loudly: "start".

Next, without a break, the 11 cooperating children took part in two $\sim$ 5-min Structured interaction mode parts to do the facial imitation actions: once paired with the robot and one time paired with the human mediator. The order of the parts was counterbalanced. Fig. 13 show snapshots of the Structured interaction session in the robot–child mode for one of the participants with ASD.

4.2.2 Manual Assessment of the Imitation Actions

Figure 14a–d, show samples of the participants while they were imitating Mina. As previously mentioned, two video coders independently rated each of the 11 children's imitation performance for both robot–child and human mediator–child modes. The Pearson's correlation coefficient (r) of their scores were 0.924 (p value $=$ 0.000) which indicates a strong positive correlation between the two coders' scores; therefore, the mean of their scores was selected as the child's performance in each action.

Next, to investigate the research hypothesis, a two-way ANOVA analysis was performed to study whether the independent factors "Action Number" and "Imitation Mode" and their interaction have a statistically significant effect on the "Imitation Score" of the participants with ASD. The "Imitation Mode" contains 2 levels (i.e. the robot–child and human mediator–child modes) and the "Action Number" includes 8 levels (i.e. actions no. 1–8). To this end, the General Linear Model (GLM) tool of Minitab Software [51] was used to describe the statistical relationship between the two mentioned factors as well as their interaction. After fitting

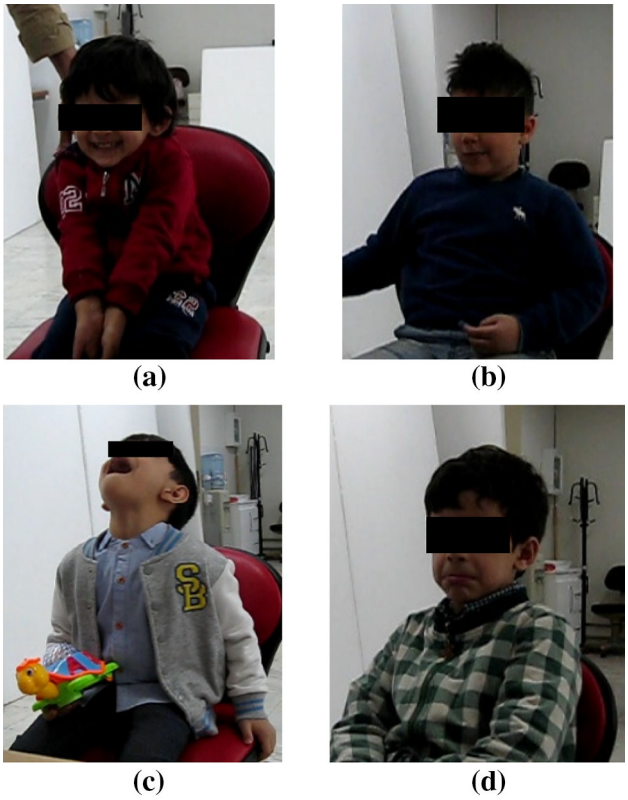

Fig. 14 Participants try to imitate Mina's facial gestures. **a** Action no. 1, **b** action no. 5, **c** action no. 6 and **d** action no. 7

Table 7 Analysis of Variance for imitation scores of the eleven collaborating children using Minitab software

Source	DOF	F-value	p value	Partial η^2
Factor1: Imitation Mode	1	11.81	0.001	0.069
Factor2: Action Number	7	8.45	0.000	0.270
Interaction of Imitation Mode and Action Number	7	0.75	0.632	0.032

a general linear model on the data, multiple comparisons between the factor level means and the significant differences were calculated (Table 7). The interaction plot for the mean of the manual scores is also presented in Fig.15.

According to Table 7, the p values for the "Imitation Mode" and "Action Number" are both less than 0.05, separately. The effect of both independent factors "Imitation Modes" and "Action Numbers" on the imitation scores are statistically significant. Therefore, it can be concluded that the performance of our participants with ASD in facial imitation actions in the human mediator–child mode was significantly better than the robot–child mode which means that our hypothesis is rejected. Moreover, in this study, no interaction was observed between the factors "Imitation Mode" and "Action Number" (p value $= 0.632 > 0.05$). However, unlike our observation, the researchers in [41,52] reported that their subjects with autism had a better performance in facial and emotional expressions tasks in the robot–child mode than the human-child mode. We believe that the main reason for the significant difference in the children' performance in the two modes in this study could be the robot's lack of verbal communication with the subjects during the actions; which is one of the differences between our setup and references [41,52]. However, we both faced similar restrictions of having a small number of participants and a non-homogenous group of subjects which makes it difficult to generalize our observations and confirm any strong claims.

