



Influence of Reaction Time in the Emotional Response of a Companion Robot to a Child's Aggressive Interaction

Ahmad Yaser Alhaddad^{1,2} · John-John Cabibihan¹ · Andrea Bonarini²

Accepted: 18 January 2020 / Published online: 20 February 2020
© The Author(s) 2020

Abstract

The quality of a companion robot's reaction is important to make it acceptable to the users and to sustain interactions. Furthermore, the robot's reaction can be used to train socially acceptable behaviors and to develop certain skills in both normally developing children and children with cognitive disabilities. In this study, we investigate the influence of reaction time in the emotional response of a robot when children display aggressive interactions toward it. Different interactions were considered, namely, pickup, shake, drop and throw. The robot produced responses as audible sounds, which were activated at three different reaction times, namely, 0.5 s, 1.0 s, and 1.5 s. The results for one of the tasks that involved shaking the robotic toys produced a significant difference between the timings tested. This could imply that producing a late response to an action (i.e. greater than 1.0 s) could negatively affect the children's comprehension of the intended message. Furthermore, the response should be comprehensible to provide a clear message to the user. The results imply that the designers of companion robotic toys need to consider an appropriate timing and clear modality for their robots' responses.

Keywords Companion robots · Machine learning · Autism spectrum disorders · Children · Robots for therapy · Meltdown · Long short-term memory network (LSTM)

1 Introduction

The recent advances in robotics accelerated the integration of robots to new areas, such as in healthcare. More specifically, social robots or rehabilitation robots are being developed to monitor and improve health, to assist with difficult tasks, and to prevent the declining of one's health [48]. Assisting in therapy is an application of robots in healthcare that has shown a promising potential. For example, social robots were found to be effective in improving the outcomes of therapy sessions, especially among children with autism [21,54].

Aggression is a behavior that is done by a living agent, such as a human or an animal, that causes harm and violates the rights of others [23]. The American Psychological Association (APA) defines aggression as a behavior that is aimed

at hurting others either physically or psychologically [8]. APA categorizes aggression as hostile aggression, which is intended to cause harm; instrumental aggression, which is not intended to cause harm; and affective aggression, which is emotionally motivated toward the source of distress. The frequency of physical aggression among children was reported to peak during the years before school [41]. Kicking, biting, and hitting are examples of the physical aggressive behaviors that might occur during the early years of childhood [7]. Aggression among children is considered as one of the most common reasons for mental health referrals [56]. The occurrence of aggression or disruptive behavior was reported to be higher among children with psychiatric disorders. For example, the prevalence rate of such behaviors was reported to reach 62.3% among children with anxiety disorders, while it could reach 45.8% among those with mood disorders [42].

Considering all the children with or without developmental disabilities, challenging behaviors appear to have higher prevalence rates among those affected by autism spectrum disorder (ASD) [26,29]. Even within the spectrum itself, those with severe autism have displayed challenging behaviors at a higher rate as compared to those with less ASD severity [39,40]. Even in their infancy, children with

✉ John-John Cabibihan
john.cabibihan@qu.edu.qa

¹ Department of Mechanical and Industrial Engineering, Qatar University, 2713 Doha, Qatar

² Department of Electronics, Information and Bioengineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy

autism have displayed more challenging behaviors as compared to their neuro-typical peers [27,34]. Previous studies reported high prevalence rates of challenging behaviors (e.g. 49–69% [12,15,32]). Aggression against others, meltdowns, tantrums, withdrawal, and stereotyped behaviors are some of the forms of the challenging behaviors that are exhibited by children on the spectrum [31,37,38]. These behaviors pose a risk on themselves and others around them, such as family members, companions, and care givers [31,46]. The mitigation of challenging behaviors is possible with early intervention [49].

The current progress in technology is offering new improvements to intervention and therapy sessions, such as hands-on learning, independent learning, individualizing, and others [25]. The interest in integrating social robots into therapy is increasing due to the reported evidence of their efficacy [21,52]. However, the presence of social robots could pose a risk during the exhibition of challenging behaviors, such as throwing objects, hitting, banging on objects, and kicking objects [38]. Children showed some aggression toward the robots as reported in previous studies [6,14,19]. For smaller companion robots, children might pick it up and mishandle it (Fig. 1). The throwing of such objects (i.e.

small robotic toys) might hit another person's head and cause harm [2]. Due to safety and legal concerns, robot designs must account for such scenarios and adopt new methods and ways to mitigate any potential harm [4,5,18,22,58].

Social robots represent a new type of stimuli that are meant to elicit behaviors and initiate interactions, and that might trigger unwanted ones. To date, the studies to characterize the unwanted and aggressive interactions are limited [14,35,51]. Additionally, limited work has been done to investigate the proper reactions once such behaviors are detected [3]. The ability of a robot to detect and respond to unwanted interactions will provide many benefits, such as the prevention of potential harm, monitoring, promoting safety culture, and to prevent the progression of an aggressive behavior [19]. Furthermore, it could be used as a therapeutic tool to address aggressive behaviors.

In this study, we investigate the effects of reaction time and sound modality employed in robotic toys on the perceived perception by children interacting with the robots. A recognition architecture based on Long Short-term Memory Cell (LSTM) was adopted to classify the behaviors based on the acceleration data received. Different reactions with different timings were produced once a pickup, a shake, a

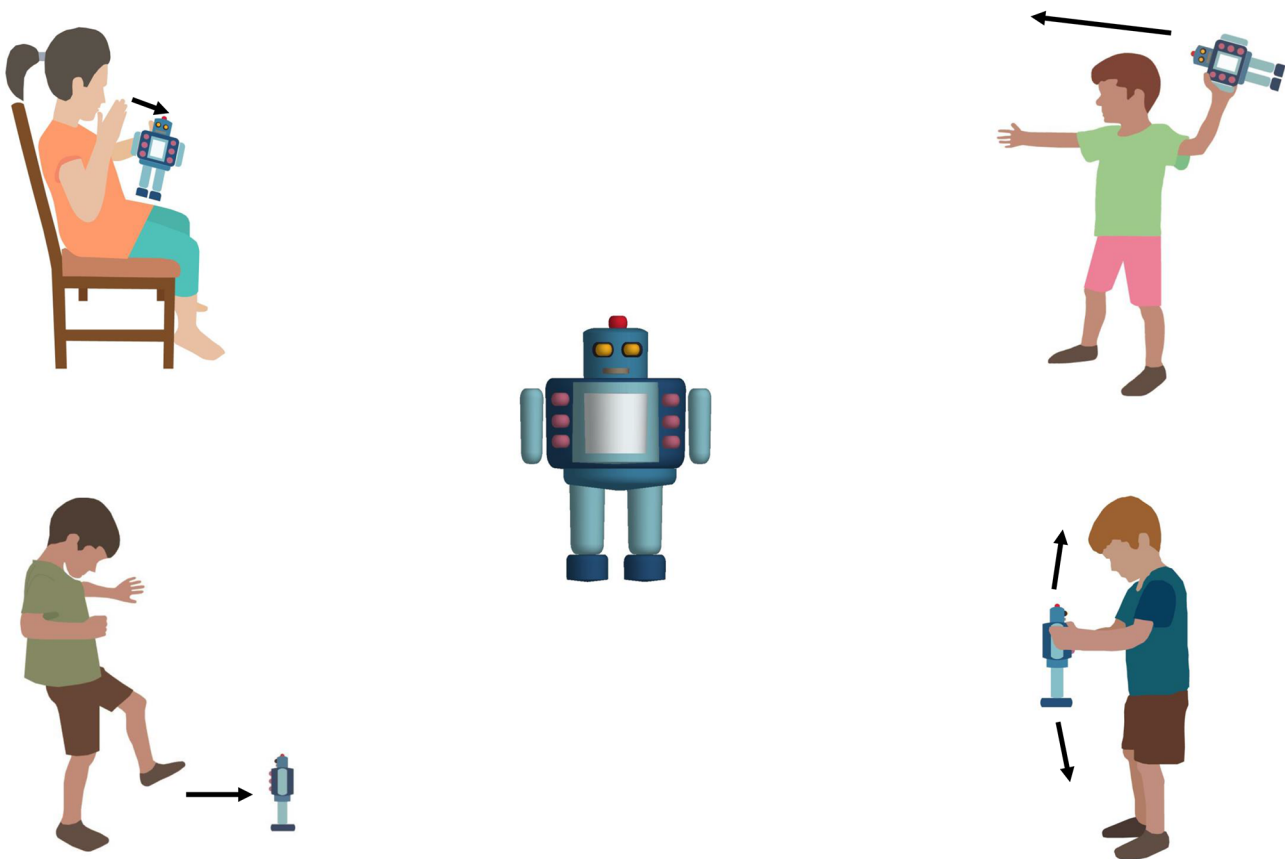


Fig. 1 Some of the unwanted and aggressive interactions that might be exhibited by children toward a companion robotic toy

