


# Effects of robots' intonation and bodily appearance on robot-mediated communicative treatment outcomes for children with autism spectrum disorder

Caroline L. van Straten<sup>1</sup> · Iris Smeekens<sup>2,3</sup> · Emilia Barakova<sup>4</sup> · Jeffrey Glennon<sup>2</sup> · Jan Buitelaar<sup>2,3</sup> · Aoju Chen<sup>5</sup> 

Received: 23 December 2016 / Accepted: 7 July 2017 / Published online: 1 August 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** Previous research has suggested that robot-mediated therapy is effective in the treatment of children with Autism Spectrum Disorder (ASD), but not all robots seem equally appropriate for this purpose. We investigated in an exploratory study whether a robot's intonation (monotonous vs. normal) and bodily appearance (mechanical vs. humanized) influence the treatment outcomes of Pivotal Response Treatment (PRT) sessions for children with ASD. The children (age range 4–8 years) played puzzle games with a robot which required communication with the robot. The treatment outcomes were measured in terms of both task performance and affective states. We have found that intonation and bodily appearance have an effect on children's affective states but not on task performance. Specifically, humanized bodily appearance leads to more positive affective states in general and a higher degree of interest in the interaction than mechanical bodily appearance. Congruence between bodily appearance and intonation triggers a higher degree of happiness in children with ASD than incongruence between these two factors.

**Keywords** Children · Autism Spectrum Disorder · Robot-mediated therapy · Intonation · Bodily appearance · Socially Assistive Robotics

## 1 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that is characterized by persistent deficits in social interaction and communication and restricted, repetitive patterns of behavior, interests or activities [1]. Regarding social communication skills, children with ASD often encounter problems with functional initiatives (e.g., requesting objects or activities, asking for help), social initiatives (e.g., sharing information and feelings) and appropriate social responses, which may contribute to deficits in reciprocal social interaction [39]. Since social communication deficits are regarded as core deficits in children with ASD, the majority of intervention studies focus on improvement of these skills, incorporating a variety of treatment models [52]. According to these intervention studies, technology-aided intervention shows promising results in improvement of social communication skills in ASD [52].

A technological application that has received increased attention over the last decade is the use of Socially Assistive Robotics (SAR) [20]. A number of reasons have been proposed for why the use of SAR may be promising. First, technological artifacts may be intrinsically appealing to children with ASD. Their great interest in such objects may in turn provide them with a high level of motivation to interact with the robots they are presented with during therapy sessions [20, 2, 49]. Furthermore, robots' behavior is simple and may thus be more predictable than the complex behavior that humans exhibit. Predictability has been shown to be appealing to children with ASD [23]. Additionally, children with ASD have strong desires to have control over a situation [17]. Interacting with a robot can provide them with such a feeling of being in control. Positive outcomes of SAR-mediated training methods

✉ Aoju Chen  
aoju.chen@uu.nl

<sup>1</sup> Amsterdam School of Communication Research, University of Amsterdam, PO Box 15791, 1001 NG Amsterdam, The Netherlands

<sup>2</sup> Department of Cognitive Neuroscience, Radboud University Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands

<sup>3</sup> Karakter, Centre for Child and Adolescent Psychiatry, Reinier Postlaan 12, 6525 GC Nijmegen, The Netherlands

<sup>4</sup> Industrial Design Department, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands

<sup>5</sup> Department of Languages, Literature and Communication and Utrecht Institute of Linguistics, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands

have been reported in various studies (albeit with small sample sizes). For instance, it has been found that children with ASD tend to produce more verbal utterances in therapy sessions when interacting with a robot than with a human interlocutor [21]. Children with ASD also appear to learn target social skills faster from feedback provided through a robot than from identical feedback provided by a therapist [13].

