International Journal of Intelligent Robotics and Applications (2025) 9:592–607
https://doi.org/10.1007/s41315-024-00401-7

REGULAR PAPER

Using intelligent robots to detect body language and improve social development in children with autism spectrum disorder

Wei-sheng Lin[1] · Xiang-yu Peng[1] · Yu-fang Cheng[1]

Received: 5 August 2024 / Accepted: 29 October 2024 / Published online: 24 November 2024

Abstract

Socially assistive robotics (SAR) is an expanding field within special education, proven to benefit the social development of children with autism spectrum disorder (ASD). This population often faces challenges in social interaction, particularly in understanding body language, leading to feelings of isolation and impacting their social lives. The use of robot-agent for practice is increasingly being adopted to address these challenges. In this study, an intelligent robot system was designed to enhance social interactions, particularly focusing on body language, among children with ASD. The system integrated an intelligent robot with a Kinect sensor to detect and execute body language actions, utilizing these developed modules for detecting body movements. This study explored the effects of the developed NAO-Recognitions of Body Languages and Imitation (NAO-RBI) system and analyzed the performance of three participants with ASD in these targeted body languages. The body-language gestures involved: raising a hand, high-five, handshake, "Please give me," "Come on!," arms akimbo with an angry face, and "Please!". A multiple-probe design was used to monitor improvements in the targeted actions, with a scoring module developed to gather scores from each participant. Results showed that the NAO-RBI system successfully recognized and demonstrated the seven body language gestures. This highlights the potential of SAR, combined with Kinect technology, to enhance body language skills in children with ASD. Comparing skeletal point data captured by Kinect with predefined standard body skeleton coordinates effectively executed the module functions. Robot-assisted practice proved to be a valuable tool for improving body language understanding and use for social interactions in this population. Further research is needed to optimize the robot and enhance its effectiveness.

Keywords Socially-assistive robots · Multiple probe designs · Body-languages · Autism spectrum disorder · A kinect tool

1 Introduction

In the past decade, robotics technology has developed at an amazing pace, and all types of robots, from industrial robots to social robots, have been used in a wide variety of applications (Sheridan 1992). Specifically, Socially-Assistive Robot (SAR) has been implemented in children with serve multiple functions, including emotional therapy, cognitive training, social facilitation, companionship, and physiotherapy (Sheridan 1992). These are often applied to meet the social and psychological needs of children with autistic spectrum disorder (ASD) (Lee 2020).

✉ Yu-fang Cheng
yfcheng@cc.ncue.edu.tw

[1] Department of Industrial Education and Technology, National Changhua University of Education, No.1 Jin-De Road, Changhua 500, Taiwan, ROC

Children with ASD have difficulties with social interactions, communication, and expression; repetitive behaviors; strong obsessions; and uniformity in the preschool years (Muhle et al. 2004a). The impairments of social communication and interaction include the following (Cadette et al. 2016): (a) social-emotional reciprocity deficits: inability to have normal conversations with others and less sharing of emotions or expressions; (b) social non-verbal communication behavioral deficits: difficulty coordinating verbal and non-verbal communication (e.g., eye contacts and body-languages); and (c) impairment in relationship and developmental skills: lack of interest in peers and difficulties in making friends. These communication skills are essential for developing social relationships in this population. One of these non-verbal communication skills is body language, a major social cognitive skill; it plays critical cognitive, linguistic, communicative, and social roles (Cummings 2021; Doody and Bull 2011). Humans often use body language such as gestures to support verbal expression, pointing at objects, or speaking. Psychologists believe that humans communicate 60% more through non-verbal cues than through other methods (Duh 2006). Most children with ASD have difficulty understanding the meaning of the body language of the general population, communication and interaction with others might hinder normal body-language recognition, isolating these children mentally and behaviorally from general social expectations (Muhle et al. 2004a). Therefore, the study(Alshurman and Alsreaa 2015) has proved that children with ASD could make improvements in nonverbal communication skills if they received early training in their lives. For instance, they could increase their capability of expressing their emotions and feelings to others, such as pointing to a happy picture or pointing to a specific object. This population sometimes could use hand gestures or body movements to perform their needs, such as pointing to a water cup to present thirst; or pointing to a toilet stool to convey their need to go to the bathroom. Because of lacking verbal communication skills, some educators developed various pictures to communicate with others, e.g., showing a picture of an apple to indicate they want to eat it (Febriantini et al. 2021).

Some studies have attempted to help non-verbal communication interventions for children with ASD (Lee 2020; Chevalier et al. 2017; Kumazaki et al. 2019), they have impairment of execution functions in motor, which stems from integrating sensory information delayed, and difficulty adjusting mistakes from their internal experiences synchronization (Vishne et al. 2021) Research shows that their motor functions are not noisy, but they perform poorly in quickly integrating sensory information for action planning, especially in tasks that require rapid responses, such as catching a ball, which requires quick integration of visual information; such as catching a ball, which requires the rapid integration of visual information (Vishne et al. 2021) In synchronous tasks, individuals with autism are less likely to use recent sensory information for error correction, which affects their ability to synchronize movements(Vishne et al. 2021) Additionally, they have a weaker ability to integrate multiple sensory inputs and feedforward models, resulting in poorer performance in more challenging motor tasks. These factors significantly affect the performance of individuals with autism in tasks that require rapid responses (Vishne et al. 2021).

Regarding the provision of prosthetic tools for practice within this population, the study indicated that using Socially Assistive Robotics (SAR) in treatment and education is beneficial for children with ASD. SAR can increase their motivation and engagement, helping to improve target behaviors and comprehend essential concepts in teaching or practicing social skills (Kaboski et al. 2015). For instance, adaptable teaching methods in a customizable learning environment can meet children's individual needs, provide repetition and consistency for building confidence and skill mastery, and offer real-time feedback in a stress-free learning environment (Geminiani et al. 2019). The studies of Chevalier et al. (2017); Costa et al. 2018) stated that robot—child interactions in the form of non-verbal communication tasks, robots could teach children body language better than humans through non-verbal communication (Robins et al. 2005). Many researchers believe that humanoid robots with human-like behaviors can help with the treatment and education of children with ASD (Bleuer et al. 2017). Some researchers (Muhle et al. 2004b; Shi et al. 2022) have also stated that anthropomorphic robots are more appropriate for children with ASD because of their humanoid appearance (e.g., limb mobility is more effective in stimulating social responses). Such as the NAO robot. It is a 58-cm-tall humanoid robot with a small rounded body, arms, and legs, is similar in appearance to humans, can walk, recognize humans or objects, listen to, and talk to people, and use body language and speech to express emotions (Tariq et al. 2016). The study (Chevalier et al. 2017) utilized the NAO robot to conduct an imitation task, where the robot interacted with 12 children with ASD in seven imitation sessions over eight weeks. The results demonstrated that these participants exhibited increased engagement in the robot-based environment and showed improvements in target behaviors, such as the ability to locate body parts, make eye contact, follow others' gaze, and engage in joint attention. These findings suggest that robots could effectively stimulate and maintain children's interest without causing distraction (Tariq et al. 2016).

In the context of using a robot with a Kinect sensor to track the body language of children with ASD, a capture system that tracks body movements is necessary. However, it's important to note that children with ASD are particularly sensitive to being touched, and the use of wearable devices may lead to the rejection of such hardware (Geminiani et al. 2019; Robaczewski et al. 2021). One effective tool is visual capture systems, such as low-cost depth cameras and the Kinect v2 sensor developed by Microsoft. The Kinect sensor integrates depth sensing, RGB cameras, and audio sensing for comprehensive information capture, providing effective real-time human tracking capabilities. When coupled with three-dimensional (3D) pose estimation algorithms, Kinect can estimate the 3D joint positions of the human body. Additionally, researchers like Firman et al. (2016) used Kinect's depth camera in computer vision to estimate spatial occupancy and surface shapes (Firman et al. 2016). Stoll et al. (2020) also applied Kinect in automatic sign language production, combining neural machine translation and generative adversarial networks to generate realistic sign language videos (Stoll et al. 2020). These applications demonstrate Kinect's versatility and precision across various disciplines.