drop or a throw was detected. This paper is organized as follows. Section 2 presents the background. Section 3 describes materials and methods. Section 4 provides results and Sect. 5 discusses the results.

2 Background

Species in nature offer a lot of biologically-inspired concepts and ideas to roboticists. One of these mechanisms is the reflex system that can be adopted in the design and development of robots [1]. Reflexes are meant to ensure the survival of the living organism externally while ensuring the balance of operations internally. Reaction to a stimulus is usually carried out by the reflex arc that consists of several stages, namely, arrival of stimulus, activation of sensory neuron, information processing, motor neuron activation, and peripheral effector response. The implementation of reflexes in a robotic system should operate without affecting the main objectives of the robot (Fig. 2). Once an unwanted interaction is detected, the robot may respond with the appropriate reaction to deliver the corresponding message to the user [19]. The timing of the reaction and its modality should be felt as natural to provide a clear implication about the interaction.

Few robots were developed that demonstrate some reactions to a human interactions. PARO is one of the commercially available robots that reacts to physical interactions [53]. PARO is a seal-looking interactive therapeutic toy that is covered with white fur and emits voices similar to that of a baby seal. Different embedded sensors enable PARO to interact with its environment. The light sensor enables it to recognize dark and light. The audio sensor gives PARO the ability to recognize the direction of voices. The tactile sensor gives PARO the ability to feel any stroke or pressure. PARO

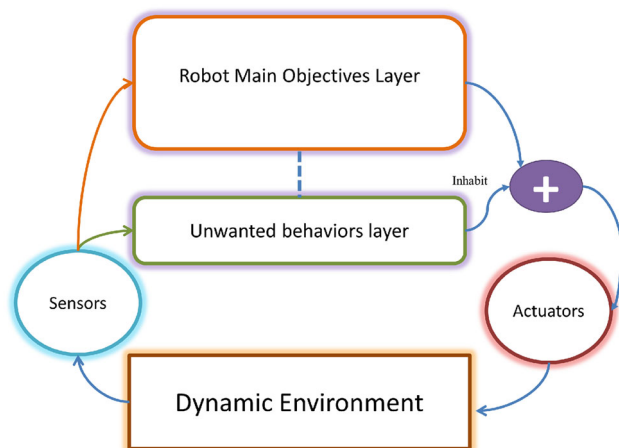


Fig. 2 The proposed reflex model to respond to unwanted interactions. A layer to detect the unwanted interactions will temporarily inhibit the system to produce an appropriate response

interacts with people by making sounds and moving some parts of its structure, such as the head, paddle and eyelid.

Roball is another robot that was developed to react to certain physical interactions [51]. The robot is shaped like a ball with a diameter of 0.27 m and weighs around 2 kgs. It is equipped with accelerometers and tilt sensors that allows it to interact and navigate in its environment. Based on the sensors' readings, several interaction modes are possible, such as being alone, general interaction, being carried, and being spun.

Teo is a mobile soft robot, which was developed to interact with children with ASD [16]. It can sense distance and touch, and it can distinguish different dynamic interactions, like hug, push, punch, getting close, and others. Based on the interpretation of sensory data, the robot can react with sounds, words, movements, and coloured lights.

Different sensors and wearable devices were considered in human activity recognition research [10,20]. A frequently used sensor is the accelerometer, which is a relatively low-cost sensor that is able to detect acceleration on three orthogonal directions. When associated to a gyroscope, the rotational speed can be detected along the same axis. One of the earliest works classifying different daily physical activities, such as walking and running, used five wearable small accelerometers on different body parts of 20 participants [13]. The data collected were from subjects performing a sequence of different daily tasks. The best classifier selected (i.e. a decision tree) was able to recognize the actions with an accuracy rate of 84%. Another study considered using acceleration and sound data to recognize workshop related activities to develop a proactive system [36]. The data collected were based on tasks performed in a wood shop. The system was able to recognize different activities with an accuracy of 84.4% on continuous simulated stream of data. Nowadays, accelerometers are used in smart phones to detect a wide range of activities [24].

Accelerometers were also considered in devices that detect the fall of the elderly [11]. One study considered a wearable device that contains an accelerometer to detect falls [55]. To facilitate the therapy for those with special needs, one study considered using accelerometers to detect problem behaviors among this population [47]. In this study, the data to develop the recognition model were simulated by trained clinical staff. Their approach was able to achieve an accuracy of 69.7% when evaluated with realistic data. For more advanced and interactive applications, accelerometers were considered in robot games to model players and recognize activities [43,44]. One study considered using a tri-axial accelerometer module embedded in a player's chest to acquire the motion data [45]. Their work showed promising results in detecting different activities with the robot, such as running, walking or dodging, and blocking the robot's path.

3 Materials and Methods

In this section, the adopted methods and approaches to conduct the investigation in our experiments are presented. The section starts with the model by describing the recognition architecture, data format, and the evaluation of the model. Then, we proceed to the experimental setup that describes the robotic toys, recognition device, and the employed reactions. Finally, we present the participants, the evaluation of reactions, and the data analysis methods.

3.1 The Model

3.1.1 Recognition Architecture

The recognition network that was adopted in our work was proposed by an earlier study that relied on Long Short-term Memory network (LSTM) in combination with bidirectional and residual connections [61]. In their proposed model, the network was able to produce improved results (i.e. 93.5%) on the public domain (i.e. UCI Machine Learning Repository) dataset on human activity recognition as compared to other configurations [9]. We considered that the recognition problem in our study would benefit from this network due to the similarity in the characteristics of the activities that needs to be recognized. In this section, we provide a brief description about this recognition network.

LSTM network is a special structure based on a Recurrent Neural Network (RNN) that is used to process a data stream. In RNN, the prediction depends on the history information that is maintained within the internal memory of the network. A typical RNN consists of three layers, namely, an input layer x , a hidden layer h , and an output layer y . The relations among these layers are defined as follows:

$$h(t) = f(Ux(t) + Wh(t-1)) \quad (1)$$

$$y(t) = g(Vh(h)) \quad (2)$$

where U is the connection weights matrix from the input layer to the hidden layer, W is the connection weights matrix within the hidden layers, and V is the connection weights matrix between the last hidden layer and the output. Furthermore, f and g represent the activation functions.