Despite the promising effects of robots on children with ASD in treatment sessions, not every robot seems to be equally appropriate to be used for this purpose. More specifically, Robins, Dautenhahn and Dubowski [45] reported that children with ASD performed better during interaction experiments with doll-like robots with non-humanlike facial appearance than during identical experimental sessions with the same robots with humanlike facial appearance, and tended to avoid looking at the robot's face when it contained humanlike features. They suggested that these results might be explained by the fact that the absence of humanlike facial features and related complexity in the movement of facial muscles made the robot's appearance far less complex to process. This feature of robots can constitute an advantage to children with ASD, considering that they tend to show more sensory overstimulation than children without ASD [30]. However, other studies suggest that bodily appearance should also be taken into account. For instance, Billard, Robins, Nadel and Dautenhahn [10] found that children with ASD responded positively to a doll-like robot with a mechanical-looking body but a humanlike face and avoided looking at the robot's face to a lesser extent than reported in Robins et al.'s [45] study. Moreover, the robot used in Billard et al.'s study provoked comparable responses from children with ASD to responses triggered by a non-humanoid, completely mechanical-looking robot [9]. These findings suggest that an appropriate combination of humanlike and mechanical features in the appearance of the robots used in treatment programs is important to children with ASD [12].

However, little is known on the effects of the robots' vocal features on children with ASD. Previous studies of communicative ASD treatment either used a pre-recorded human female voice (e.g., [5, 27]) or a synthetic voice (e.g., [26]). Current speech synthesis technology makes it possible to generate natural-sounding speech [15]. The stereotypical voice of robots as featured in science fiction films and TV series is, however, a mechanical-sounding voice that has little intonation. Intonation refers to variation in pitch, duration, and intensity in speech and conveys meaning at the utterance-level [16, 37]. Pitch is the most important parameter of intonation. For example, the one-word utterance "Coffee" can be said with a rising pitch pattern or a falling pitch pattern. The former means a question (e.g., Is this coffee? Do you want coffee?), whereas the latter means a statement (e.g., This is coffee. Or the speaker would like to have coffee.). In research on language development, language games that use virtual robots to

elicit spontaneous verbal responses from typically developing children (4 to 11 years) also adopt a monotonous voice for the virtual robots (e.g., [3, 47]). In these games, children are supposed to teach a robot to speak human language properly. It has been reported that children always reconstructed the robots' utterances with intonation as used in their own speech but they hardly corrected infelicitous uses of word order (e.g., the robot says "The wall is building the girl" instead of "The girl is building a wall"). This result suggests that typically developing children appear to consider the presence of intonation as a crucial feature that distinguishes human speech from robot speech. Children with ASD have been reported to produce atypical intonation, which has been variably described as sounding "robotic," "monotone," "overprecise," or "sing-song" ([18] and references therein), and have difficulties with using intonation appropriately in communication [40]. However, children with ASD may show strong pitch discrimination capacities in both speech and non-speech material [43] and can make use of intonational information to process syntactically ambiguous sentences [19]. These findings suggest that children with ASD and normal (verbal) intelligence can perceive the acoustic differences between monotonous speech and speech with normal intonation and may associate human speakers with speech with normal intonation. It remains to be investigated whether the presence and absence of normal intonation in a robot has any influence on its interaction with children with ASD.

Against this background, the current study examines the effects of intonation (normal vs. monotonous) combined with different bodily appearance of an embodied robot (mechanical vs. humanized) on treatment outcomes of robot-mediated therapy sessions for children with ASD. In so doing, we will gain more insight into what features of social robots are likely to generate positive treatment outcomes in these children in general. Although the reported positive impact of robots in treatment of social communication with children with ASD awaits further assessment due to the large heterogeneity among children with ASD and the small sample size in previous studies [29], we consider it valuable to obtain more knowledge on the effect of features of social robots on children with ASD because of the increasing trend to use robots in ASD treatment (e.g., [13, 20, 21, 49]). We operationalize treatment outcomes as not only children's performance in interaction tasks, as commonly done in clinical research, but also their affective states during treatment sessions. As motivation is a "pivotal" (key) area in the treatment for children with ASD, measuring whether children with ASD enjoy their interaction is useful in assessing whether the treatment is producing clinically significant changes [33].