As it was expected, the performance of the children directly depended on the number of involved facial components of the imitation tasks (i.e. the tasks' complexity). Based on Fig. 15, the overall mean scores of the children in actions no. 1–4 (with only one involved facial component) are higher than actions no. 5–8; and a decreasing trend in the scores is visible in both the robot–child and human mediator–child modes. Through the different nature and complexity of the designed imitation actions, we have reported the scores for each action separately. One may observe that the mean value of all actions' scores of each child would not be an accurate representation of the subject's performance.

As a qualitative viewpoint, all of the participants were experiencing their first interaction through facial expressions with a robot. The three common behaviors of the children we observed were: (1) three males and one female of the subjects had attempted to initiate verbal communication with Mina. Regardless of their imitation scores, it seemed that the robot is more attractive than the human mediator for these participants. (2) Two of the children spontaneously waved for Mina when they left the room. (3) Two other male participants were highly dependent on their mother's verbal encouragement to do the imitation actions.

4.2.3 Automatic Assessment of the Imitation Actions

The facial point' data was partially detected or not detected at all by the Kinect for 7 out of the 11 participants; because they had many unexpected and/or unpredictable major movements. Therefore, it was impossible to apply the automatic assessment algorithm for these 7 participants. In these cases, manual assessment seemed to be the appropriate method to rate their facial imitation performance. Therefore, this traditional and accurate method cannot be ignored or replaced. However, for (only) 4 of our high-functioning subjects, the Kinect was able to completely capture all 8 robot–child mode actions' facial data. Next, using the developed algorithm mentioned in Sect. 3.3.3, $(4 \times 8 =)$ 32 grades as automatic scores were calculated. In order to compare the manual and automatic scores, the Pearson's r on the equivalent grades for the mentioned four participants were computed as $r = 0.894$

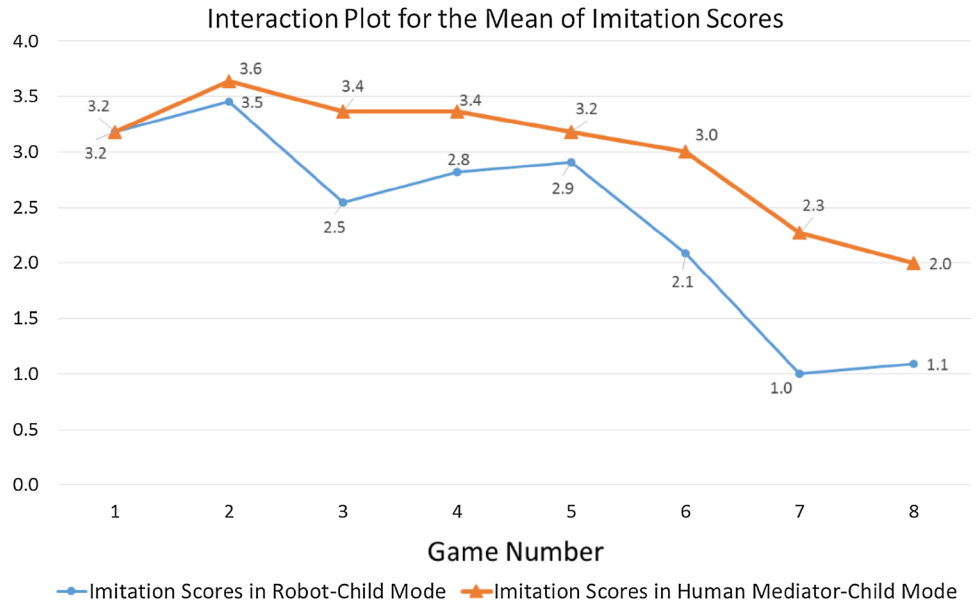

Fig. 15 Interaction plot for the mean imitation scores of the subjects with ASD (out of 4) in the robot–child and human mediator-child modes for each action in phase 2