Compared to standard RNN structure, LSTM showed stability and powerful performance in the modeling of long sequences (e.g. [57]). The structure of LSTM is unique due to a memory cell c_t that accumulates the state information [60]. Furthermore, this structure allows one to deal with the vanishing gradients problem [30]. The LSTM cell contains three controlling gates, namely, input gate, forget gate, and output gate (Fig. 3). These gates control what information that should be kept, updated, or forgotten. More complex structures can be formed by combining multiple LSTM cells. The

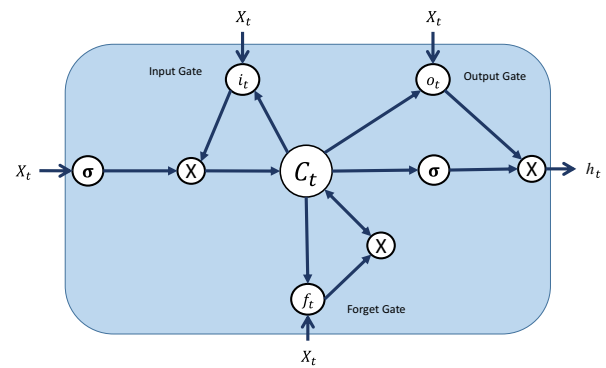


Fig. 3 A graphical representation of the Long Short-term Memory (LSTM) cell. The LSTM cell consists of three gates, namely, the input gate i , the output gate o , and the forget gate f . These gates control the information within the cell

internal parameters of an LSTM cell are defined as follows [28]:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (4)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (6)$$

$$h_t = o_t \tanh(c_t) \quad (7)$$

where i is the input gate, f is the forget gate, o is the output gate, σ is the logistic sigmoid function, and c is the cell activation vectors.

The recognition network also made use of bidirectional LSTM due to its advantages over standard LSTM. For example, the output of bidirectional LSTM is related to previous and subsequent information, hence, a better overall performance. The output of the proposed algorithm is determined by concatenating the results of the forward and backward sequences through a hidden layer that reduces the number of features [61]. Finally, the algorithm uses a residual network that provide different advantages, such as efficient training and easier optimization.

3.1.2 Data Format

The data that were used in training and testing the recognition model were acquired from an earlier study [3]. The data for the acceleration were in the form of the resultant acceleration computed as the square root of the sum of the squares of the individual accelerations. The relation is defined as follows:

$$|A| = \sqrt{A_x^2 + A_y^2 + A_z^2} \quad (8)$$

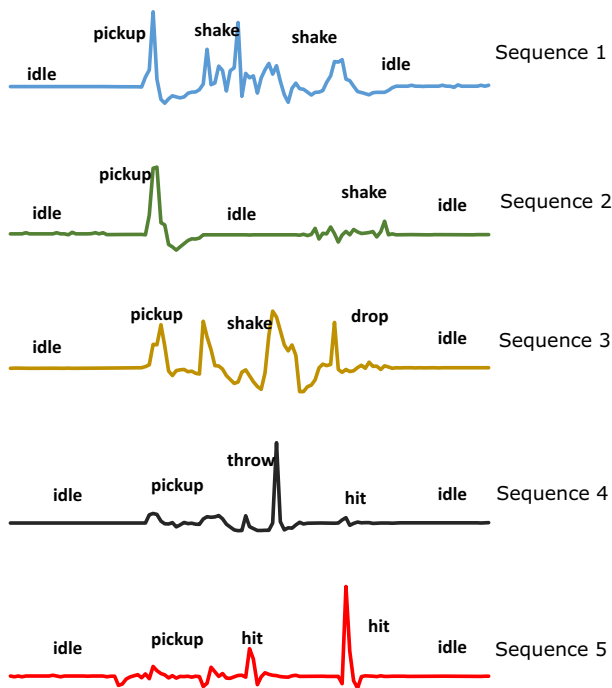


Fig. 4 Five samples of the artificially created sequences from the data samples obtained from an earlier study [3]. The sequences were selected based on their likelihood of occurring in realistic scenarios. The behaviors in the sequences were obtained randomly from the available pool of samples from each participant

where A_x represents the acceleration in the X axis, A_y represents the acceleration in the Y axis, and A_z represents the acceleration in the Z axis.

The training data were acquired from adult participants performing the behaviors of interest while the test data were acquired from children participants. To create a temporal data stream from these discrete data samples, artificial sequences were created from the data samples randomly (Fig. 4). The sequences were selected based on the likelihood of their occurrence in realistic interaction scenarios. This approach will support the creation of more variability in the data and decrease subject-dependent learning. For example, a sequence could contain samples from any of the participants and from any of the robotic toys used. This procedure was applied to both the training and testing data.

3.1.3 Model Evaluation

Three training parameters were tested to identify the model with the most promising results. The tested range for the bias mean was 0.1–1.0 while for weights SD was 0.3–0.5. As for the number of neurons per layer, the range was 10–40. Several models were trained and the best one (i.e. accuracy close to 90%) was considered. The configuration of the selected

Table 1 The classification report for the recognition algorithm when tested with the children’s data

	Precision	Recall	F1-score	Support
Drop	0.98	1.00	0.99	120
Hit	0.72	0.83	0.77	210
Idle	0.98	1.00	0.99	390
Pickup	0.86	0.73	0.79	270
Shake	0.82	0.91	0.86	150
Throw	0.99	0.86	0.92	120
Avg/total	0.89	0.89	0.89	1260

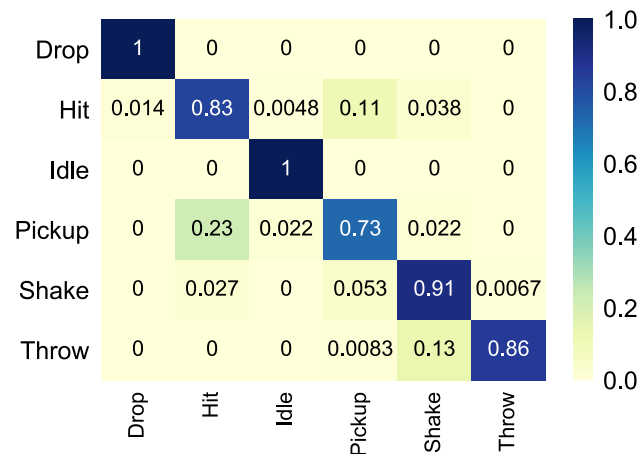


Fig. 5 The confusion matrix for the recognition algorithm when tested with the children data. The recognition performance of the model is higher than 90% for drop, idle, and shake. The recognition performance is less than 90% for hit, pickup, and throw

model included a bias mean of 0.3, weights SD of 0.3, and 28 hidden neurons per layer. The configuration of the architecture was 2×2 , where there are 2 hidden layers that contains 2 bidirectional layers each. More details about the architecture can be found in [61]. The model achieved promising results that considered precision, recall, and f1-score metrics (Table 1). The confusion matrix revealed that the model might confuse some of the behaviors (Fig. 5). For example, it might confuse *hit* as *pickup*. For the purpose of this study, we will focus on detecting *pickup*, *shake*, and *throw* or *drop*. Once these behaviors are detected, the robot will produce the corresponding responses. All other interactions will be ignored and will not produce any response once they are detected.

3.2 Experimental Setup

3.2.1 Robotic Toys

Three different toys embedded with recognition devices were considered. The toys were a stuffed panda (KRAMIG Soft toy, IKEA, Sweden), a stuffed toy robot (LATTJO soft toy,



Fig. 6 The companion toys used in the study. **a** The three different toys that were considered in our experiments. From left to right, a soft toy panda, a soft toy robot, and an excavator toy. **b** The data collection system used in this study. It consists of a Sense Hat board mounted on a Raspberry Pi board

IKEA, Sweden), and an excavator toy (Fig. 6a). The masses and dimensions of the selected toys were in the range that allowed the ease of carrying and manipulation for the targeted users. The same toys were previously used to collect the data that were then used to train the recognition model [3].