Very few studies adopted the Kinect sensors connected with the NAO robot for improving the body language in children with ASD. The NAO robot, combined with the Kinect sensor, offers multiple advantages. Firstly, Kinect can capture and transmit human skeletal data in real time. After processing, the NAO robot can imitate these movements instantly. This enables the NAO robot to interact with users quickly and naturally, with high accuracy and response speed in simulating human movements. (Yavsan and Uçar 2016). Regarding physical treatment for body language, the NAO robot has particularly shown advantages for practicing executive function in this population. Research showed that using Kinect and the NAO robot could effectively teach body language to children with ASD. The NAO robot could not only mimic user movements but also provided real-time feedback, which played a crucial role in enhancing the engagement, communication, and social skills of children with autism. For example, in the IOGIOCO ((Interactive mirroring Games with Social Robot)) therapy, the NAO robot can accurately recognize and mimic 19 body language gestures of children with ASD, achieving an accuracy rate of 95% (Chartomatsidis et al. 2016). These studies demonstrated that the NAO robot combined with the Kinect sensor could provide a non-invasive interactive platform that helps children with autism improve their ability to imitate and infer body language. Therefore, this study developed an identifiable and imitation of the body language system, which integrates the Kinect sensor and the NAO social robot to enhance the imitation and inference abilities of children with ASD in various social contexts and to promote their social interaction (Ivani et al. 2022).

The study (Yavsan and Uçar 2016) proposed a system (involving a computer, Xbox 360 Kinect, and NAO robot) that could instantly recognize and mimic human upper body movements. Their study pointed out that with the different sizes of the human body and robots, the Cartesian coordinates obtained from Kinect could not be converted directly to the NAO coordinate space; so that the data of skeletal points cannot be directly provided to the NAO robot. Therefore, the analyzed data in the skeletal points of humans collected from the Kinect were then converted to the joints of the NAO robot into the control commands in the computer and sent these data to the NAO robot via wireless fidelity (Wi-Fi) or a conventional network. The results showed that the NAO robot could capture human movements and imitate these behaviors through the Kinect sensor, requiring only a small amount of communication effort and time. And the user's interaction with the NAO robot was natural and fast, similar to that between other humans. This could be used with children or the elderly without the need for external intervention by a physical therapist. Another example of Chartomatsidis et al. (2016), this study used Microsoft Kinect sensors to control the movements of an NAO humanoid robot to mimic human movements in real-time. It built an application that used the Kinect sensor to capture the skeletal points of human movement, process the skeletal point information, and send it back to the robot in real-time so that the robot could successfully mimic the captured motion in real-time. The user could record his movements and store in a skeleton object file, and then the NAO performed these actions from the stored files by executing certain commands. It processes involved four steps: Kinect tracking the user's body and saving its skeletal joints; calculating the angle formed between the following three joints of the human body (shoulder, elbow, and wrist) and the offset of the angle repeatedly; and finally sending this data-values to the NAO robot. This allowed the user to train the NAO robot by performing these behaviors with multiple executions. On the other hand, in real-time motion, the user needs to perform the same action accurately multiple times in ordering the NAO robot following and learning; and require the data of angle and offset calculations correctly. The results indicated that the NAO had a high success rate in executing movements. However, controlling its palm proved challenging because the Kinect sometimes struggled to accurately detect whether the user's hand was open or closed during certain movements.

More importantly, a few studies focused on this special group. The study developed an interactive training system for children with ASD, which utilized a robot, Kinect sensors, and depth cameras to mimic actions and provide synchronous verbal commands (Liu et al. 2016). The child could imitate the robot's actions, and conversely, the robot could mimic human actions while guiding the child in adjusting movements to synchronize with the robot through verbal commands. These findings indicate the potential of social robots as effective tools for interventions in people with ASD. Similarly, an interactive mirror robot game called CopyRobot is specifically designed for children with ASD (Santos et al. 2020). Common characteristics of children with ASD include difficulties in social interaction, repetitive behaviors, and motor skill impairments. Their study utilized an NAO robot, Kinect cameras, and a computer to enhance imitation abilities and motor skills in this population. The experiment in clinical trials was divided into two protocols: one where the robot led the interaction, and the other where the therapist led to facilitate interaction between children and therapists. The results indicated the significant therapeutic value of these imitation activities in children with ASD. This study underscores the importance of motor interaction and stimulation in ASD therapy and demonstrates the potential of robots as therapeutic tools.

This study, therefore, developed a programmed system involving the Kinect sensor and the NAO robot to help children with ASD improve their body language and imitate the performance of the participant. The NAO-recognition of body language and imitation (NAO-RBI) system was developed in this study. The system was designed to perform and recognize seven body-language gestures ("Raising a hand," "High-five," "Handshake," "Please give me," "Come on!", "Arms akimbo with an angry face," and "Please"). Thus, the study addressed following the research question:

- Whether the Kinect sensor and the NAO robot be implemented autonomously between different modules of the NAO-RBI system?
- Whether the developed system could be successfully exhibited for children with ASD?
- How was the identification of body-language gestures ("Raising a hand," "High-five," "Handshake," "Please give me," "Come on!", "Arms akimbo with an angry face," and "Please") implemented in the NAO-RBI system?

2 System development

The NAO intelligent robot was used, a humanoid robot developed by the French company Aldebaran Robotics in 2006 (Qidwai et al. 2020). It has a human-like appearance and can execute a task properly. The hardware of the NAO includes a battery with a life of 45 min, speakers, LEDs in the eyes and ears, infrared emitters, sonar, tactile sensors, force-sensing resistors, two cameras, gyroscope, accelerometer, Wi-Fi, and Ethernet. The NAO robot uses the NaoQi operating system, which can be programmed using the Choregraphe software by using C + +, Python, Java, MATLAB, Urbi, C#, and.Net programming languages. The NaoQi also includes the basic tools: joint control, walking, talking, and face tracking. The body-language gestures designed in the NAO robot included "Raising a hand," "High-five," "Handshake," "Please give me," "Come on!", "Arms akimbo with an angry face," and "Please!" The NAO robot could be speaking by asking questions, providing guidance and prompts, and displaying encouraging behavior.

The NAO-RBI system involved the NAO robot and the Kinect sensor via Wi-Fi. connection, which contained different software modules for detecting body movements. The user could stand in front of the Kinect sensor and perform targeted actions according to the provided still pictures of social scenarios on the computer screen. The skeletal data was collected by Kinect and then delivered these data to the NAO-RBI system to save the skeletal points of the participants' movements. The Python programming language was implemented in the NAO robot module, and the Kinect was programmed in C# language. It was equipped with different sensors and actuators, which allowed it to display full body movement, face and object recognition, and automatic voice recognition. The Kinect captured the body-language position and direction and exported its 3D Cartesian skeletal point coordinates into a Microsoft Excel file. The NAO-RBI system could read and interpret the skeletal point coordinates, and determine whether the body language was correct, then send the data back to the NAO robot.

Since the NAO robot could not imitate human movements directly from the data obtained from the Kinect, the skeleton data collected from the Kinect was not directly delivered to the NAO robot. Because the skeletal point coordinates of humans and robots differed in size, the data of skeletal points captured by the Kinect was first compared with the pre-defined standard body skeletal coordinates. Then, the error of each body language (LOSS) value was calculated, and the smallest LOSS value was selected as the result of the judgment gesture to determine whether the body language was correct.