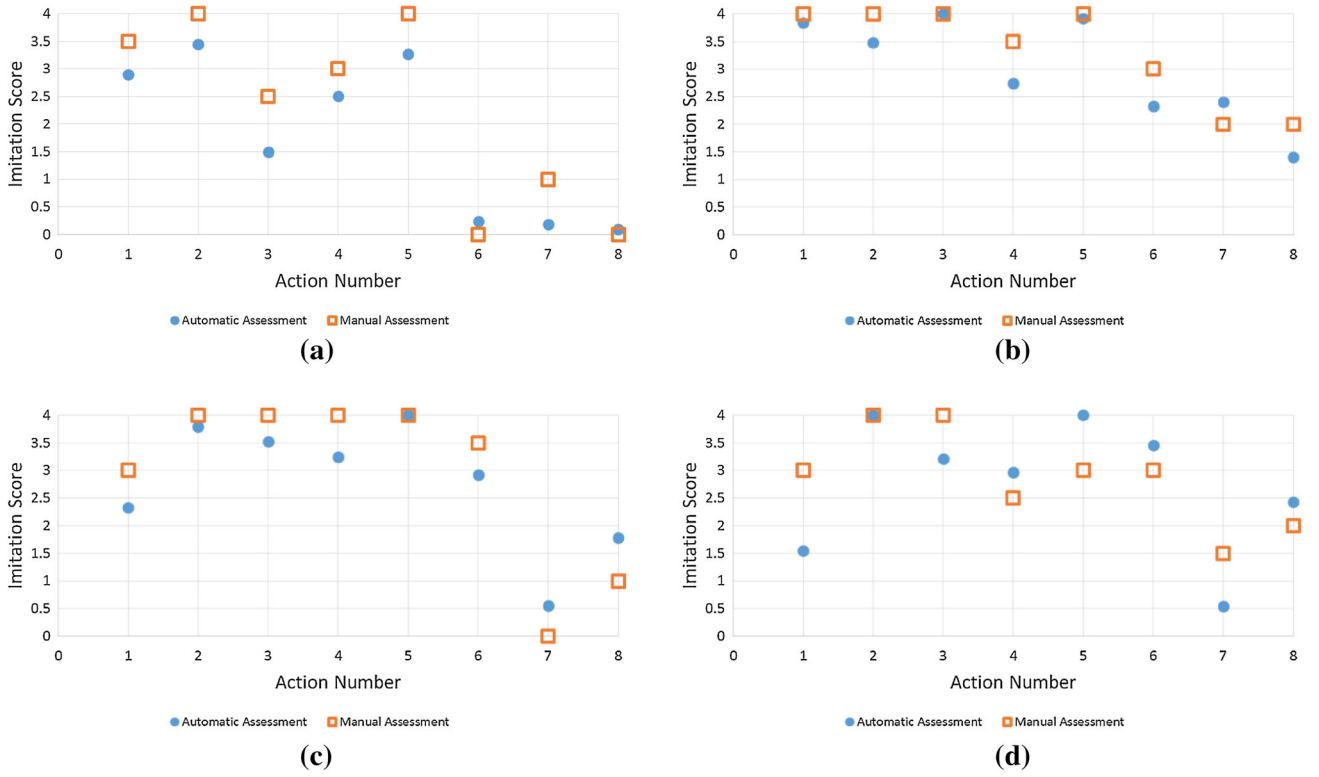

Fig. 16 Scatter plot for the automatic versus manual assessment for four of the participants with high-functioning in the robot–child mode actions (**a**–**d**). The imitation data for all 8 actions of these participants were successfully captured by the Kinect sensor

(p value $= 0.000$) which shows a sufficient agreement on the automatic and manual scores. Figure 16a–d show the scatter plots of the automatic versus manual scores for the mentioned four participants with high-functioning autism. By defining the absolute difference between the video coders' scores (as the reference scores) and the proposed automatic algorithm scores as the Error, we have observed the lowest and highest Error mean of 0.23 and 0.74 for action 2 and action 1, respectively. The mean and standard of deviation of the Errors for each action are presented in Table 8. Therefore, we can observe an acceptable performance for our developed platform in the automatic assessment process. In this study,

Table 8 The Error mean and standard deviation (SD) of the automatic scores in comparison to the manual scores

	Action 1	Action 2	Action 3	Action 4	Action 5	Action 6	Action 7	Action 8
Error mean	0.74	0.23	0.38	0.69	0.27	0.57	0.62	0.65
Error SD	0.53	0.22	0.33	0.15	0.49	0.09	0.24	0.17