3.2.2 Recognition Device

The recognition device used was a small computing device (Raspberry Pi 3 Model B+, Raspberry Pi Foundation, UK). This device is powered by a 1.4GHz quad-core processor and supports wireless, Bluetooth, and Ethernet communication. The availability of such communication channels make it easier to access, program, and configure with other devices. Furthermore, it contains many peripherals that make it possible to augment it with other devices. The official operating system (Raspbian v4.19, Debian Project) was installed on a micro SD card (16 GB, Edge, Sanddisk). The selected storage should provide more than enough space for the operating system, trained recognition mode, collected data, and for any needed packages. A remote access software (TeamViewer Host for Raspberry Pi, US) was installed to allow ease of access to the device and more flexibility for debugging and testing. The kernel, firmware, and packages were all upgraded to their latest versions.

The standard Raspberry Pi does not contain any on-board board sensors, however, the 40-pin can support different boards with different functionalities. A Sense Hat board (Raspberry Pi Foundation, UK), which contains different sensors and a display, was mounted on the Raspberry Pi. The built-in accelerometer (LSM9DS1, STMicroelectronics, Switzerland) was used in the recognition model to acquire the raw acceleration data at a rate of around 30 Hz and at a magnitude of up to 16 g. This rate and magnitude were shown to be adequate enough for the recognition of human activities [17,33]. The entire device was placed in a dedicated

enclosure with a small fan mounted on the side for cooling (Fig. 6b). For the experiments, the devices were embedded inside the toys and each was powered by a dedicated power bank (Slim 2, 5000 mAh, POWERADD).

3.2.3 Reactions

We believe a companion robot should exhibit the feeling of pain once the robot is thrown or dropped. Hence, the responses for these behaviors were selected to be similar once an event is detected. The detection of being picked up or carried would produce a response to imply an event related to being surprised. As for being shaken, the robot would produce a response corresponding to being annoyed by the shaking action. The detection of the idle case produces no response as it means that there is no physical interaction that has occurred. For simplicity and to avoid redundancy with the throw and drop cases, the detection of hit does not produce any response. The reason for that is the logical response after being hit is to express pain, which is already covered by the other two cases. Hence, the reactions triggering actions were limited to *pickup*, *shake*, and *drop* or *throw*.

The robotic toys showed reactions when manipulated by the user. For example, a robot would display discomfort when shaken. The reactions were implemented as different short sounds. The samples were obtained from <https://freesound.org> and were modified for the experiments. The sound samples were cut and shortened to less than one second and were saved as *wav* files. For the behaviors considered in the experiments, 6 different sound samples for each behavior were selected to provide a variety. For example, when a *pickup* is detected, one sound sample is randomly selected from the pool of the available samples for *pickup* and then played (See supplementary material¹). A Bluetooth speaker

¹ See <https://youtu.be/uY1dpTIREIE>.

(AQL Sparkle, Cellularline, Italy) was used to emit the sound samples for the behaviors. The speaker was activated by the system embedded in the robot.

To investigate the effects of response time on the interactions, three different timings were considered. The three robotic toys were configured with reactions at different timings, namely, 0.5 s, 1 s, and 1.5 s. The timing of each toy was altered once after performing half of the experiments with each toy. For example, the timing of the panda toy was changed from 0.5 to 1.5 s. A scheduled task that periodically checks the detected behaviors was used to control the tested reaction times. This task generates a reaction based on the detected manipulation with a delay equal to the selected time. However, a condition has been implemented that prevents the generation of two consecutive responses in less than one second. This was designed to make the toy more natural in terms of response rate and more pleasant to interact with.

3.3 Experimental Procedure

3.3.1 Participants

The experiments conducted in this study were focused on the evaluation of the appropriateness of the reactions implemented in the robots, in particular on the reaction timing. Subjects (9 females and 21 males) volunteering in the experiments were students aged 8–13 years old (10.26 ± 1.48 years old). The consent from the parents was secured by their school and the children were accompanied by their teachers to the experiment site. The children were introduced in the experimental room one at a time. In the room, one researcher and one assistant were present. The procedures for these experiment did not include any invasive or potentially hazardous methods and were in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

3.3.2 Reaction Evaluation

Robotic toys or social companion robots should provide a timely feedback, a reaction, to the user performing an interactive act. A late and less frequent response might render the interaction slow and uninteresting while a very fast and more frequent response might be felt as eerie and unnatural. The frequency and the speed of a response should be natural and comfortable to the user. To evaluate the effects of these, a set of experiments were performed with a group of children individually. The three robotic toys were configured with reactions at different timing, namely, 0.5 s, 1 s, and 1.5 s. The participants were divided into three groups accordingly. A robotic toy was placed on a small table and a child was encouraged to interact with it. The evaluated behaviors were limited to pickup, shaking, and throwing

or dropping (Fig. 7). All tasks were requested in the form of imaginative scenarios that the children need to perform with the robotic toys (Table 2). After each session, a questionnaire containing five simple questions was given to the child (Table 3). The questions were related to the interactions and the possible answers were in Likert scale showing five different levels of agreement (i.e. from total agreement to total disagreement). All sessions were recorded with a webcam (C310 HD, Logitech, Switzerland) and then annotated with an open-source software (BORIS, version 3.12, Torino, Italy).

3.3.3 Data Analysis

The data collected from the participants were based on questionnaires containing five different questions. To visualize the collected responses, histogram plots were generated for each question to check for the peaks, spread, and symmetry. A Mann-Whitney U test was performed to check the effect of gender at $p < 0.05$. Furthermore, Kruskal–Wallis tests were performed on each question to check for any statistically significant differences between the medians of the three groups at $p < 0.05$. Furthermore, the test was performed to check for any effect due to gender differences.

4 Results

In this section, a summary of all the responses for each question are presented as histogram plots for the different groups. Then, the statistical analysis is provided for the effect of gender and the response time.

4.1 Summary of the Questionnaire

The first statement in the questionnaire was: *The robot reacted to my interaction*. The frequency of answers for each group were presented as a histogram plot in Fig. 8. The majority (i.e. 80%) of the responses for each group fell into the agreement region. This clustering of the responses created a right-skewed symmetry for all the groups. The peak of the data was at the *Strongly agree* response for group 3 (i.e. reaction time of 1.5 s). There was only one subject's response in the disagreement region for group 3. This could be due to the slow reaction time compared to other groups (i.e. 1.5 s vs 1.0 s or 0.5 s) that gave the wrong impression of the robot's responses to the subject. Alternatively, this could have been simply an outlier.

The distribution of the responses changed when the subjects were asked about the second statement of the questionnaire, which was: *The robot reacted quickly to my interaction*. Similar to Q1, the majority of the participants have answered