With respect to the effect of intonation alone, as a monotonous voice emphasizes a robot's non-human, technological nature and may appeal to children with ASD, we predict that a monotonous voice will lead to better treatment outcomes in

terms of better performance in interaction tasks and higher positivity in affective states than a voice with normal intonation (as used in speech of speakers with no speech or hearing deficits). With respect to the effect of bodily appearance alone, it has been suggested that the use of a robot with humanlike bodily appearance (combined with non-humanlike facial appearance) triggers better treatment outcomes than the use of a robot that is humanized in both its facial and bodily appearance [45]. Relatedly, there were indications that children with ASD pay more attention to the mechanical body parts of a robot than to the interaction task [36], suggesting that humanized bodily appearance may be less distracting. Also, humanization of the robot's bodily appearance may induce a peer-like perception in children with ASD [42, 44]. We thus predict that humanizing a robot's bodily appearance (without changing its mechanical facial appearance) will result in better treatment outcomes, compared with presenting a robot to children in its regular, mechanical appearance.

Considering the two factors together, we see several possible scenarios. First, assuming that the effect of intonation and the effect of bodily appearance are additive, we predict that the combination of a monotonous voice and humanized bodily appearance will lead to better treatment outcomes than the other combinations of voice and bodily appearance. However, the effect of intonation and the effect of bodily appearance may not be simply additive. Two factors may play a role in the interaction between intonation and bodily appearance. On the one hand, it has been found that children with ASD are attracted to predictability [23]. If children with ASD associate a monotonous voice with a robot, like typically developing children, a monotonous voice combined with mechanical bodily appearance would be more predictable to them than a voice with normal intonation combined with either humanized bodily appearance or mechanical bodily appearance. From the perspective of predictability, the use of a robot with a monotonous voice and mechanical bodily appearance would thus result in the best treatment outcomes. On the other hand, it has been shown that individuals with ASD exhibit stronger neural responses to audiovisual incongruence (e.g., a picture of a pig combined with the barking sound of a dog) than healthy controls [46]. This finding suggests that congruency between intonation and bodily appearance in a robot can matter to the treatment outcomes. Congruence can be reached in two ways: both intonation and bodily appearance emphasize the robot's mechanical nature, or both express its human-likeness. From the perspective of congruence, the use of a robot whose intonation and bodily appearance are congruent would lead to better treatment outcomes than the use of a robot whose intonation and bodily appearance are incongruent.

These predictions have been tested in an exploratory study of children with ASD who participated in a larger randomized clinical trial, investigating the effectiveness of using a robot within Pivotal Response Treatment (PRT) for improving social

communication in children with ASD (the “PicASSo” project). PRT is a well-studied and promising treatment model for ASD, which focuses on training “pivotal” (key) skills in ASD with the goal of facilitating improvements in both these key skills and collateral gains in other areas of functioning [33].

## 2 Method

As mentioned above, this study was embedded in the “PicASSo” project, which focuses on investigating the effectiveness of PRT for young children with ASD. Adjustments to the PRT treatment sessions for the purpose of the current study were restricted to the bodily appearance and voice of the robot and thus kept minimal. A mixed repeated measures experimental design was adopted, with intonation as a within-subject factor and appearance as a between-subject factor. This resulted in four experimental conditions: (1) mechanical bodily appearance-normal intonation, (2) mechanical bodily appearance-monotonous intonation, (3) humanized bodily appearance-normal intonation, and (4) humanized bodily appearance-monotonous intonation.

### 2.1 Participants

Eight children from the PicASSo project participated in this study. The children were diagnosed with ASD according to the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM) [1]. Diagnosis was provided by a child psychiatrist and a health care psychologist, based on clinical evaluation and information provided by parents (and other caregivers) on patient history. The diagnoses were additionally confirmed by the Autism Diagnostic Observation Schedule (ADOS) [8]. The children could speak with single words at minimum and at least one of the parents spoke Dutch to the child. The children were aged between 4 to 8 years ( $M = 6.70$ ,  $SD = 1.52$ ), as previous research demonstrated the responsiveness of children in this age range to PRT sessions [4]. All children were native speakers of Dutch with no history of audiovisual or additional neurological deficits. They had an intelligence quotient (IQ) of 70 or higher ( $M = 98.13$ ,  $SD = 11.93$ ) as measured by the Wechsler Intelligence Scale for Children (WISC) [35] or the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) [25]. The medication dose that the children received was fixed during the course of the study to minimize any potential effects on treatment outcomes.