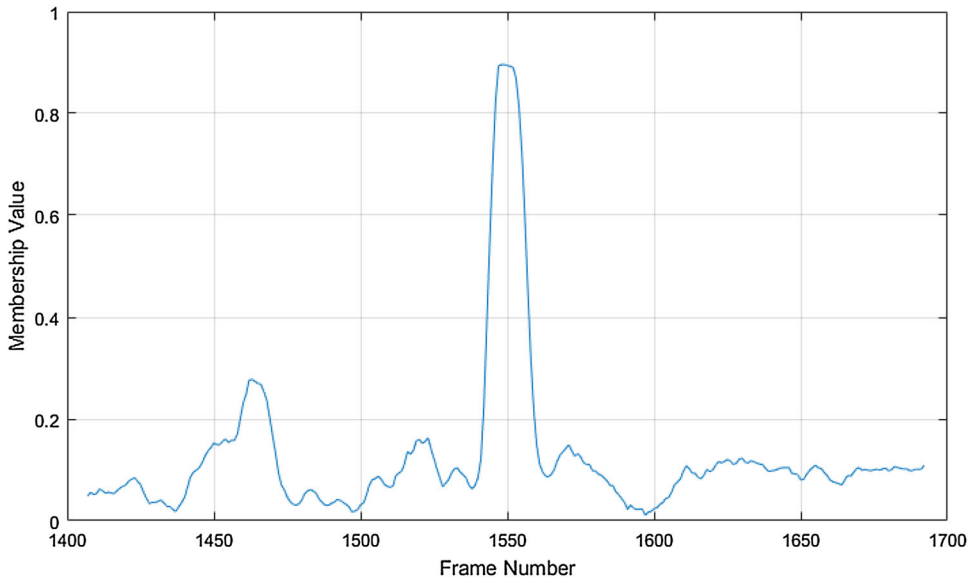

Fig. 17 The membership value of action no. 3 for one of the participants versus frame number (data captured at 30 frames per second)

the minimum and maximum possible values for the Error are 0 and 4, respectively.

Figure 17 shows a sample of membership value (related to action no. 3) for one of the participants versus time during the robot–child mode actions. According to Fig. 17, the automatic score for this participant in the third action is 3.52. The equivalent manual score by the video coders for this participant in action no. 3 was "4" (Fig. 16c), which is in line with the automatic assessment score. According to Fig. 16, we observe that in 29 score pairs out of 32 ($\sim$ 91% of the available data), the absolute difference between the manual and automatic assessment scores are less than 1 point. From another perspective, in 20 score pairs out of 32 (62.5% of the available data), the manual assessment score is higher than the paired automatic one. In fact, the psychologists have been trained to evaluate the imitation actions on the Likert scale and they usually decide to rate children's performance based on their experience and cognitive knowledge. On the other hand, in the proposed calculation-based automatic assessment algorithm, the imitation score of the participants is the multiple of the membership value resulted from the FCM algorithm output and it gives continuous scores in the interval 0 to 4. Therefore, according to the high correlation of the manual and automatic scores, we can conclude that the presented classification algorithm seems to work appropriately. Now, we can answer the research question of the study in this way: the proposed fuzzy algorithm can be used for automatic assessment of facial imitation. All in all, while "manual" imitation assessment is more accurate and reliable, it is a laborious and time-consuming task that could be handled by HRI platforms.

5 Limitations and Future Works

As it was mentioned earlier, such reciprocal platforms have the potential for cognitive rehabilitation of children with autism. Regarding the preliminary acceptance rate of the current HRI as well as our qualitative observations, the next step of our study is to design and conduct a set of therapeutic intervention sessions for children with ASD. Moreover, the case study approach of this research and the small number of the participants with diverse autism severity make it difficult to generalize the findings; therefore, it is recommended that the developed platform could be run with more population and compared to the typical reciprocal imitation training programs in order to find out the exact advantages and disadvantages of the robot-assisted platforms. In our experimental conditions, for all of the subjects, one of the parents was also present in the class which could possibly affect his/her child's cooperation. In future scenarios that we envision, similar to the regular autism therapy, parents do not need constantly present with their ASD children during the robot-assisted intervention sessions.