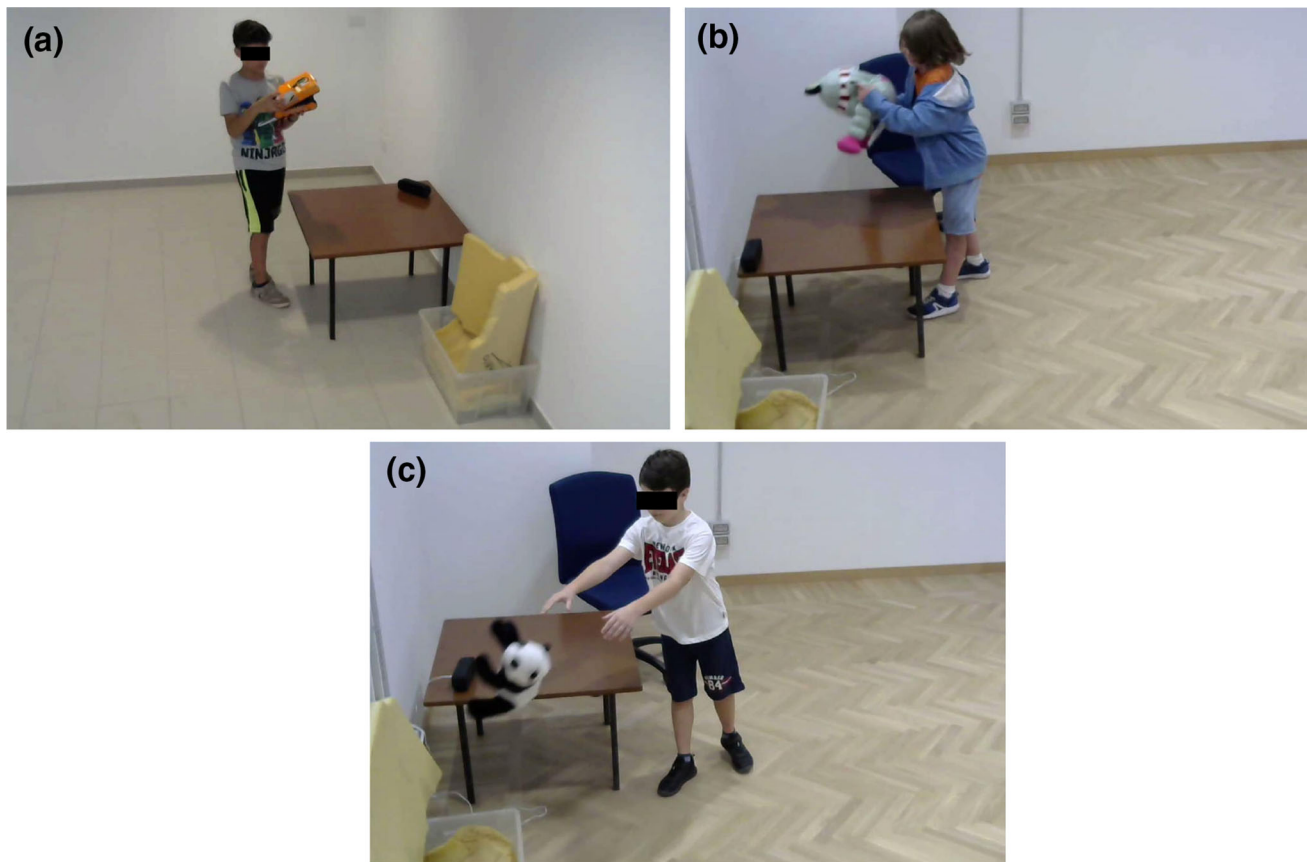


Fig. 7 Samples of the conducted experiments. **a** A child exploring the toy. **b** A child shaking the toy. **c** A child throwing the toy

Table 2 The experimental protocol for the experiments conducted in this study

Number	Reaction evaluation
Task 1	Pick up the robot up and explore it
Task 2	The robot is sleeping and in order to wake it up you need to shake it
Task 3	The robot would like to go to a specific place, toss it there

Table 3 The questions stated in the questionnaire

Number	Questionnaire statement
Q1	The robot reacted to my interaction
Q2	The robot reacted quickly to my interaction
Q3	The robot liked it when I picked it up
Q4	The robot liked it when I shook it
Q5	The robot liked it when I threw it

in agreement to the statement, with group 2 being the highest (i.e. 80% of the subjects) and group 3 the lowest (i.e. 60% of the subjects) (Fig. 9). The data for each group appear to be skewed to the right. There were three peaks for each group at

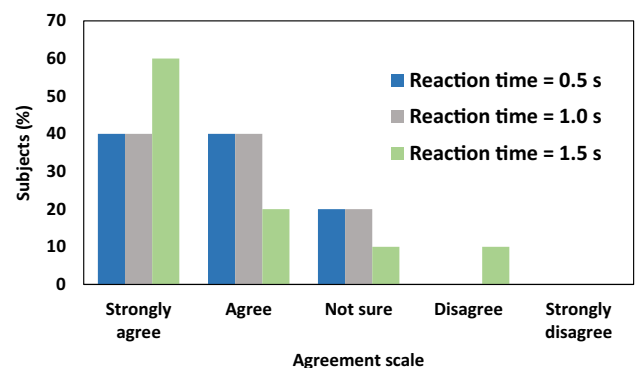


Fig. 8 A histogram summarizing the responses for the first question of the questionnaire: *The robot reacted to my interaction*

the *Strongly agree* and *Agree* scales. More responses were in the disagreement region as compared to the previous question. Group 3 contained the highest number (i.e. 40% of the subjects) of responses in the disagreement scales. This could be attributed to the relatively late response of the robot for this group as compared to the other groups.

The distributions for the third question (i.e. *The robot liked it when I picked it up*) showed different spread for each group (Fig. 10). The responses for group 2 (i.e. reaction time of

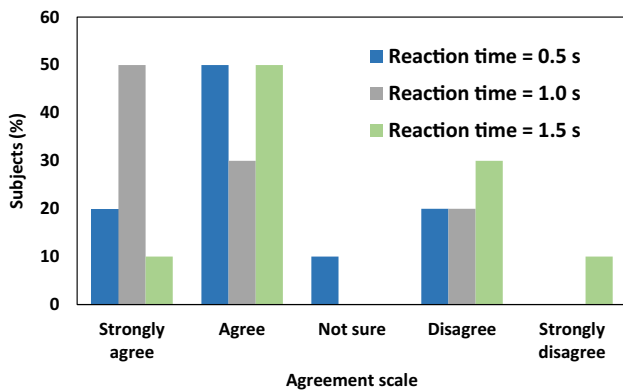


Fig. 9 A histogram summarizing the responses for the second question of the questionnaire: *The robot reacted quickly to my interaction*

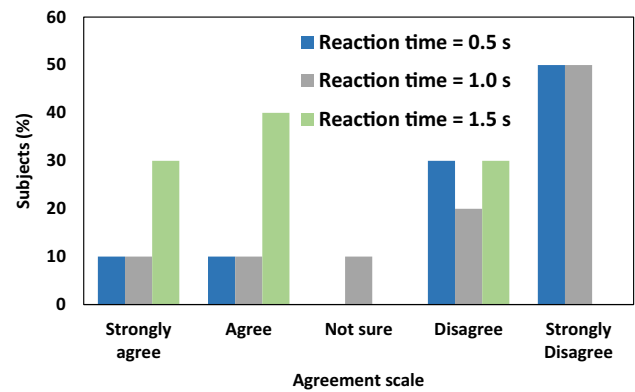


Fig. 11 A histogram summarizing the responses for the fourth question of the questionnaire: *The robot liked it when I shook it*

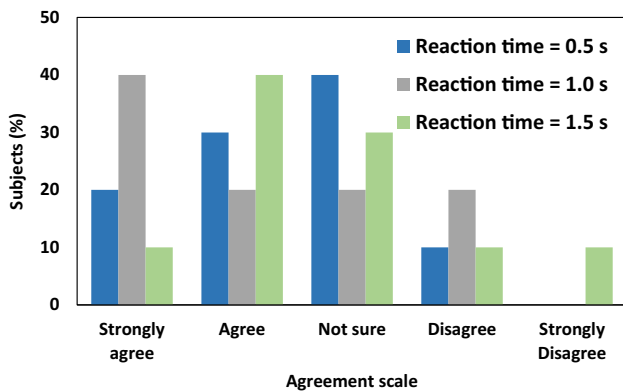


Fig. 10 A histogram summarizing the responses for the third question of the questionnaire: *The robot liked it when I picked it up*

1.0 s) appears to be right-skewed with 60% of the responses in the agreement region. Group 3 (i.e. reaction time of 1.5 s) also appears to be right-skewed, but with 50% of the subjects in agreement with the statement. The peak for group 2 was at *Strongly agree* selection while for group 3 the peak was at the *Agree* selection. As for group 1 with a reaction time of 0.5 s, the overall responses appear to be scattered in the agreement region (i.e. 50% of the subjects), however, the peak is at the *Not sure* scale. There were some responses in the disagreement region mainly for reaction time of 1.0 s and 1.5 s (i.e. 20%). The discrepancy in the responses could be attributed to the perceived understanding of the robot’s reactions due to the subjects’ interaction. The robot voice reaction to being picked up was similar to that of being surprised, but in a joyful manner. This could have confused some of the participants which made more responses leaning toward the *Not sure* scale or even into the disagreement region.