The NAO-RBI system assessed the accuracy of the participant's body language movements and transmitted the data back to the NAO robot. The participant performed actions corresponding to the presented social situation. If the participant did not respond within 5 min, the NAO robot could provide tips. Upon correct execution of the body language, the NAO robot would offer praise or feedback; if incorrect, the NAO robot could encourage the participant to try again. Therefore, the NAO-RBI system divided its data into five "*.py files*," each of which was responsible for a different function (as shown in Fig. 1) as follows:

- *main.py:*
  - This was the entry point of the program, mainly responsible for creating the graphical user interface (GUI), creating the body-language labels (seven actions), determining whether to go to the next action (based on whether the action was correct or not) and calling the files "InformationWindow.py," "record_video.py," "Body_Predict.py," "Predict.py," and others.
- *InformationWindow.py:*
  - Users entered basic information such as "Name," "Age," "Gender," "WISC Scale Score," and "Obstacle Type and Level," which was recorded in the MySQL database.
- *record_video.py:*
  - This launched the Kinect sensor to capture 25 skeletal coordinates (X, Y, Z) of the participant's body and exported the skeletal coordinate data into a Microsoft Excel file to facilitate the subsequent judgment of the correct body-language movements.
- *Body_Predict.py:*
  - After "record_video.py" detected the skeletal points, it exported the Microsoft Excel file and read the data, compared the data with the pre-set standard body skeletal coordinates, calculated the LOSS value for each posture, selected the smallest LOSS value as the

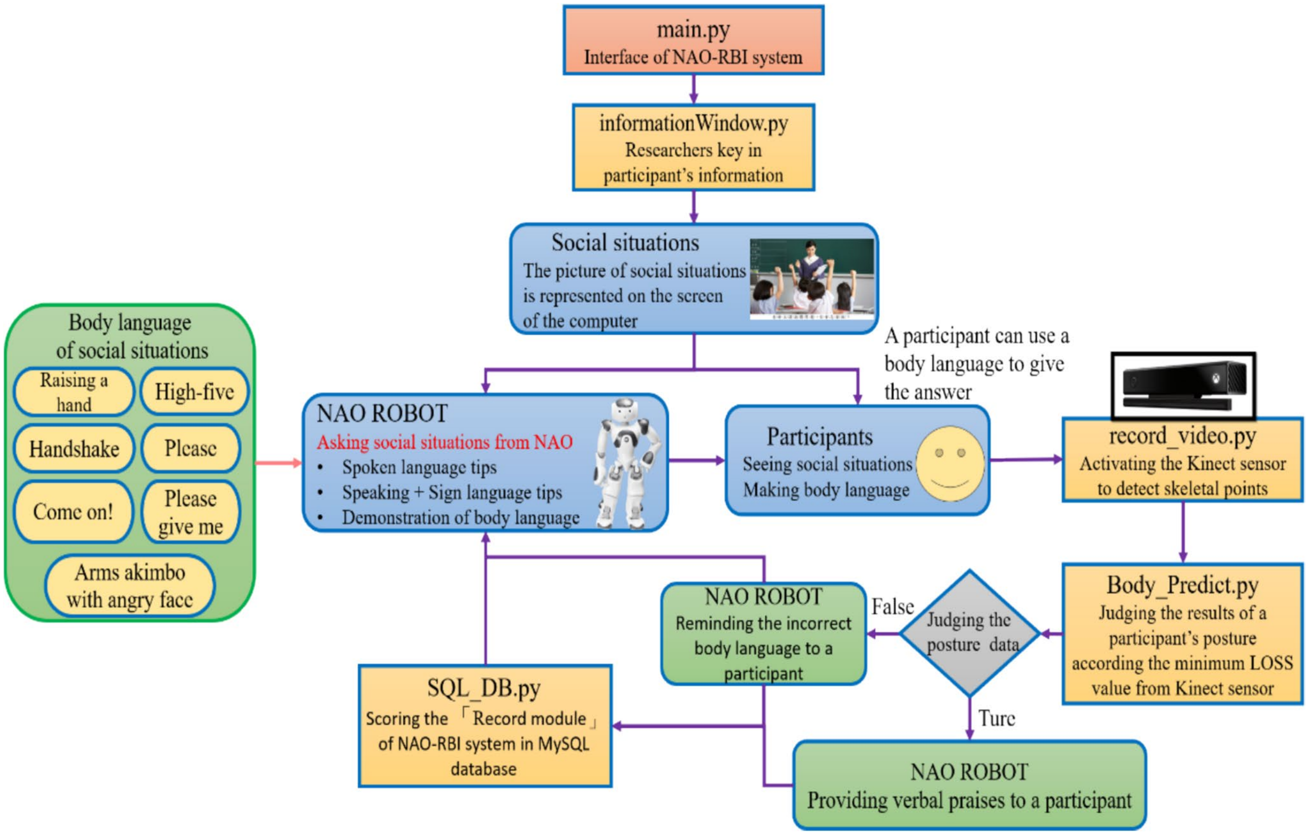

Fig. 1 Flow chart of "NAO-RBI system"

posture result to judge whether the body posture was correct, and returned the result to the NAO robot.

– *SQL_DB.py:*
  - For the "NAO-RBI system—scoring module," the database was created with MySQL, and the data sheet, NAO-RBI system, and MySQL connection were constructed with the Python language. The seven targeted behaviors of "Raising a hand," "High-five," "Handshake," "Please give me," "Come on!", "Arms akimbo with angry face," and "Please" were recorded in the MySQL database, along with the basic information entered in InformationWindow.py.

First, the system ran *main.py* to open the researcher interface and call InformationWindow.py; the basic information of the participant such as "name," "age," "gender," "WISC Scale Score," and "Obstacle Type and Level" was entered, and the window was closed. Next, it could click "Start imitation" on the researcher interface and call record_video.py to start the Kinect sensor to capture 25 skeletal points (each with X, Y, and Z coordinates, so a total of 75 skeletal coordinates) of the participant's body, and export the skeletal coordinates data into a Microsoft Excel file. Then, calling the "*Body_Predict.py*" combined the data with the pre-set standard body skeletal coordinates. If the recognition was correct, the NAO robot gave verbal praise and clapped for the participant. If the recognition was incorrect, the NAO robot verbally reminded the participant of the incorrect body posture and encouraged the participant to perform it again. This is shown in Fig. 2.

Seven targeted body languages were collected during the system operation, as shown in Fig. 3. The number of responses recognized in the system and the basic information entered in "InformationWindow.py" were recorded in the MySQL database.

3 Research methods

This study employed a "cross-subject multi-baseline design," consisting of three phases: baseline, intervention, and maintenance (Reynhout and Carter 2011). The method was to assess the progress of changes in the participant with ASD's performance. in which all participants were subjected to the intervention procedure, and the results were obtained by visual analysis. Since this population was different from typically developing children, they had heterogeneity in characteristics and personalize, this method allows to the measurement of precise and personalized assessments and