As another limitation, we did not have the male version of a humanoid robot with expressive face in order to investigate

whether children with autism interact differently with a robot with opposite gender. There were also some restrictions on the number and location of the active DOF in Mina's face which made it difficult to design distinct facial gestures in the Structured interaction mode. In addition, the Mina robot did not have verbal communication ability in this study which could have negative effects on the children's perception of instruction from the robot. In this regard, the ability of verbal communication could be added to the current HRI platform in the future.

Moreover, enriching the facial expression' database by gathering approved data from typically developing children is highly recommended to be done in the future. Regarding the automatic assessment, the proposed algorithm worked for the posed gestures, but temporal algorithms were needed for gestures including dynamic movements. The main limitations of the used algorithm for automatic assessment are: (1) facial gestures should be designed in a way that they can be classified through the mentioned facial features, and (2) participants should obtain the minimum instruction perception level to stay focused on the robot during the actions.

An ultimate goal for the future of this research is to empower the reciprocal HRI architecture to imitate the entire body and gesture movements, vocalizations, and even participants' actions with objects. Our hope is such studies in autism treatment reduce the applicable costs in Iran [53–55].

6 Conclusion

In general, the emotional state is a combination of two or more basic emotions. To have better HRI, detecting the share of each basic emotion in the users' current emotional state is considered valuable. The method used in this study made it possible to assign a membership value to the facial expression of the user, meaning that the user's emotional state could be related to more than one basic emotional state. In addition, two methods for facial feature extraction were discussed and basic emotions were recognized with an overall accuracy of more than 90% for 5 out of 6 basic emotions. Then, the identified facial expression was given to a finite state machine developed for emotional interaction. To expose the proper facial expression, Mina was programmed to turn her head to face the user. Finally, the HRI system was shown to be capable of producing a combinatorial facial expression output. Some extra actions were designed for some of the states and would take place if Mina stayed in that state for the specified time. The system was also able to select and generate different facial expressions with variable intensities.

Furthermore, a facial expression imitation assessment for children with Autism Spectrum Disorder (ASD) was compiled. Then, different facial expressions were produced by Mina and a human-mediator in two different modes while the children were asked to imitate their facial gestures through the Structured interaction mode. It was observed that the performance of the participants with ASD in the human mediator–child mode was significantly better than the robot–child mode. Moreover, for the automatic facial expression imitation assessment, the same FCM method is used with a different database. Unfortunately, it was not possible to use the recorded Kinect data for 7 participants with ASD since they had many unexpected major movements. In these cases, manual assessment seems to be the best possible method. We observed that the automatic assessment algorithm worked for the designed posed (i.e. non-dynamic) gestures with high correlation to manual scores.

The proposed HRI platform could be used as a facilitator in autism treatment. However, in this stage, attendance of child psychologists in the robot-assisted intervention sessions is considered as an essential issue. Besides taking advantage of Mina's attractive appearance, handling the unpredicted situations of autism classes needs the knowledge, expertise, and experience of psychologists, the qualities robotic platforms do not have.

Acknowledgements Our profound gratitude goes to the "Center for the Treatment of Autistic Disorders (CTAD)" and its psychologists for their contributions to the clinical trials with the children with autism. This research was funded by the "Cognitive Sciences and Technology Council" (CSTC) of Iran (http://www.cogc.ir/). We also appreciate the Iranian National Science Foundation (INSF) for their complementary support of the Social & Cognitive Robotics Laboratory (http://en.insf.org/).

Compliance with Ethical Standard

Funding This study was funded by the "Cognitive Sciences and Technology Council" (CSTC) of Iran (Grant Number: 95p22)

Conflict of interest Author Ali Meghdari has received research grants from the "Cognitive Sciences and Technology Council" (CSTC) of Iran. The authors Ali Ghorbandaei Pour, Alireza Taheri, and Minoo Alemi declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Ethical approval for the protocol of this study was provided by Iran University of Medical Sciences (No. IR.IUMS.REC.1395.95301469), and the certification for ABA and robot-assisted Therapy with autistic children was received from the Center for the Treatment of Autistic Disorders (CTAD), Iran.