The fourth question was *The robot liked it when I shook it*. For this case, the robot produced a voice that indicated being annoyed to being shook. Hence, the responses are expected to be mostly in the disagreement zone. More than 70% of the responses for group 1 and group 2 fell into the disagree-

ment region (Fig. 11). Group 1 and group 2 (i.e. reaction time of 0.5 s and 1.0 s) appear to be left-skewed with two peaks occurring at the *Strongly disagree* scale. The majority of the participants in group 3 (i.e. reaction time of 1.5 s) voted in agreement (i.e. 70% of the subjects) to the fact that the robots liked being shaken. These results could be due to the relatively late response time for this group that made the robot produce delayed or incorrect reactions for the current interaction being made. For example, the robot is making the reaction for pickup while it should produce the one for shake. Clearly, a reaction time greater than one second could alter the perceived perception of a robot’s response.

The fifth question was related to the perceived understanding of the robots’ response after being thrown. The robot produced a sound indicating the feeling of pain in this case. The majority of the responses appear to be clustered in the disagreement region when the participants were asked *The robot liked it when I threw it*. The peak was found for group 1 (i.e. reaction time of 0.5 s) at the *Strongly disagree* scale followed by group 2 (i.e. reaction time of 1.0 s) at the *Disagree* scale (Fig. 12). Group 3 with a reaction time of 1.5 s achieved the highest number of responses (i.e. 40% of the subjects) in the agreement region followed by group 1 (i.e. 30% of the subjects).

4.2 Statistical Analysis

4.2.1 Gender Effect

As a secondary objective, it is interesting to find if there is an effect of gender on the responses for the different groups. For this analysis, only group 1 and group 2 were considered because of the close number of participants’ genders (i.e. total of 8 females vs. 12 males). A Mann–Whitney U test was run on 20 participants to determine if there were differences in the responses between males and females. The median response scores for males (3.5) and females (4.0) were not

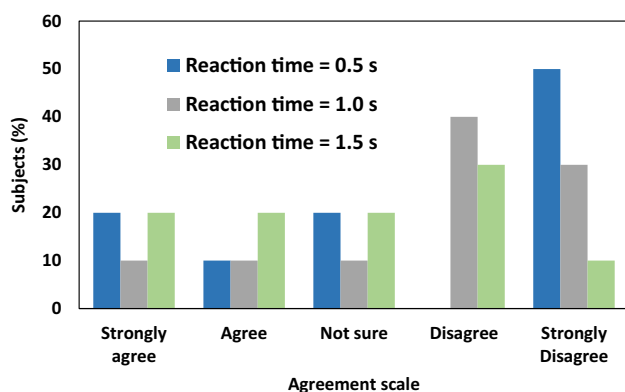


Fig. 12 A histogram summarizing the responses for the fifth question of the questionnaire: *The robot liked it when I threw it*

statistically significantly different, $p = 0.948$. These results were expected as the human perception of a response should be similar regardless of the gender.

4.2.2 Response Time Effect

A Kruskal–Wallis test for each item in the questionnaire was conducted to check for any significant difference between the three groups.

For the first question, the median values for group 1 (4.0), group 2 (4.0), and group 3 (5.0) were not statistically significantly different, $p = 0.827$.

The median values for the second questions of group 1 (4.0), group 2 (4.5), and group 3 (4.0) were not statistically significantly different, $p = 0.223$.

As for the third question, the differences between the median values of group 1 (3.5), group 2 (4.0), and group 3 (3.5) were not statistically significant, $p = 0.666$.

The median values for the fourth question of group 1 (1.5), group 2 (1.5), and group 3 (4.0) had statistically significant differences, $p = 0.023$. The average rank and median values showed that group 3 was different compared to the other groups. Group 3 was the one with the longest reaction time (i.e. 1.5 s) and that could explain the statistical difference.

As for the fifth question, the differences in the median values of group 1 (2.0), group 2 (2.0), and group 3 (3.0) were not statistically significant, $p = 0.415$. However, the average rank for group 3 (18.5) is higher than that of group 1 (14.3) and group 2 (13.8).

5 Discussion

The participants displayed different reactions while performing the tasks with the robotic toys. The first task was to pick up the robot and explore it, and the robot would respond with sounds implying a joyful reaction. For this task, many showed

curiosity and laughter about the sounds that the robots were emitting. Some of the children showed surprised expressions and stopped temporarily to explore the robots then looked at the experimenters. The second task was to shake the robot, and the robot would respond with sounds implying annoyance. For this task, many were surprised, stopped shaking the robot, and then placed it back after hearing the robots' reactions. A few resumed shaking after stopping temporarily. The last task was to throw the robot to a specific target, and the robot would emit a sound, which implied pain. Many showed surprised expressions about the responses while some of them gazed at the experimenter with astonished looks.

The results of the questionnaire implied that there is an effect for the reaction timings on the perceived understanding of the robots' responses. Group 3 (i.e. reaction time of 1.5 s) scored more incorrect responses across most of the questions as compared to other groups. This was very evident in the responses for the fourth item in the questionnaire (Fig. 11). The delay in producing a reaction to an interaction might have given the wrong impression about the causation effect, hence, making it difficult to understand the aim or goal behind a robot's response. In other words, the longer the duration to make a reaction, the more likely it will deliver an incorrect message to the user for the intended interaction. Producing a response within one second from detecting a stimuli should produce more favorable results. The Kruskal–Wallis test results for the fourth question support these findings.

Another dimension that might have influenced the responses is the modality of the response itself. The sounds for the responses were considered to indicate three different expressions, namely, joyful surprise, being annoyed, and feeling pain. These responses were selected by adults to target children as the primary users. Some of the incorrect responses to the questions could be attributed to a possible confusion about the intended message behind each sound (i.e. response). This implies the need for more commonly-accepted responses that could be easily understood regardless of age, culture, or geographical region.

The number of participants in our study was limited to 30 subjects. Hence, experiments with a larger sample size are required to make a better generalization. Furthermore, the experiments in our study were conducted with neurotypical children. The findings cannot be necessarily generalized to include those with special needs or cognitive disorders. More tailored and individualized experiments need to be conducted to study and address the needs of those populations. The experiments in this study were limited to three different responses corresponding to three different interactions. However, more responses could exist to imply different emotions and reactions. Sound was the only modality that was considered to convey the robots' responses. Different modalities could be considered and integrated to provide clearer responses. Children were the only participants in our

experiments because of the targeted end-users of this study. However, adults participants could be considered to obtain a more comprehensive and in-depth feedback about the experiments. Finally, the recognition model could be improved to increase its capabilities in recognizing more behaviors accurately and quickly.

6 Conclusions

We have presented an approach to detect and respond to three types of manipulation of robotic toys, namely, being picked up, being shaken, and being thrown. Furthermore, we have evaluated the perception of a response provided at different reaction timings through the emission of sounds. The results showed that the reaction time affects the understanding of a robot's response to an interaction. Furthermore, sound as a modality to be used in a robot's response provided a sufficient message to be understood by the majority of the participants.

Ideally, the response to an action for robotic toys should occur not more than one second after the detection of an aggressive behavior or an unwanted interaction. This implies the need for fast recognition algorithms that must provide a quick prediction about an interaction. The modality of the response should be clear enough to provide the right message intended from the interaction. Multiple modalities could be fused together to provide a stronger response and clearer message to the user. Hence, it would reduce the likelihood of user's misinterpreting the intended message behind a response.