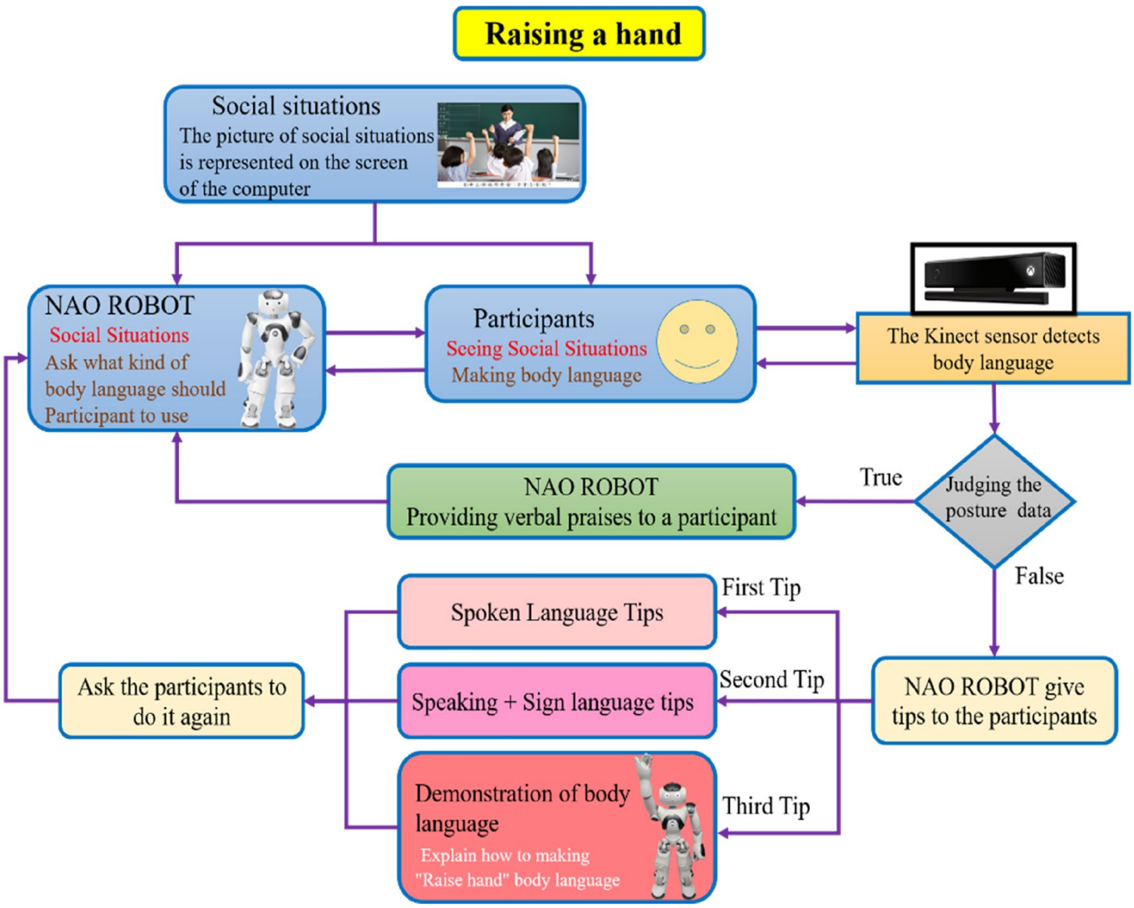

Fig. 2 Part of a flowchart showing a scenario with possible—Raising a Hand

is adapted to the needs of individual progress (Muhle et al. 2004a). Further, the baseline phase was designed to observe the participants' behavioral changes before the experimental intervention. The intervention phase was designed to observe whether the participant's body language changed through the NAO-RBI system during the experiment. The maintenance phase was concerned with observing whether the participant could maintain the performance during the intervention phase or return to the baseline phase after the NAO-RBI system was removed (Fig. 4).

The participant was set up in a university laboratory, where they could be involved in a quiet environment without any disturbances. Three sessions were involved in this experimental study. The NAO-RBI system was only used in the intervention phase for helping the body language, and the baseline and maintenance phases were designed to investigate the changes in these targeted behaviors. In the intervention, the NAO-RBI system was controlled by a technician using the "Wizard of Oz" approach. The NAO robot displayed body language to the participant, and the Kinect sensor connected the NAO and controlled it remotely via Wi-Fi. The body language and voice were presented by the NAO robot according to different teaching contexts. The Kinect sensor captured the participants' body bone points and sent these data to the NAO-RBI system to determine whether the body language was correct or not. At the same time, the resulting data was sent back to the NAO robot for giving feedback or applause. Finally, all the performances and the scores of each participant were recorded by using the single-subject method-visual observation and scoring module embodied in the NAO-RBI system.

4 Participants

This experiment was conducted on three participants with ASD. Their parents were informed of the purpose of the study and provided a score on the Wechsler Intelligence

Fig. 3 NAO Robot—Seven Body Language Behavior Expressions

#	Target Behavior	NAO Behavior Description	#	Target Behavior	NAO Behavior Description
1	Raising a hand		**5**	Come on!	
2	High-five		**6**	Arms akimbo with angry face	
3	Handshake		**7**	Please	
4	Please Give Me				

Scale (WISC) above 70 for each participant. The participants with ASD were then enrolled in the study, and their participating consent was obtained before the experiment.

Before the experiment, the researcher assessed the performance of body language for each participant. Participant A could execute some gestures but struggled with using both hands to request something from others, for example, saying, "Please give me this!" Participant B encountered difficulty standing with legs straight and moving the body properly, which hindered their ability to perform certain body-language gestures accurately. Participant C demonstrated a limited understanding of body language. She struggled to comprehend most of the given instructions and required additional time to act accordingly. The relevant data for each participant is presented in Table 1.

5 Measurement tools

5.1 The pictures of social scenarios (PSS) card and body-language observation scale (BLO scale)

The PSS cards were used as teaching material for displaying social events, they displayed pictures of a social scenario evoked by body language (including “Raising a hand,” “High-five,” “Handshake,” “Please give me”, “Come on!”, “Arms akimbo with an angry face,” and “Please”). The 24 social scenarios were used with the targeted body language in these three sessions (baseline, intervention, and maintenance).

Fig. 4 Participants interacted with the NAO robot

Table 1 Participant's information

	Eden	Carl	Heidi
Diagnosis	ASDs	ASDs	ASDs
Gender	Male	Male	Female
Age	6	10	6
WISC	101	82	73

This BLO scale was developed regarding the "social skills training course examples" (Hong et al. 1999) and the early social communication scales (ESCS) (Mundy et al. 2003) and was self-administered and expert-checked to record and analyze the changes in the body language of the participants during the baseline, intervention, and maintenance phases. Scores ranged from 0 to 5, based on the participant's first response to the targeted actions. The score 5: participants complete a posture independently without any assistance or prompting. The score 4: the participant needs to be reminded with partial oral prompts, and then able to achieve the targeted actions. Score 3: the participant needs to be provided the verbal prompt and the instructions for the gestures. Score 2: the participant needs to watch the demonstration of targeted gestures; then he/she can imitate this action. The score 1: the participant needs to have physical assistance from the researcher to complete the targeted actions. The score was 0: the participant did not have any response.

Finally, the participant's performance was assessed in each session using the PSS card and the BLO scale, which were completed with the researchers at the end of the experiment.

5.2 NAO-RBI system—scoring module

The NAO-RBI system was used to interact with the participants during the intervention phases of the experiment. The recorded data in the scoring module includes skeletal points of the body-language actions exhibited by the participants during the experiment, as well as the number of correct responses recorded in the database This scoring module recorded the performed body language of each participant; seven body-language actions was tested up to three times individually. If a participant's body-language action appropriately corresponded with the specific social scenario on the PSS card, the system would advance to the next body-language action question. The NAO-RB system detected the body language for a score of 5 and discriminated the posture according to the last body language. The scoring system had four scoring levels (displayed in Table 2). The score 0: The

Table 2 The recorded data of the scoring module

Target behaviors Usage count	Raise your hand	High five	Hand shake	Please give me	Come on	Insert waist Angry	Please (Request)
Eden							
Usage count (I)	0	1	4	2	0	0	0
Usage count (II)	0	1	4	3	1	0	0
Usage count (III)	1	0	0	4	4	0	0
Usage count (IV)	0	0	4	1	3	0	0
Usage count (V)	0	0	4	4	0	0	0
Usage count (VI)	4	1	4	4	1	0	0
Carl							
Usage count (I)	0	0	4	3	1	0	0
Usage count (II)	0	1	4	4	1	1	0
Usage count (III)	0	0	1	4	1	0	0
Usage count (IV)	1	0	0	4	0	0	0
Usage count (V)	0	1	4	4	2	0	0
Usage count (VI)	0	0	4	4	4	0	0
Heidi							
Usage count (I)	2	2					
Usage count (II)	2	0	3	3	1	0	0
Usage count (III)	1	1	4	4	0	0	0
Usage count (IV)	0	1	4	4	1	0	0
Usage count (V)	0	0	4	4	1	0	0
Usage count (VI)	0	0	4	4	4	0	0

participant could complete the body language by himself/herself. The score 1: Failed once to display the correct body-language gesture but could perform it once the NAO robot provided verbal tips. The score 2: Failed twice to display the body-language gesture but could do it after the NAO robot provided verbal and gesture tips. The score 3: Failed three times to display the correct body-language gesture but could do it after a demonstration by the NAO robot. The score 4: Unable to perform the body-language action correctly and needed physical assistance from the researcher.

6 Experimental design and process

The study aimed to investigate the participants' ability to recall NAO robot teaching content during the intervention phase and whether learning effects in the maintenance reverted to baseline states. At the being, the participant was brought to a university laboratory for an undisturbed experimental environment. At the end of each phase (baseline, intervention, and maintenance periods), the researchers used the PSS cards and BLO scale to measure the performance and behaviors of each participant with ASD. The study was a non-continuous experiment, with weekly sessions of approximately 40–60 min each, for 12 weeks over three months.

During the baseline phase (weeks 1–3), the participant was allowed to randomly select three of the PSS cards. Subsequently, they were verbally presented with questions related to body language, such as, "*If we just met today and became new friends, what could we do for each other?*" and "*What could I do if I don't understand something in class?*" Participants were then instructed to respond using only body language.