The children were divided into two groups. One group was assigned to the “mechanical bodily appearance” conditions, the other group to the “humanized bodily appearance” conditions. The two groups were matched for both chronological age and verbal IQ (VIQ) scores (Table 1). Independent samples  $t$  tests showed no significant differences in age ( $t(6) = .037$ ,  $p > .05$ ) and VIQ ( $t(6) = .360$ ,  $p > .05$ ) between the groups.

## 2.2 Experimental task

Each child participated in two treatment sessions (hereafter trials). On each trial, the child played a puzzle game with a robot. In this game, the child completed three animal-shaped puzzles of his choice. Crucially, the child could not complete any of the puzzles on his own and needed to cooperate with the robot. For instance, the puzzle boxes were kept in a large trunk, which the child could not open. In order to access the puzzle boxes, the child needed help from the robot. Further, the child could find only some pieces of each puzzle in the puzzle boxes and the robot had the rest of the pieces and the pins to attach them. The child needed to ask the robot for the remaining puzzle pieces and pins appropriately. A step-by-step description of the game is provided together with illustrating pictures in the [appendix](#).

In each puzzle game, learning moments were created using prompts (i.e., encouragement cues) based on the child's interests and abilities to train the child in social communication skills. Seven different levels of the game were created adhering to specific treatment goals in communication for each child (i.e., varying from simple functional initiatives to more complex social initiatives). For each learning moment, the game scenario includes three prompts of increasing explicitness, i.e., an open question prompt (i.e., "What could you ask/tell me now?"), a fill-in prompt (i.e., "Fill in, May I..."), and a tell prompt (i.e., "Say after me, may I have the puzzle pieces?").

## 2.3 Apparatus

The robot used in the PicASSo project is the humanoid robot NAO (V4, Fig. 1), equipped with the NaoQi 1.14.1 software development kit (Aldebaran Robotics). NAO is 58 cm tall and has 25 degrees of freedom that allows for flexible, humanlike movements. It has simple features, making it appropriate for usage with children with ASD. Its speech is produced by a speech synthesizer and played via integrated loud speakers. In the PicASSo project, NAO has adopted the synthetic voice "Jasmijn."

For the current experiment, the voice of Jasmijn was adapted by raising the mean pitch from 166 to 232 Hz for the normal-intonation condition. This manipulation reserved some features of voice Jasmijn and yet made the voice sound sufficiently different to be presented as another speaker. Also, the increased pitch made the voice sound more child-directed than it did originally. The voice used in the monotonous-intonation conditions was created using the following procedure. First, NAO's utterances produced in the normal-intonation condition were recorded with an H1-Zoom digital recorder positioned on an adjustable tripod stand. Second, the pitch of the recording was removed using Praat [11]. Third, a flat pitch contour with a pitch height of 232 Hz was imposed on each utterance, setting the pitch range to 0 Hz and making the voice monotonous. Since recording the robot's speech that was played via its built-in speakers resulted in reduced loudness, the loudness of

the utterances was amplified three times to match the loudness of the utterances in the normal-intonation condition. Finally, the recording was segmented into individual utterances using Audacity 2.0.6, which were uploaded onto the robot and invoked in the speech synthesizer. The voices used in different intonation conditions thus differed only in pitch-related features, while the temporal properties were the same.

NAO's bodily appearance was humanized by covering his body parts in human clothing (Fig. 2). He wore a long sleeved T-shirt and a pair of trousers, which entirely covered his torso and limbs. The colors of both garments were neutral colors that would appeal to both boys and girls.

## 2.4 Procedure

The participants completed the trials in an outpatient treatment facility for children with psychiatric disorders (Karakter expert center for child and adolescent psychiatry in Nijmegen, the Netherlands). The trials were programmed and executed on an ASUS Zenbook UX31A (OS Windows 8.1) using Tino's Visual Programming Environment (TiViPE) 2.1.2 [6, 38]. There was an interval of 1 to 2 weeks between the two trials for each participant. On each trial, a parent of the participant and a therapist were also present in the treatment room. Both the therapist and parent were seated at a certain distance from the child in order to encourage the child to focus his attention on the robot. The parent was instructed to minimize interaction with the child during the puzzle game.