References

1. Pantic M, Pentland A, Nijholt A, Huang TS (2007) Human computing and machine understanding of human behavior: a survey. In: Huang TS, Nijholt A, Pantic M, Pentland A (eds) Artificial intelligence for human computing. Lecture notes in computer science, vol

4451. Springer, Berlin. https://doi.org/10.1007/978-3-540-72348-6_3
2. Valstar MF (2008) Timing is everything: a spatio-temporal approach to the analysis of facial actions. Imperial College London, London
3. Mavridis N (2015) A review of verbal and non-verbal human–robot interactive communication. Robot Auton Syst 63:22–35
4. Tardif C, Lainé F, Rodriguez M, Gepner B (2007) Slowing down presentation of facial movements and vocal sounds enhances facial expression recognition and induces facial-vocal imitation in children with autism. J Autism Dev Disord 37(8):1469–1484
5. Dawson G, Webb SJ, McPartland J (2005) Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. Dev Neuropsychol 27(3):403–424
6. Baron-Cohen S, Leslie AM, Frith U (1985) Does the autistic child have a "theory of mind"? Cognition 21(1):37–46
7. Baron-Cohen S (2001) Theory of mind in normal development and autism. Prisme 34(1):74–183
8. Haviland JM, Lelwica M (1987) The induced affect response: 10-week-old infants' responses to three emotion expressions. Dev Psychol 23(1):97
9. Tonks J, Williams WH, Frampton I, Yates P, Slater A (2007) Assessing emotion recognition in 9–15-years olds: preliminary analysis of abilities in reading emotion from faces, voices and eyes. Brain Inj 21(6):623–629
10. Pouretemad H (2011) Assessment and treatment of joint attention deficits in children with autistic spectrum disorders. Arjmand Book, Tehran (**in Persian**)
11. Ingersoll B (2010) Brief report: pilot randomized controlled trial of reciprocal imitation training for teaching elicited and spontaneous imitation to children with autism. J Autism Dev Disord 40(9):1154–1160
12. Alemi M, Meghdari A, Ghazisaedy M (2015) The impact of social robotics on L2 learners' anxiety and attitude in English vocabulary acquisition. Int J Soc Robot 7(4):523–535
13. Tamura T, Yonemitsu S, Itoh A, Oikawa D, Kawakami A, Higashi Y et al (2004) Is an entertainment robot useful in the care of elderly people with severe dementia? J Gerontol Ser Biol Sci Med Sci 59(1):M83–M85
14. Alemi M, Ghanbarzadeh A, Meghdari A, Moghadam LJ (2016) Clinical application of a humanoid robot in pediatric cancer interventions. Int J Soc Robot 8(5):743–759
15. Taheri A, Alemi M, Meghdari A, Pouretemad H, Basiri NM, Poorgoldooz P (2015) Impact of humanoid social robots on treatment of a pair of Iranian autistic twins. In: International conference on social robotics. Springer, pp 623–632
16. Scassellati B, Admoni H, Matarić M (2012) Robots for use in autism research. Annu Rev Biomed Eng 14:275–294
17. Taheri A, Meghdari A, Alemi M, Pouretemad H, Poorgoldooz P, Roohbakhsh M (2016) Social robots and teaching music to autistic children: myth or reality? In: International conference on social robotics. Springer, pp 541–550
18. Hopkins IM, Gower MW, Perez TA, Smith DS, Amthor FR, Wimsatt FC, Biasini FJ (2011) Avatar assistant: improving social skills in students with an ASD through a computer-based intervention. J Autism Dev Disord 41(11):1543–1555
19. Feil-Seifer D, Mataric MJ (2008) B 3 IA: a control architecture for autonomous robot-assisted behavior intervention for children with Autism Spectrum Disorders. In: The 17th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 328–333
20. Meghdari A, Alemi M, Pour AG, Taheri A (2016) Spontaneous human–robot emotional interaction through facial expressions. In: International conference on social robotics. Springer, pp 351–361
21. Zacharatos H, Gatzoulis C, Chrysanthou YL (2014) Automatic emotion recognition based on body movement analysis: a survey. IEEE Comput Graph Appl 34(6):35–45
22. Xiao Y, Zhang Z, Beck A, Yuan J, Thalmann D (2014) Human–robot interaction by understanding upper body gestures. Presence Teleoper Virtual Environ 23(2):133–154
23. Aly A, Tapus A (2015) Multimodal adapted robot behavior synthesis within a narrative human–robot interaction. In: IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 2986–2993
24. Kwon DS, Kwak YK, Park JC, Chung MJ, Jee ES, Park KS et al (2007) Emotion interaction system for a service robot. In: The 16th ieee international symposium on robot and human interactive communication (RO-MAN), pp 351–356
25. Brown L, Howard AM (2014) Gestural behavioral implementation on a humanoid robotic platform for effective social interaction. In: The 23rd IEEE international symposium on robot and human interactive communication (RO-MAN), pp 471–476
26. Noh JY, Neumann U (1998) A survey of facial modeling and animation techniques. USC technical report, pp 99–705
27. Mavadati S (2015) Spontaneous facial behavior computing in human machine interaction with applications in autism treatment. Doctoral dissertation, Electrical and Computer Engineering Department, University of Denver, Denver
28. Halder A, Konar A, Mandal R, Chakraborty A, Bhowmik P, Pal NR, Nagar AK (2013) General and interval type-2 fuzzy face-space approach to emotion recognition. IEEE Trans Syst Man Cybern Syst 43(3):587–605
29. Dahmane M, Meunier J (2014) Prototype-based modeling for facial expression analysis. IEEE Trans Multimed 16(6):1574–1584
30. Kotsia I, Pitas I (2007) Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Trans Image Process 16(1):172–187
31. Li Y, Wang S, Zhao Y, Ji Q (2013) Simultaneous facial feature tracking and facial expression recognition. IEEE Trans Image Process 22(7):2559–2573
32. Holthaus P, Wachsmuth S (2013) Direct on-line imitation of human faces with hierarchical ART networks. In: The 22nd IEEE international symposium on robot and human interactive communication (RO-MAN), pp 370–371
33. Li Y, Mavadati SM, Mahoor MH, Zhao Y, Ji Q (2015) Measuring the intensity of spontaneous facial action units with dynamic Bayesian network. Pattern Recogn 48(11):3417–3427
34. Chakraborty A, Konar A, Chakraborty UK, Chatterjee A (2009) Emotion recognition from facial expressions and its control using fuzzy logic. IEEE Trans Syst Man Cybern Part A Syst Hum 39(4):726–743
35. Abdat F, Maaoui C, Pruski A (2011) Human–computer interaction using emotion recognition from facial expression. In: Fifth UKSim european symposium on computer modeling and simulation (EMS), pp 196–201
36. de Carvalho Santos V, Romero RAF, Coca SRDM (2012) Imitation of facial expressions for a virtual robotic head. In: Robotics symposium and latin american robotics symposium (SBR-LARS), 2012 Brazilian, pp 251–254
37. Cid F, Prado JA, Bustos P, Nunez P (2013) A real time and robust facial expression recognition and imitation approach for affective human–robot interaction using gabor filtering. In: IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 2188–2193
38. Chumkamon S, Masato K, Hayashi E (2014) The robot's eye expression for imitating human facial expression. In: The 11th international conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON), pp 1–5