In the intervention period (weeks 4–9), the NAO-RBI system was set up, and the system and related kits were installed on a computer. The computer screen showed the social event selected from the PSS cards in a still picture, and the participants could perform the corresponding body language according to the presented social event. Meanwhile, the Kinect sensor captured the participants' movements of the body. The NAO-RBI system and Kinect sensors facilitated learning body language. A technician controlled the system using the "Wizard of Oz" method, providing feedback on participants' gestures.

If a participant's body language posture was correct, the NAO robot would provide feedback through verbal praise and clapping hands. If a participant performed incorrectly, the NAO robot would offer verbal tips, demonstrate the correct body language, and encourage the participant to try again. Meanwhile, the researchers aided the participants if they could not understand or did not know how to perform the gestures. Subsequently, the NAO robot displayed the corresponding body language and asked the participants, "*Which body language could be performed in this social event?*" The participants were required to respond/perform to the NAO robot using their body language. Upon the participants' response, the NAO robot provided feedback by applauding or offering hints. It could provide verbal and non-verbal hints as needed. Simultaneously, the scoring module recorded the body language performed by the participants for further analysis.

For the maintenance phase (weeks 10–12), the NAO-RBI system used in the intervention phase was removed. The PSS cards were used to represent the social event and required the participant to correspond to suitable body language. The teaching process was likely the baseline phase. This session aimed to assess whether the participants with ASD could maintain the learning outcomes from the intervention and whether they exhibited any changes compared to the baseline phase.

All three participants underwent the intervention procedure, and the results of targeted behaviors were assessed through visual analysis and scoring. A video recorder was used for data analysis, observation, and evaluation of inter-observer consistency and procedural integrity.

7 Data collection and analysis

To measure the changes in the targeted body language of the participants, the score of each participant was collected by the researchers-observers. The performance of each participant and the execution of the operating process of the NAO-RBI system were recorded in the scoring module. The following two types of data were obtained in this study:

(1) The recorded data of the scoring module: The data of the performed body language gestures from the NAO-RBI system were obtained from the participant's body movement during the intervention phase. The data from the seven targeted body languages were recorded in the scoring module.
(2) The score of the participant's performance was collected and analyzed in three phases (baseline, intervention, and maintenance). The scores were assigned by the three observers using the BLO scale, and a visual analysis (Inter-observer agreement) was used to represent a graph. The IOA was used to calculate the inter-observer reliability by observing whether the participants were attentive or not. All average IOA scores were above 87%. The formula was as follows: Consistency percentage = (amount of concordances/number of concordances + number of inconsistencies) × 100%.

8 Results

8.1 Results of the participants' performance

This study used the cross-subject multi-baseline design of the single-subject method to conduct preliminary analysis in the baseline, intervention, and maintenance phases. It also used visual analysis to draw a line graph of the data points obtained using the BLO scale (Fig. 5).

The performance of the three participants, Eden, Carl, and Heidi, varied across the baseline, intervention, and maintenance phases, as illustrated in Fig. 5. Eden showed progress in the baseline phase, with scores rising from approximately 45 to nearly 70. In the intervention phase, Eden's scores stabilized between 70 and 80, indicating steady performance in the targeted behaviors (as highlighted in the red box in Fig. 5). During the maintenance phase, although there was a slight decrease in scores, they remained above 70, approaching 80, demonstrating that Eden was able to retain most of the learned skills. Similarly, Carl exhibited remarkable progress, with scores rapidly increasing from 30 to 70 during the baseline phase, followed by further improvement and stabilization around 80 during the intervention phase (as highlighted in the red box in Fig. 5). In the maintenance phase, Carl's scores slightly decreased to between 70 and 75, yet this still reflected successful retention of the acquired skills. In contrast, Heidi's progress was more gradual, with fluctuating scores during the baseline phase. However, her performance gradually improved during the intervention phase, rising from around 30 to nearly 40. By the maintenance phase, her scores further increased to between 60 and 75, indicating that although her learning pace was slower, she still exhibited a certain level of skill retention (as highlighted in the yellow box in Fig. 5). Overall, the results indicated that while all participants benefited from the intervention, the magnitude and pace of improvement varied, reflecting individual differences in response to the intervention.

8.2 The recorded data of the scoring module:

The data recorded in the NAO-RBI system-SQL_DB.py is listed in Table 2. If a participant failed to answer the question or took more than 2 times to respond to the question, the NAO robot would give "verbal tips," "verbal with gesture tips," or "demonstrated the body language"; it was depending on the number of wrong answers by

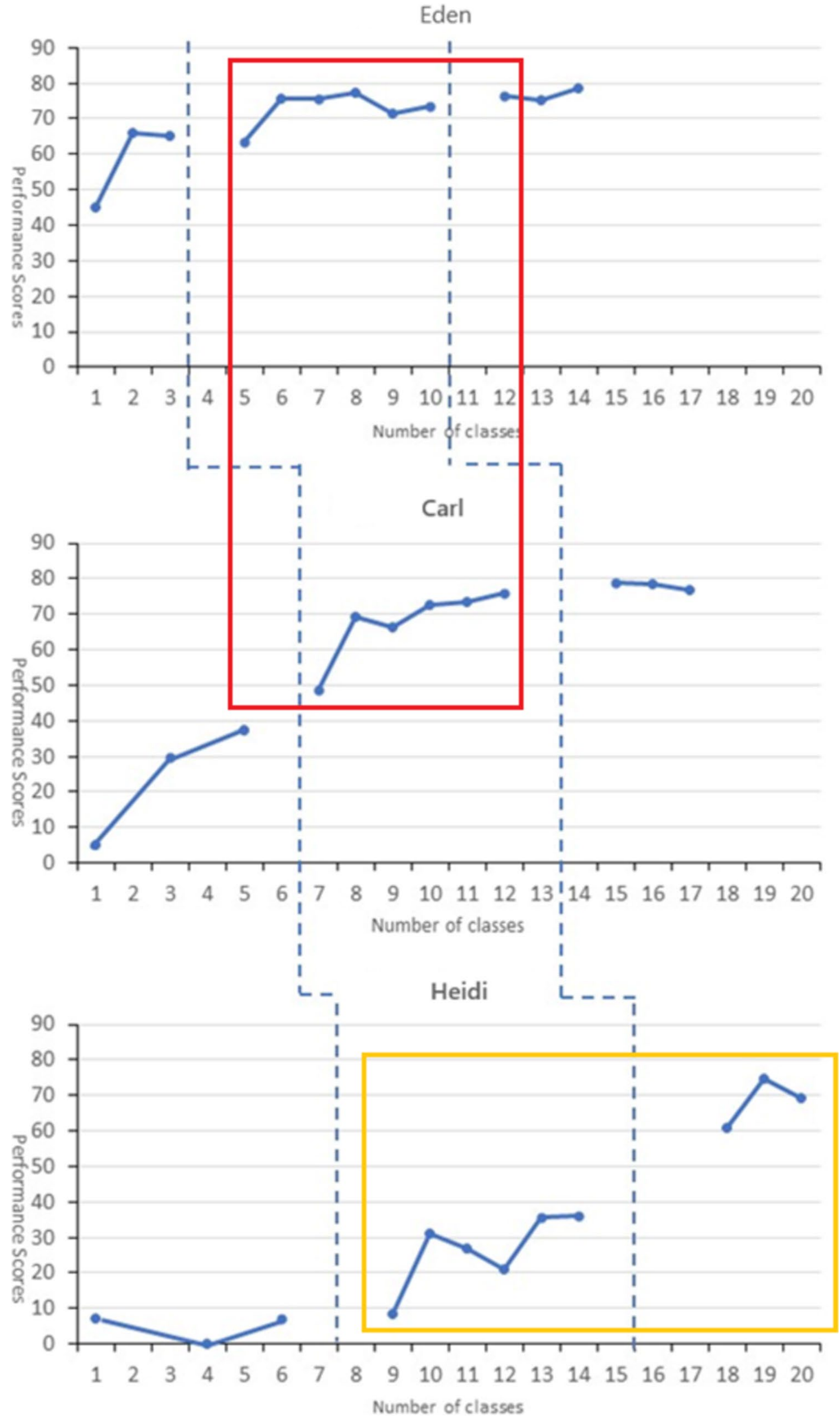

Fig. 5 The baseline, intervention, and maintenance phases of the participants' body language performance