To justify the differences in voice and appearance between the current treatment sessions and previous treatment sessions in which the children had encountered NAO, the therapist introduced the robot as NAO's friend and with the name of "NEO" (first trial) or "NIO" (second trial). The only difference between NAO and his friends was their voice or clothing (Fig. 3).<sup>1</sup> The therapist was further instructed to answer questions using standard answers given in a manual and to respond to unforeseen questions briefly and as neutrally as possible in terms of form and content. This way, variation between treatment sessions was minimized.

The completion of one puzzle game resulted in 15-min child-robot interaction on average. All trials were videotaped by means of a Sony Handycam HDR-CX240E video camera, mounted on an adjustable tripod, which was placed behind the robot. In addition, the participants answered three questions on their affective states, one question at the beginning of a trial (i.e., "How happy are you right now?") and two questions at the end of a trial (i.e., "How happy are you right now?" and "How do you like the robot?"). The participants indicated their answers on a five-point scale illustrated with smilies and were asked to briefly motivate

<sup>1</sup> The reader might at this point notice a gender discrepancy between the use of male pronouns and the female voice that the robot was equipped with. We chose to refer to NAO with he/him because participants in the regular sessions referred to NAO as a male robot.



**Table 1** Information on participants

Experimental group	Gender	Chronological age	VIQ score	Mean age (group)	Mean VIQ (group)
1	M	4; 04	91	6; 8	99.75
	M	6; 05	114		
	F	7; 11	93		
	M	8; 02	101		
2	F	5; 02	85	6; 8	96.5
	M	5; 07	118		
	M	7; 09	90		
	M	8; 02	93		

their ratings. The questions, however, proved to be too difficult for many participants to motivate, which is in line with the communication and emotion recognition difficulties often encountered by children with ASD. We have therefore decided to exclude the responses to the questions from further analysis.

### 3 Analysis and results

#### 3.1 Measures

The participants’ task performance was assessed via the percentage of prompts that a participant received during a trial. It was calculated by dividing the total number of prompts received by the total number of prompts available during the learning moments. As mentioned in Sect. 2.2, the participants could receive three kinds of prompts from the robot at each learning moment if

they did not take an appropriate initiative spontaneously. Each open question-prompt received a score of 1, each fill-in prompt got a score of 2, and each tell prompt got a score of 3. A lower percentage of prompts indicated better task performance (i.e., more instances of spontaneous communication).

The participants’ affective states were assessed using the Affect Rating Scale [32]. We conducted the rating using the video recordings made from each trial. The Affect Rating Scale runs from 0 to 5 and can be employed to evaluate affective states on three dimensions: demonstrated interest (towards the game as a whole, i.e., not solely towards the robot), happiness (judging from a child’s facial expressions and other behavioral observations), and appropriate behavior (i.e., performing the task without showing disruptive behavior). We divided each trial into 1-min segments and scored the child in each 1-min segment on the three dimensions, following the descriptions in Koegel, Singh and Koegel [34]. We then



**Fig. 1** NAO (Aldebaran Robotics)



**Fig. 2** NAO’s humanlike appearance

**Fig. 3** Meeting NAO's friend and playing puzzle games with him



averaged the scores from the individual segments for each dimension and obtained the mean affect scores for each child on each dimension. A higher score on each dimension corresponds to more interest, a higher degree of happiness, or more appropriate behavior during the trial. We also computed a general affect score for each child, which was the average of the three mean affect scores. The general affect score was an indication for a child's overall affective state.

A trained rater scored all the data and obtained the two types of measures from each participant. In order to ensure reliability of the scoring, a second rater scored 25% of the video data, following the recommendation of Fisher, Piazza and Roane [22]. The data consisted of four randomly selected trials, each of which represented one of the four experimental conditions, and took place at various time points during the experiment. We used the intraclass correlation coefficient (ICC) test in SPSS (IBM SPSS statistics version 22) to assess inter-rater agreement between the two raters. The ICCs for required prompting and affect scores were 1 and 0.79, respectively, suggesting excellent inter-rater agreement for all measures and thus high reliability of the scoring by the first rater [14].