39. Meghdari A, Shouraki SB, Siamy A, Shariati A (2016) The real-time facial imitation by a social humanoid robot. In: The 4th international conference on robotics and mechatronics (ICROM), pp 524–529
40. Tanaka JW, Wolf JM, Klaiman C, Koenig K, Cockburn J, Herlihy L et al (2010) Using computerized games to teach face recognition skills to children with autism spectrum disorder: the let's face it! program. J Child Psychol Psychiatry 51(8):944–952
41. Duquette A, Michaud F, Mercier H (2008) Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. Auton Robots 24(2):147–157
42. Salvador MJ, Silver S, Mahoor MH (2015) An emotion recognition comparative study of autistic and typically-developing children using the zeno robot. In: IEEE international conference on robotics and automation (ICRA), pp 6128–6133
43. Wainer J, Robins B, Amirabdollahian F, Dautenhahn K (2014) Using the humanoid robot KASPAR to autonomously play triadic games and facilitate collaborative play among children with autism. IEEE Trans Auton Ment Dev 6(3):183–199
44. Hanson D, Mazzei D, Garver C, Ahluwalia A, De Rossi D, Stevenson M, Reynolds K (2012) Realistic humanlike robots for treatment of ASD, social training, and research; shown to appeal to youths with ASD, cause physiological arousal, and increase human-to-human social engagement. In: Proceedings of the 5th ACM international conference on pervasive technologies related to assistive environments (PETRA'12)
45. Kinect for Windows SDK (2016) https://msdn.microsoft.com/en-us/library/
46. http://www.robokindrobots.com/ (2016)
47. Ekman P, Friesen W (1978) Facial action coding system: a technique for the measurement of facial movement. Consulting Psychologists, Palo Alto
48. Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10(2–3):191–203
49. Popescu M, Keller J, Bezdek J, Zare A (2015) Random projections fuzzy c-means (RPFCM) for big data clustering. In: IEEE international conference on fuzzy systems (FUZZ-IEEE), pp 1–6
50. Yan J, Ryan M, Power J (1994) Using fuzzy logic: towards intelligent systems, vol 1. Prentice Hall, Upper Saddle River
51. Minitab INC (2000) MINITAB statistical software. Minitab Release, 13
52. Giannopulu I, Montreynaud V, Watanabe T (2014) PEKOPPA: a minimalistic toy robot to analyse a listener-speaker situation in neurotypical and autistic children aged 6 years. In: Proceedings of the second international conference on human-agent interaction. ACM, pp 9–16
53. Taheri A, Meghdari A, Alemi M, Pouretemad H (2017) Human-robot interaction in autism treatment: a case study on three pairs of autistic children as twins, siblings, and classmates. Int J Soc Robot. https://doi.org/10.1007/s12369-017-0433-8
54. Taheri A, Meghdari A, Alemi M, Pouretemad H (2017) Teaching music to children with autism: a social robotics challenge. Int J Sci Iran Trans G Socio Cognit Eng. https://doi.org/10.24200/SCI.2017.4608
55. Elahi MT, Korayem AH, Shariati A, Meghdari A, Alemi M, Ahmadi E et al (2017) "Xylotism": a tablet-based application to teach music to children with autism. In: International conference on social robotics. Springer, Cham, pp 728–738