the participant. Meanwhile, the researcher also observed whether the participants understood the body language shown by the NAO robot and gave suitable prompts. If the participants still could not perform the body language actions correctly after the demonstration by the NAO robot, then the researcher provided physical assistance to the participants. More specifically, almost all these participants successfully performed the targeted gestures. The participants' performance of the gestures "Handshake" and "Please give me" was suboptimal, and several factors likely contributed to the system’s difficulty in recognizing these specific gestures accurately. First, the clarity of the gestures played a critical role. Participants may not have fully extended their arms or performed clear hand movements, which made it difficult for the Kinect sensor to accurately detect the hand positions. For example, if a participant did not fully extend their arm or moved their hand during a handshake, the system might struggle to classify the gesture due to the unclear action. Second, the angles and distance between the participant's body and the sensor were key factors. The Kinect sensor needed the participant to maintain specific body positioning, arm angles, and movement ranges to capture gestures correctly. If participants stood too far from or too close to the sensor, or if their arm angles did not match the system’s predefined skeletal model, the system might misinterpret the gesture. This issue was especially critical for gestures like "handshake" and "please give me," which require more precise alignment and movement for the system to accurately detect them. Third, the complexity of the gestures themselves could also affect recognition accuracy. These gestures required a higher level of hand–eye coordination and precise movements. Children with ASD often experience difficulties with motor planning and execution, which can impair their ability to make quick adjustments to their movements. This, in turn, reduces the accuracy of the gestures, making them harder for the system to recognize. Fourth, the limitations of the sensor should also be taken into account. The Kinect sensor sometimes struggles to distinguish whether a hand is open or closed, which is essential for detecting fine motor movements, like those involved in a handshake. If the participant's hand movements are subtle or occur at the edge of the sensor's detection range, the system may have difficulty correctly identifying the gesture. These technical challenges, combined with the participants' motor difficulties, may further contribute to the system's inability to consistently recognize these specific gestures. Lastly, variations in the participants' understanding of the social meaning and execution of these gestures could further impact recognition accuracy. Some participants may have a limited grasp of the social significance behind gestures like "Handshake" and "Please give me," leading to improper execution even after receiving prompts from the robot. This misunderstanding can complicate the system’s ability to detect these gestures accurately, as the participants may perform the actions incorrectly or inconsistently. Table 2 highlights the NAO-RBI system's detection of body language performance across the three participants, with five body language gestures being correctly identified with significant accuracy. However, the recognition accuracy for "Handshake" and "Please give me" was notably lower. This suggests that these gestures posed greater challenges for the system, possibly due to a combination of sensor limitations, gesture complexity, motor difficulties, and varying levels of participant comprehension. As a result, the system struggled to consistently identify these specific actions.

9 Discussion

This study aimed to improve body language skills; and effectiveness in addressing these targeted behaviors in children with ASD. This study investigated the identification and imitation performance of children with ASD for body language after the NAO-RBI system intervention. It evaluated the effectiveness of the executive functions of the NAO-RBI system, which cooperated with the skeletal points captured by the Kinect sensor. The results revealed that the NAO-RBI system successfully recognized and demonstrated seven body language gestures. It effectively executed module functions to transmit the skeletal coordinates captured by the Kinect sensor to the NAO robot. The advantage of this system over other research (Lee 2020; Liu et al. 2016; Doody and Bull 2013) lies in its ability to effectively recognize body movements and provide feedback with greater speed and accuracy through the NAO-RBI system. This is achieved by not directly using the data acquired by Kinect for the NAO robot's imitation of human movements. Instead, the system involves comparing the skeletal point data captured by Kinect with predefined standard body skeleton coordinates. Additionally, the NAO-RBI system facilitated the practice of imitating body language for children with ASD, thereby eliciting their engagement and motivation. The results indicated that participants improved their targeted body language skills in this experimental study.

The NAO-RBI system offered accessibility and user-friendliness for this population. The system accurately recognized five body language gestures ("Raising a hand," "High-five," "Come on!", "Arms akimbo with an angry face," and "Please!") exhibited by the NAO robot. In this process, the NAO-RBI system utilized the body language skeletal points captured by the Kinect sensor to determine their accuracy and provide the final data back to the NAO robot. In contrast, two body language gestures (“Handshake” and “Please give me”) were hardly recognized in the

NAO-RBI system. Adjusting the angle setting of the hand position with the standard body skeleton coordinates predesigned in the "Body_Predict.py" module was necessary. For example, the participant's body language performance was compared using these skeleton coordinates, and the LOSS value for each posture was calculated. The smallest LOSS value was selected to determine whether body language was correct. Therefore, the skeleton coordinates for each gesture needed to be adjusted based on the following parameters: (a) Determining the degree of arm elevation required for clear detection of body language. (b) Identifying specific ranges of skeleton coordinates within each gesture requires more precise assignment, such as the performance of straightened arms. Furthermore, some of the gestures (e.g., “Handshake” and “Please give me”) were not recognized properly by the NAO-RBI system, because the gestures performed by the participants were probably not obvious. For instance, the participants’ arms were not straight or out of the detecting range, and the precise position between the participant and the Kinect sensor was not defined accurately. These issues will improve in the future.

Further, the Kinect sensor was used to capture 25 skeletal coordinates (X, Y, and Z) of each body language gesture of participants with ASD. The collected data were used to determine the posture of each participant through the NAO-RBI system, and the processed results could be sent back to the NAO robot to execute the posture. The NAO robot could exhibit appropriate body language immediately according to the social event caused by participants. The system successfully identified whether the performance of a gesture by a participant was correct. Compared to other similar research systems, the primary advantage of this system lies in its ability to accurately assess the correctness of behaviors in children with ASD using minimal sensors. In addition, the developed modules were adaptable and able to be customized for different functions in further. The robot can effectively demonstrate actions; the NAO-RBI system supports repetitive learning, enabling children to practice without stress. This system offers convenience, ease of use, and comprehensive functionality, contributing to increased learning confidence in children with ASD. It also provided instant feedback to enhance the learning experience for these children.

Moreover, concerning the transmission data of the Kinect sensor, many studies (Chartomatsidis et al. 2016; Doody and Bull 2013; Assad-Uz-Zaman et al. 2021) have processed Kinect sensors connected to the NAO robot. Their studies usually linked the human skeletal points captured by the Kinect sensor and mapped these data to the skeletal points of the NAO robot. Thus, the limb movements of the NAO were controlled by the Kinect sensor. The collected data of skeletal points of the human were analyzed, computed first, and then converted into the control commands corresponding with each joint of the NAO robot to display body language movements. The skeleton data could not be directly delivered to the NAO robot due to differences in the body shapes of humans and robots. However, this effective approach of the NAO-RBI system used the Kinect sensor was used in this study. The design of the NAO-RBI system used the Kinect sensor to collect data on skeletal points from the participants and then to determine whether the performed gesture was correct. The NAO robot was presented as an agent only to interact with these participants in this context. It was not necessary to require the collected data of skeletal points in the Kinect sensor to be transmitted to the NAO robot. This method reduced complex calculations, the time of data analysis, and the processing for the frequency of the “delay” phenomenon.