Companion robots would benefit from having the capability of detecting and reacting to aggressive interactions. This layer to detect unwanted interactions would operate independently from the robot's main objectives. Having such capabilities to detect undesired behaviors could be used to make children experiment with the consequences of their actions and their effects on others. For example, a robot displaying sad emotion after being hit can influence a child to believe that this behavior is not appropriate in social interactions. Furthermore, this also has the potential to be extended to target aggression among neurotypical and neurodivergent children.

The perception of children with special needs and cognitive disorders toward an emotional response might differ compared to neurotypical children and might even differ among the same disorder group. For example, children with autism are different in their symptoms depending on the degree and diagnosis of ASD [59]. This diversity among these populations opens the possibility for more personalized models of various timings and settings of robotic designs to meet their requirements [50].

Future studies can investigate the emotional appropriateness of sounds along with other modalities. Furthermore,

a potential future work would consider monitoring some aspects of the participants' reactions to determine more quantitative analysis. For example, aspects, such as gaze, emotions, and others, can be considered. Moreover, further improvements on the recognition algorithm should be considered to ensure smoother interactions, which should reach a much higher performance to become acceptable as a product in the mass market.

Acknowledgements Open Access funding provided by the Qatar National Library. The work is supported by a research grant from Qatar University under the grant No. QUST-1-CENG-2019-10 and by a scholarship grant from the Italian government (IRE 2018-2019). The statements made herein are solely the responsibility of the authors.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Alhaddad AY, Cabibihan JJ (2016) Reflex system for intelligent robotics. In: Qatar foundation annual research conference proceedings, vol 2016, no 1, p HBSP2914. Hamad bin Khalifa University Press (HBKU Press), Doha (2016)
2. Alhaddad AY, Cabibihan JJ, Bonarini A (2019) Head impact severity measures for small social robots thrown during meltdown in autism. *Int J Soc Robot* 11:255–270
3. Alhaddad AY, Cabibihan JJ, Bonarini A (2019) Recognition of aggressive interactions of children toward robotic toys. In: 2019 28th IEEE international conference on robot and human interactive communication (RO-MAN), pp 1–8
4. Alhaddad AY, Cabibihan JJ, Hayek A, Bonarini A (2019) Influence of the shape and mass of a small robot when thrown to a dummy human head. *SN Appl Sci* 1(11):1468
5. Alhaddad AY, Cabibihan JJ, Hayek A, Bonarini A (2019) Safety experiments for small robots investigating the potential of soft materials in mitigating the harm to the head due to impacts. *SN Appl Sci* 1(5):476
6. Alhaddad AY, Javed H, Connor O, Banire B, Al Thani D, Cabibihan JJ (2019) Robotic trains as an educational and therapeutic tool for autism spectrum disorder intervention. In: Lepuschitz W, Merdan M, Koppensteiner G, Balogh R, Obdržálek D (eds) *Robotics in education*. Springer, Cham, pp 249–262

7. Alink LR, Mesman J, Van Zeijl J, Stolk MN, Juffer F, Koot HM, Bakermans-Kranenburg MJ, Van IJzendoorn MH (2006) The early childhood aggression curve: development of physical aggression in 10-to 50-month-old children. *Child Dev* 77(4):954–966
8. American Psychological Association (2018) *Apa dictionary of psychology*. <https://dictionary.apa.org/aggression>. Last Accessed 26 Aug 2018
9. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL, (2013) A public domain dataset for human activity recognition using smartphones. In: (2013) proceedings. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)
10. Ann OC, Theng LB (2014) Human activity recognition: a review. In: 2014 IEEE international conference on control system, computing and engineering (ICCSCE). IEEE, pp 389–393
11. Bagala F, Becker C, Cappello A, Chiari L, Aminian K, Hausdorff JM, Zijlstra W, Klenk J (2012) Evaluation of accelerometer-based fall detection algorithms on real-world falls. *PloS ONE* 7(5):e37062
12. Baghdadli A, Pascal C, Grisi S, Aussilloux C (2003) Risk factors for self-injurious behaviours among 222 young children with autistic disorders. *J Intell Disabil Res* 47(8):622–627
13. Bao L, Intille SS (2004) Activity recognition from user-annotated acceleration data. In: International conference on pervasive computing. Springer, Berlin, pp 1–17
14. Boccanfuso L, Barney E, Foster C, Ahn YA, Chawarska K, Scassellati B, Shic F (2016) Emotional robot to examine differences in play patterns and affective response of children with and without ASD. In: The eleventh ACM/IEEE international conference on human robot interaction. IEEE Press, pp 19–26
15. Bodfish JW, Symons FJ, Parker DE, Lewis MH (2000) Varieties of repetitive behavior in autism: comparisons to mental retardation. *J Autism Dev Disord* 30(3):237–243
16. Bonarini A, Garzotto F, Gelsomini M, Romero M, Clasadonte F, Yilmaz ANÇ (2016) A huggable, mobile robot for developmental disorder interventions in a multi-modal interaction space. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE, pp 823–830
17. Bouten CV, Koekkoek KT, Verduin M, Kodde R, Janssen JD (1997) A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Trans Biomed Eng* 44(3):136–147
18. Cabibihan JJ, Javed H, Sadasivuni K, Alhaddad AY (2018) Smart robotic therapeutic learning toy. WIPO Patent WO2018033857. World Intellectual Property Organization
19. Cabibihan JJ, Chellali R, So CWC, Aldosari M, Connor O, Alhaddad AY, Javed H (2018) Social robots and wearable sensors for mitigating meltdowns in autism: a pilot test. In: Ge SS, Cabibihan JJ, Salichs MA, Broadbent E, He H, Wagner AR, Castro-González Á (eds) *Social robotics*. Springer, Cham, pp 103–114
20. Cabibihan JJ, Javed H, Aldosari M, Frazier TW, Elbashir H (2017) Sensing technologies for autism spectrum disorder screening and intervention. *Sensors* 17(1):46
21. Cabibihan JJ, Javed H, Ang M Jr, Aljunied SM (2013) Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *Int J Soc Robot* 5(4):593–618
22. Cabibihan JJ, Javed H, Sadasivuni KK, Alhaddad AY (2019) Smart robotic therapeutic learning toy. US Patent App. 16/326,169
23. Connor DF (2012) *Aggression and antisocial behavior in children and adolescents: research and treatment*. Guilford Press, New York
24. del Rosario M, Redmond S, Lovell N (2015) Tracking the evolution of smartphone sensing for monitoring human movement. *Sensors* 15(8):18901–18933
25. Ennis-Cole DL (2015) *Technology for learners with autism spectrum disorders*. Springer, Berlin
26. Estes A, Munson J, Dawson G, Koehler E, Zhou XH, Abbott R (2009) Parenting stress and psychological functioning among mothers of preschool children with autism and developmental delay. *Autism* 13(4):375–387
27. Fodstad JC, Rojahn J, Matson JL (2012) The emergence of challenging behaviors in at-risk toddlers with and without autism spectrum disorder: a cross-sectional study. *J Dev Phys Disabil* 24(3):217–234
28. Graves A, Jaitly N, Mohamed AR (2013) Hybrid speech recognition with deep bidirectional LSTM. In: 2013 IEEE workshop on automatic speech recognition and understanding. IEEE, pp 273–278
29. Gurney JG, McPheeters ML, Davis MM (2006) Parental report of health conditions and health care use among children with and without autism: National survey of children's health. *Arch Pediatr Adolesc Med* 160(8):825–830
30. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
31. Johnson NL, Lashley J, Stonek AV, Bonjour A (2012) Children with developmental disabilities at a pediatric hospital: staff education to prevent and manage challenging behaviors. *J Pediatr Nurs* 27(6):742–749
32. Kanne SM, Mazurek MO (2011) Aggression in children and adolescents with ASD: prevalence and risk factors. *J Autism Dev Disord* 41(7):926–937
33. Kern N, Schiele B, Schmidt A (2003) Multi-sensor activity context detection for wearable computing. European symposium on ambient intelligence. Springer, Berlin, pp 220–232
34. Kozłowski AM, Matson JL (2012) An examination of challenging behaviors in autistic disorder versus pervasive developmental disorder not otherwise specified: significant differences and gender effects. *Res Autism Spectr Disord* 6(1):319–325
35. Li B, Boccanfuso L, Wang Q, Barney E, Ahn YA, Foster C, Chawarska K, Scassellati B, Shic F (2016) Human robot activity classification based on accelerometer and gyroscope. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN). Presented at the 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 423–424
36. Lukowicz P, Ward JA, Junker H, Stäger M, Tröster G, Atrash A, Starner T (2004) Recognizing workshop activity using body worn microphones and accelerometers. In: International conference on pervasive computing. Springer, Berlin, pp 18–32
37. Machalicek W, O'Reilly MF, Beretvas N, Sigafos J, Lancioni GE (2007) A review of interventions to reduce challenging behavior in school settings for students with autism spectrum disorders. *Res Autism Spectr Disord* 1(3):229–246
38. Matson JL, Gonzalez ML, Rivet TT (2008) Reliability of the autism spectrum disorder-behavior problems for children (ASD-BPC). *Res Autism Spectr Disord* 2(4):696–706
39. Matson JL, Shoemaker M (2009) Intellectual disability and its relationship to autism spectrum disorders. *Res Dev Disabil* 30(6):1107–1114
40. Murphy O, Healy O, Leader G (2009) Risk factors for challenging behaviors among 157 children with autism spectrum disorder in Ireland. *Res Autism Spectr Disord* 3(2):474–482
41. Network N.E.C.C.R., Arsenio WF, et al (2004) Trajectories of physical aggression from toddlerhood to middle childhood: predictors, correlates, and outcomes. *Monographs of the Society for Research in Child Development*, pp i–143
42. Nock MK, Kazdin AE, Hiripi E, Kessler RC (2007) Lifetime prevalence, correlates, and persistence of oppositional defiant disorder: results from the national comorbidity survey replication. *J Child Psychol Psychiatry* 48(7):703–713
43. Oliveira E, Orrù D, Nascimento T, Bonarini A (2017) Modeling player activity in a physical interactive robot game scenario. In:

- Proceedings of the 5th international conference on human agent interaction. ACM, pp 411–414
44. Oliveira EL, Orrù D, Morreale L, Nascimento TP, Bonarini A (2018) Learning and mining player motion profiles in physically interactive robogames. *Fut Internet* 10(3):22
 45. Oliveira EL, Orrù D, Nascimento T, Bonarini A (2017) Activity recognition in a physical interactive robogame. In: 2017 Joint IEEE international conference on development and learning and epigenetic robotics (ICDL-EpiRob). IEEE, pp 92–97
 46. Plant KM, Sanders MR (2007) Reducing problem behavior during care-giving in families of preschool-aged children with developmental disabilities. *Res Dev Disabil* 28(4):362–385
 47. Plötz T, Hammerla NY, Rozga A, Reavis A, Call N, Abowd GD (2012) Automatic assessment of problem behavior in individuals with developmental disabilities. In: Proceedings of the 2012 ACM conference on ubiquitous computing. ACM, pp 391–400
 48. Robinson H, MacDonald B, Broadbent E (2014) The role of health-care robots for older people at home: a review. *Int J Soc Robot* 6(4):575–591
 49. Rogers SJ (1996) Brief report: early intervention in autism. *J Autism Dev Disord* 26(2):243–246
 50. 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* 3:19
 51. Salter T, Michaud F, Létourneau D, Lee D, Werry IP (2007) Using proprioceptive sensors for categorizing human–robot interactions. In: 2007 2nd ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 105–112
 52. Scassellati B, Admoni H, Mataric M (2012) Robots for use in autism research. *Annu Rev Biomed Eng* 14:275–294
 53. Shibata T (2012) Therapeutic seal robot as biofeedback medical device: qualitative and quantitative evaluations of robot therapy in dementia care. *Proc IEEE* 100(8):2527–2538
 54. So WC, Wong MY, Cabibihan JJ, Lam CY, Chan RY, Qian HH (2016) Using robot animation to promote gestural skills in children with autism spectrum disorders. *J Comput Assist Learn* 32(6):632–646
 55. Sucerquia A, López JD, Vargas-Bonilla JF (2018) Real-life/real-time elderly fall detection with a triaxial accelerometer. *Sensors* 18(4):1101
 56. Sukhodolsky DG, Smith SD, McCauley SA, Ibrahim K, Piasecka JB (2016) Behavioral interventions for anger, irritability, and aggression in children and adolescents. *J Child Adolesc Psychopharmacol* 26(1):58–64
 57. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In: Advances in neural information processing systems, pp 3104–3112
 58. Teo HT, Cabibihan JJ (2015) Toward soft, robust robots for children with autism spectrum disorder. In: FinE-R@ IROS, pp 15–19
 59. Tidmarsh L, Volkmar FR (2003) Diagnosis and epidemiology of autism spectrum disorders. *Can J Psychiatry* 48(8):517–525
 60. Xingjian S, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC (2015) Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: Advances in neural information processing systems, pp 802–810
 61. Zhao Y, Yang R, Chevalier G, Xu X, Zhang Z (2018) Deep residual bidir-lstm for human activity recognition using wearable sensors. *Math Probl Eng*. <https://doi.org/10.1155/2018/7316954>

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

Ahmad Yaser Alhaddad obtained his Bachelor degree in robotics and automation from Multimedia University (MMU), Malaysia in 2011; and obtained his Master degree in Mechatronics and Automatic Control from Universiti Teknologi Malaysia (UTM), Malaysia in 2013. He has worked on many Engineering Projects involving circuit design, microcontroller and control systems. He is currently doing his Ph.D. in Mechanical and Information Technology at Qatar University and at Politecnico di Milano, as part of the double doctorate programme. His research interests include social robotics, haptics, control engineering, and mechatronic systems.

John-John Cabibihan received the Ph.D. degree in Bioengineering, with specialization in Biorobotics, from the Scuola Superiore Sant'Anna, Pisa, Italy, in 2007. From 2008 to 2013, he was an Assistant Professor at the Electrical and Computer Engineering Department, National University of Singapore. He is currently an Associate Professor at the Mechanical and Industrial Engineering Department, Qatar University. He is also the Lead/Co-Lead Principal Investigator of several projects under the National Priorities Research Program of Qatar National Research Fund. He was the General Chair of the 10th International Conference on Social Robotics (ICSR) 2018 in Qingdao, China; Program Chair/Co-Chair of the IEEE Conference of Robot and Human Interactive Communications (Ro-Man) 2019 in New Delhi, India; ICSR 2019 in Madrid, Spain; ICSR 2017 in Tsukuba, Japan; ICSR 2016 in Kansas City, USA; ICSR 2012 in Chengdu, China; and ICSR 2010 in Singapore. Over the years, his work has been focused toward assistive and social robotics for the therapy of children with autism, lifelike prosthetics, bioinspired tactile sensing, and human-robotic touch and gestures.

Andrea Bonarini is full Professor and Coordinator of the AI and Robotics Lab at Politecnico di Milano. His current research interests are on machine learning and emotional human–robot interaction, with special focus on robotic games and toys for entertainment and social, psychological and physical development of people with and without special needs. Since 1989, he has developed with his collaborators and students more than 50 autonomous robots. In 2015, he co-founded Nova Labs, to share experience and tools to produce professional robots in a short time at low cost. He is collaborating with toy companies to introduce technology in their products. He has an active role in the COST European network “LUDI: play for children with disabilities”. He published more than 150 peer-reviewed papers.