Ali Ghorbandaei Pour was born in Tehran, Iran. He received the B.Sc. in Electrical Engineering from University of Tehran and his M.Sc. in Mechatronics from Sharif University of Technology in Iran. His focus is on Robotics, Machine Learning, and Automation. Since 2016, he has been a member of Social and Cognitive Robotics Laboratory at Sharif University of Technology.

Alireza Taheri received his Ph.D. in Mechanical Engineering with emphasis on Social Robotics from Sharif University of Technology, Tehran, Iran. During 2015–2017, he spent a total of 1-year sabbatical as a research scholar at Yale University and University of Denver, USA.

Minoo Alemi received her Ph.D. in Applied Linguistics (TEFL) from Allameh Tabataba'i University in 2011. She is an Assistant Professor and Division Head of Applied Linguistics in Islamic Azad University, West Tehran Branch. She is the founder of Robot-Assisted Language Learning (RALL), and the co-founder of Social Robotics in Iran which she achieved as a Post-Doctoral research associate at the Social and Cognitive Robotics Laboratory of Sharif University of Technology. Her areas of interest include discourse analysis, interlanguage pragmatics, social robotics, and RALL. Dr. Alemi has been the recipient of various teaching and research awards from Sharif University of Technology, Allameh Tabataba'i University, Islamic Azad University, and Int. Conf. on Social Robotics (ICSR2014), Sydney-Australia. She has published over 75 papers and books in refereed national and international conferences, and journals. She is also on the editorial boards of the British Journal of Educational Technology (BJET), and the *Scientia-Iranica*: Transactions on Socio-Cognitive Engineering.

Ali Meghdari is a Professor of Mechanical Engineering at Sharif University of Technology (SUT) in Tehran. Professor Meghdari has performed extensive research in various areas of robotics; social and cognitive robotics, mechatronics, bio-robotics, and modelling of biomechanical systems. He has been the recipient of various scholarships and awards, the latest being: the 2012 Allameh Tabatabaei distinguished professorship award by the National Elites Foundation of Iran (BMN), the 2001 Mechanical Engineering Distinguished Professorship Award from the Ministry of Science, Research and Technology (MSRT) in Iran, and the 1997 ISESCO Award in Technology from Morocco. He is currently the Director of the Center of Excellence in Design, Robotics and Automation (CEDRA), an affiliate member of the Iranian Academy of Sciences (IAS), and a *Fellow* of the American Society of Mechanical Engineers (ASME). He is also the Editor of Transactions on Socio-Cognitive Engineering, *Scientia-Iranica International Journal*.