The intelligent robot NAO-RBI system has significant advantages in training individuals with Autism Spectrum Disorder (ASD) by recognizing and interpreting their body movements. This system leverages the inherent characteristics of robots, including the ability for repetitive use, resistance to fatigue, and a friendly demeanor, to provide users with extensive practice opportunities. Firstly, the use of intelligent robots allows for the repetitive execution of tasks without the risk of fatigue, ensuring consistent training sessions. This consistency is particularly beneficial for individuals with ASD, who often require numerous repetitions to master specific skills. The robot's ability to provide endless practice opportunities without variation in performance helps to reinforce learning and skill acquisition. Secondly, integrating interdisciplinary course teaching helps to understand body movements. By combining the teaching of social context with extended training sessions, an effective approach can be developed to address both physical movement and cognitive learning. Consequently, this approach also can support children with developmental disorders in understanding and executing correct body movements. Furthermore, the interactive nature of training with physical robots facilitates the generalization of learned skills to daily life. As participants engage with the robot, they might transfer the learned skills during training to real-world scenarios through repetitive practice. This generalization is essential for individuals with ASD, who often find it challenging to apply learned behaviors outside of controlled environments. A similar study (Alshurman and Alsreaa 2015) also indicated that targeted behaviors in this population could be improved through repetitive practice. Another significant advantage was the provision of immediate feedback during training sessions. The robot offered verbal or visual cues to guide participants toward the correct execution of body movements and promptly corrected any errors. This immediate feedback enhanced the learning process, improving participants' confidence and motivation. In summary, the use of intelligent robots, like the NAO-RBI system in training programs for individuals with ASD offers numerous benefits. These include the ability for repetitive practice, integration of cognitive learning, facilitation of skill generalization, and provision of immediate feedback. These advantages collectively contribute to improved outcomes for individuals with ASD, considering the potential of SAR in therapeutic and educational settings. Further research needs to refine these functions of the robot and explore their long-term benefits.

Regarding the improvement of this population, most participants could perform the body gestures independently without any hints from the NAO robot. Here are some key points of using the developed NAO-Recognitions of Body Languages and Imitation (NAO-RBI) system. These participants could have independent performance and skill improvement, most participants performed the body gestures without assistance, and participants practiced with the NAO robot and repeated exercises to master body language. Especially in the "Please give me" gesture was observed significant improvement during the maintenance phase. Further, the developed system provided assisted learning and immediate feedback, some participants needed verbal or gesture hints from the NAO robot; and immediate verbal, voice, and visual feedback from the robot boosted motivation and confidence. Participants improved their body language skills by practicing with the NAO robot and repeating exercises, particularly for those who initially struggled with the gestures. For example, participant A, who initially struggled with maintaining eye contact and receiving objects with one hand, showed marked improvement. After the intervention, he maintained eye contact and expressed gratitude when receiving a cookie from the researcher during the maintenance phase, saying, "Thank you, teacher!". However, these findings highlight the effectiveness of the NAO-RBI system in enhancing body language skills among children with ASD, emphasizing the importance of continuous practice and feedback. Further research is needed to refine these interventions and explore their long-term benefits (Geminiani et al. 2019; Robaczewski et al. 2021; Chartomatsidis et al. 2016).

Finally, the system displayed user-friendliness for this population. In comparison to studies utilizing traditional cable connections (); the NAO-RBI system enhanced convenience by employing Wi-Fi transmission, which ensures rapid responsiveness and high stability. This approach aimed to reduce unnecessary delays and error rates, alleviating anxiety among this population. NAO could autonomously execute corresponding targeted behaviors without the need for complex calculations to translate human body movements into the NAO robot's limb movements. The pre-set design-maintained consistency in each repeated execution, thereby preventing distraction and minimizing potential learning interference for children with ASD during repetitive practice sessions. However, the NAO-RBI system could identify the body movement of each participant with ASD and help this population improve and practice repetitive body language interaction skills in a safe environment without being distracted. Three participants were found to be highly interested in the NAO robot, which offered a tangible, visualization, and realistic presentation. It presented a cute, friendly, and humanoid appearance, and offered immediate feedback, which attracted the participants' attention and maintained their learning motivation (Kaboski et al. 2015; Robaczewski et al. 2021; Belpaeme et al. 2018).

10 Conclusion

The purpose of this study was to investigate the effects of using the NAO-RBI system and to explore the performance of body language gestures from children with ASD. The NAO-RBI system was able to exhibit seven body language gestures ("Raising a hand," "High-five," "Handshake," "Please give me," and "Come on!" "Arms akimbo with an angry face," and "Please!"). All the body-language gestures performed by participants with ASD could also be accurately recognized by the NAO robot, and the NAO-RBI system executed them without error or delay in execution. Further research expects to extend the functional modules in the NAO-RBI system to different body language gestures (e.g., welcoming others or speaking) and to increase diversity and flexibility functions. This robot-based approach can also be used to integrate teaching strategies and social content for children with special needs.

Acknowledgements We would like to thank Ministry of Science and Technology, Republic of China, Taiwan under grant number MOST 110-2511-H-018-009 for their financial support to this study. We also deeply appreciate the students with autism spectrum disorder, their teachers, and their parents who generously participated in this study. Their willingness to volunteer was invaluable to the success of our research.

Authors Contributions All authors read and approved the final manuscript. Yu-fang Cheng*4[0000-0003-1831-8757]: Corresponding author—Conceptualization, Methodology, Data Curation, Analysis, Supervision, Writing—Original Draft. Wei-sheng Lin: System development, Experiment, Data collection, Validation. Xin-gu Peng: Investigation, Data analysis.

Funding This study was supported by Ministry of Science and Technology, Republic of China, Taiwan under grant number MOST 110-2511-H-018-009.

Data Availability The data supporting the findings of this study are available within the article "Intelligent Robots, Alongside Kinect Tools, Detect Body Language in Children with Autism Spectrum Disorder", its supplementary materials. Further inquiries can be directed to the corresponding author.

Declarations

Conflicts of Interest The authors declare no competing interests.

Ethical Approval and Consent to Participate The experimental study received approval from the Institutional Review Board (IRB) of National Changhua University of Education, Taiwan (Approval Number: NCUEREC-111–047) to ensure that research involving participants with autism spectrum disorder (ASD) was conducted ethically. Informed consent was obtained from all participants or their legal guardians prior to their inclusion in the study.

Consent for publication I, the undersigned, give my consent for the publication of identifiable details, which can include photograph(s) and/or videos and/or case history and/or details within the text ("Material") to be published in the above Journal and Article. I confirm that I have seen and been given the opportunity to read both the Material and the Article (as attached) to be published by Taylor & Francis. I have discussed this consent form with _ Wei-sheng Lin, Xin-gu Peng, _, who are authors of this paper.

References

Abdi J, Al-Hindawi A, Ng T, Vizcaychipi MP (2018) Scoping review on the use of socially assistive robot technology in elderly care. In: BMJ Open. England, vol 8, p e018815 https://doi.org/10.1136/bmjopen-2017-018815

Alshurman, W., Alsreaa, I.: The efficiency of peer taught of developing non-verbal communication to children with autism spectrum disorder (ASD). J. Educ. Pract. **6**(29), 33–38 (2015)

Assad-Zaman, M., Islam, M.R., Rahman, M.H., Wang, Y.C., McGonigle, E.: Kinect-controlled NAO robot for telerehabilitation. J. Intell. Syst. **30**(1), 224–239 (2021). https://doi.org/10.1515/jisys-2019-0126

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., Tanaka, F.: Social robots for education: A review. Sci. Robot. **3**(21), 5954 (2018). https://doi.org/10.1126/scirobotics.aat5954

Bleuer, H., Bouri, M., Mandada, F.C., Boudon, R.: New trends in medical and service robots. Springer, London (2017)

Cadette, J.N., Wilson, C.L., Brady, M.P., Dukes, C., Bennett, K.D.: The effectiveness of direct instruction in teaching students with an autism spectrum disorder to answer "Wh-" questions. J. Autism Dev. Disord. **46**(9), 2968–2978 (2016). https://doi.org/10.1007/s10803-016-2825-2

Chartomatsidis, M., Androulakis, E., Kavallieratou, E.: Training NAO used kinect. Res. Comput. Sci. **123**(1), 27–37 (2016)

Chevalier, P., Raiola, G., Martin, J.C., Isabledu, B., Bazile, C., Tapus, A.: Do sensory preferences of children with autism impact an imitation tasked with a robot? In 2017 12th ACM/IEEE International Conference on Human-Robot Interactedion HRI: 177–186. https://doi.org/10.1145/2909824.3020234 (2017)

Costa, A., Schweich, T., Charpiot, L., Stefgen, G.: Attitudes of children with autism towards robots: an exploratory study. arXiv :180607805. https://doi.org/10.48550/arXiv.1806.07805 (2018)

Cummings, L.: Handbook of pragmatic language disorders: complex and underserved populations. Springer Nature, Berlin (2021)

Doody, J.P., Bull, P.: Asperger's syndrome and the decoding of boredom, interest, and disagreement from body posture. J. Nonverbal Behav. **35**(2), 87–100 (2011). https://doi.org/10.1007/s10919-010-0103-0

Doody, J.P., Bull, P.: Evidence for impaired verbal identification but intact nonverbal recognition of fearful body postures in Asperger's syndrome. J. Autism Dev. Disord. **43**(7), 1652–1661 (2013). https://doi.org/10.1007/s10803-012-1715-5

Duh, J.-J.: Single subject research. Psychological Publishing, London (2006)

Febriantini, W.A., Fitriati, R., Oktaviani, L.: An analysis of verbal and non-verbal communication in autistic children. J. Res. Lang. Educ. **2**(1), 53–56 (2021). https://doi.org/10.33365/jorle.v2i1.923

Firman, M., Mac Aodha, O., Julier, S., & Brostow, G. J. (2016). Structured prediction of unobserved voxels from a single depth image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5431–5440).