### 3.2 Statistical analysis and results

We used the linear mixed effects model [24] in SPSS (IBM SPSS statistics V22) to assess the effects of the fixed factors, intonation, and bodily appearance, on the outcome variables that stemmed from three types of measures. The factor participant was included into the models as a random factor. There were five outcome variables: the average percentage of required prompts, the mean general affect score, the mean "interest" score, the mean "happiness" score, and the mean "behavior" score. Five linear mixed-effects models were performed. To control the false discovery rate (i.e., the proportion of significant results that are actually false positives) in multiple statistical analyses on the data obtained from the same participants, we adopted the Benjamini-Hochberg procedure using the standard false rate value 0.25 [7, 48], by means of the Excel spreadsheet developed by McDonald [41]. This procedure revealed two significant main effects of bodily

appearance and one significant interaction between intonation and bodily appearance.

First, the factor bodily appearance had a significant main effect on the mean general affect score. As shown in Fig. 4, humanized bodily appearance led to a higher general affect score than mechanical appearance ( $F(1, 6) = 6.588, p = .043$ ).

Second, the factor bodily appearance had a significant main effect on the mean interest scores ( $F(1, 6) = 8.446, p = .027$ ). As shown in Fig. 5, humanized bodily appearance led to a higher mean interest score than mechanical bodily appearance.

Finally, the interaction of intonation and bodily appearance was significant in the models with the mean happiness score as the outcome variable ( $F(1, 6) = 9.936, p = .020$ ). As shown in Fig. 6, normal intonation led to a higher mean happiness score when combined with humanized bodily appearance than when combined with mechanical appearance, whereas monotonous intonation led to a higher mean happiness score when combined with mechanical bodily appearance than when combined with humanized bodily appearance.

No significant main effects and interactions were found for intonation and bodily appearance on the outcome variable concerning task performance (i.e., the average percentage of required prompts).

## 4 Discussion and conclusions

We have examined the effect of intonation (normal vs. monotonous), combined with different bodily appearance, on the treatment outcomes of robot-mediated therapy sessions for children with ASD. On the basis of earlier findings, we hypothesized that both intonation and bodily appearance of robots could influence treatment outcomes on their own or in interaction.

We found no evidence that intonation had a significant effect on the performance measure. This null result might be due to a floor effect, as the mean percentage of prompts received across conditions only came down to 8.55%. This floor effect could in turn be a consequence of the previous experience that our participants had with PRT sessions. During their

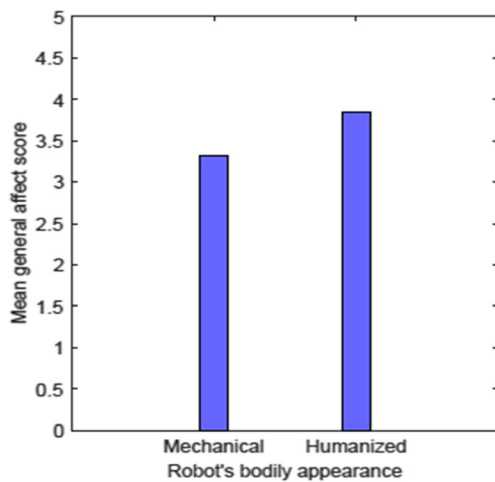


Fig. 4 Main effect of bodily appearance on general affect score

previous PRT sessions, they have learned what kind of initiatives are expected and have become more capable of showing these initiatives spontaneously. Therefore, prompting might only have been of marginal need to these children, which could have prevented experimental manipulations from further lowering the number of prompts they received. Further, the mean affect scores were slightly higher in the monotonous-intonation condition than in the normal-intonation condition in each dimension of the affect scores. This is in line with our prediction based on the attraction of children with ASD towards mechanical objects (i.e., monotonous intonation emphasizes the robot’s mechanical nature) and the difficulty with integrating different cues in speech processing in children with ASD [20, 28, 49]. The effect of intonation did not reach statistical significance in any of the models on the affect scores. This may be related to the small differences in scores between the two intonation conditions. As mentioned before, the participants had previous experience with the robot with voice Jasmijn—a female voice with normal intonation. As the voice used in the normal intonation condition resembled the voice Jasmijn more than did the voice