Geminiani, A., Santos, L., Casellato, C., Farabbi, A., Farella, N., Santos-Victor, J., Olivieri, I., Pedrocchi, A.: Design and validation of two embodied mirroring setups for interactedive games with autistic children used the NAO humanoid robot. In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 1641–1644). https://doi.org/10.1109/EMBC.2019.8857576 (2019)

Hong, L.-Y., Chen, F.-F., Huang, Y.-H., You-Fen, Xu.: Social skills training course examples. National Taipei Normal University, Department of Special Education, Taipei (1999)

Ivani, A.S., Giubergia, A., Santos, L., Geminiani, A., Annunziata, S., Caglio, A., Pedrocchi, A.: A gesture recognition algorithm in a robot therapy for ASD children. Biomed. Signal Process. Control **74**, 103512 (2022). https://doi.org/10.1016/j.bspc.2022.103512

Kaboski, J.R., Diehl, J.J., Beriont, J., Crowell, C.R., Villano, M., Wier, K., Tang, K.: Brief report: A pilot summer robotics camp to reduce social anxiety and improve social/vocational skills in adolescents with ASD. J. Autism Dev. Disord. **45**(12), 3862–3869 (2015). https://doi.org/10.1007/s10803-014-2153-3

Kumazaki, H., Muramatsu, T., Yoshikawa, Y., Corbett, B.A., Matsumoto, Y., Higashida, H., Yuhi, T., Ishiguro, H., Mimura, M., Kikuchi, M.: Job interview training targeting nonverbal communication used an android robot for individuals with autism spectrum disorder. Autism **23**(6), 1586–1595 (2019). https://doi.org/10.1177/1362361319827134

Lee, I.J.: Kinect-for-windows with augmented reality in an interactedive roleplay system for children with an autism spectrum disorder. Interact. Learn. Environ. **29**(4), 688–704 (2020). https://doi.org/10.1080/10494820.2019.1710851

Liu, X., Zhou, X., Liu, C., Wang, J., Zhou, X., Xu, N., & Jiang, A. (2016, July). An interactive training system of motor learning by imitation and speech instructions for children with autism. In 2016 9th International Conference on Human System Interactions (HSI) (pp. 56–61). IEEE. https://doi.org/10.1109/HSI.2016.7529609

Muhle, R., Trentacoste, S.V., Rapin, I.: The genetics of autism. Pediatrics **113**(5), e472–e486 (2004a). https://doi.org/10.1542/peds.113.5.e472

Muhle, R., Trentacoste, S.V., Rapin, I.: The genetics of autism. Pediatrics (2004b). https://doi.org/10.1016/j.spen.2004.07.003

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., Seibert, J.: Early social communication scales (ESCS). University of Miami, Coral Gableds, FL (2003)

Qidwai, U., Kashem, S.B.A., Conor, O.: Humanoid robot as a Taugher's assistant: helping children with autism to learn social and academic skills. J. Intell. Rob. Syst. **98**(3), 759–770 (2020). https://doi.org/10.1007/s10846-019-01075-1

Reynhout, G., Carter, M.: Evaluation of the efficacy of social stories™ using three single subject metrics. Res. Autism Spectrum Disord. **5**(2), 885–900 (2011). https://doi.org/10.1016/j.rasd.2010.10.003

Robaczewski, A., Bouchard, J., Bouchard, K., Gaboury, S.: Socially assistive robots: The specific case of the NAO. Int. J. Soc. Rob. **13**(4), 795–831 (2021). https://doi.org/10.1007/s12369-020-00664-7

Robins, B., Dautenhahn, K., Dubowski, J.: Robots as isolators or mediators for children with autism a cautionary tale. In Procs of the AISB 05 Symposium on Robot Companions. AISB. (2005)

Roshidul, H., Shariff, A.R.B.M., Blackmore, B.S., Aris, I.B., Ramli, A.R.B., Hossen, J.: High adoption of behavior based robotics in the autonomous machines. J. Inf. Syst. Technol. Plann. **3**(6), 30–41 (2010)

Santos, L., Geminiani, A., Olivieri, I., Santos-Victor, J., Pedrocchi, A. (2020). CopyRobot: Interactive Mirroring Robotics Game for ASD Children. In: Henriques, J., Neves, N., de Carvalho, P. (eds) XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019. MEDICON 2019. IFMBE Proceedings, vol 76. Springer, Cham. https://doi.org/10.1007/978-3-030-31635-8_239

Sheridan, T.B.: Telerobotics, automation, and human supervisory control. MIT press, Cambridge (1992)

Shi, Z., Groechel, T.R., Jain, S., Chima, K., Rudovic, O., Matarić, M.J.: Toward personalized affect-aware socially assistive robot tutors for long-term interventions with children with autism. ACM Trans. Hum.-Robot Interact. (THRI) **11**(4), 1–28 (2022). https://doi.org/10.1145/3526111

Stoll, S., Camgoz, N.C., Hadfield, S., et al.: Text2Sign: towards sign language production using neural machine translation and generative adversarial networks. Int. J. Comput. vis. **128**, 891–908 (2020). https://doi.org/10.1007/s11263-019-01281-2

Tariq, S., Baber, S., Ashfaq, A., Ayaz, Y., Naveed, M., Mohsin, S. (2016). Interactive Therapy Approach Through Collaborative Physical Play Between a Socially Assistive Humanoid Robot and Children with Autism Spectrum Disorder. In: Agah, A., Cabibihan, JJ., Howard, A., Salichs, M., He, H. (eds) Social Robotics. ICSR 2016. Lecture Notes in Computer Science(), vol 9979. Springer, Cham. https://doi.org/10.1007/978-3-319-47437-3_55

Vishne, G., Jacoby, N., Malinovitch, T., Epstein, T., Frenkel, O., Ahissar, M.: Slow update of internal representations impedes synchronization in autism. Nat. Commun. **12**(1), 5439 (2021). https://doi.org/10.1038/s41467-021-25740-y

Yavsan, E., Uçar, A.: Gesture imitation and recognition used Kinect sensor and extreme learned machines. Measurement **94**, 852–861 (2016). https://doi.org/10.1016/j.measurement.2016.09.026

Wei-sheng Lin is a postgraduate in the Master's Program in e-Learning at the Department of Industrial Education and Technology, National Changhua University of Education, Taiwan. His academic interests are centered around the integration of cutting-edge technologies into educational environments, with a particular focus on the development of using a humanoid robot helps children with Autism Spectrum Disorder practicing body-language skills.

Xiang-yu Peng is a postgraduate in the Master's Program in e-Learning at the Department of Industrial Education and Technology, National Changhua University of Education. His academic interests focus on integrating cutting-edge technologies into educational environments, with particular emphasis on applications of large language models in artificial intelligence, natural language processing, and computer vision in robotic systems.

Yu-fang Cheng is a professor in the Master Program of e-Learning at the Department of Industrial Education and Technology, National Changhua University of Education, Taiwan. She also serves as the Director of the Research Lab for e-Learning Technology for Children with Autism Spectrum Disorders (ASDs). Prof. Cheng received her PhD from the Department of Information and Engineering at Leeds Metropolitan University, UK. With nearly 20 years of contributions in the field of computer technology for individuals with ASDs, Prof. Cheng and her research team have successfully developed VR robot applications and interactive systems aimed at addressing the challenges faced by children with ASDs. Her research interests primarily focus on robot-assisted interactions for special education.