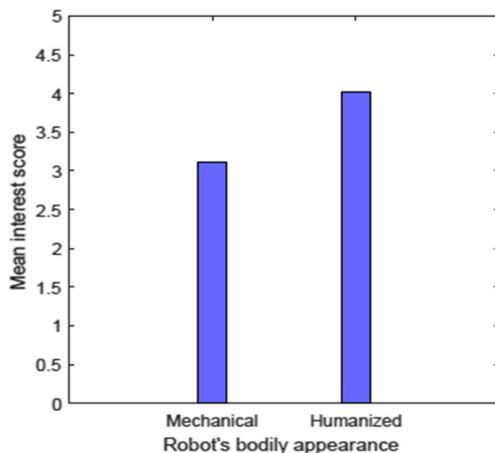


Fig. 5 Main effect of bodily appearance on interest scores

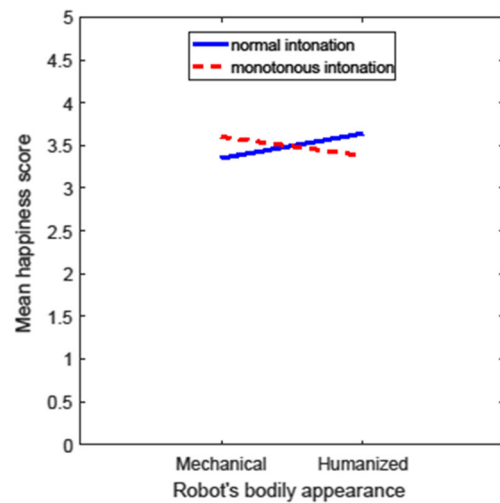


Fig. 6 Interaction effect of intonation and appearance on happiness scores

used in the monotonous condition, the preference for “sameness” in children with ASD [31] might have weakened the positive effect of a voice with monotonous intonation.

Furthermore, we found no evidence for a significant effect of bodily appearance on the performance measure, possibly due to the abovementioned floor effect. However, this factor did have a clear effect on the participants’ affective states. Namely, the participants obtained higher general affect scores as well as higher interest scores when presented with the robot in its humanized bodily appearance. These results suggest that the influence of the humanized bodily appearance, i.e., making the children focus less on the robot’s body parts [36], outperformed the “need for sameness” of children with ASD [31]. Reluctance to change might nevertheless explain why the differences between mean scores were statistically significant but were small in magnitude.

Finally, we found that intonation and bodily appearance can influence participants’ affective states. The participants who were presented with the robot with congruent appearance and intonation (either mechanical appearance + monotonous intonation or humanized appearance + normal intonation) scored higher on the happiness dimension than the participants who interacted with the robot with incongruent appearance and intonation. However, when inspecting the scores of each participant in the two intonation conditions, we have observed that three of the eight participants did not exhibit the group pattern. This observation was in line with the reported heterogeneity in individuals with ASD.

To conclude, our exploratory study suggests that humanized bodily appearance and congruence in bodily appearance and intonation can lead to better treatment outcomes concerning various dimensions of affective states. Specifically, humanized bodily appearance was related to an increase in interest and happiness in the children with ASD during the trials, compared to mechanical bodily appearance. Further, children with ASD may be sensitive to the (in)congruence of intonation and bodily appearance. Congruence in intonation and bodily appearance was related to a higher degree of

happiness than incongruence in intonation and bodily appearance. Additionally, we have observed qualitative evidence for a positive effect of monotonous intonation (in highly intelligible speech) on the affective states of children with ASD during treatment sessions. In terms of design guidelines, an implication deriving from these findings is that it may be a beneficial strategy to use a robot with congruent intonation and bodily appearance in robot-mediated therapy sessions for children with ASD.

The current study has involved a relatively small number of participants. However, it conforms to the norm of published studies of robot-mediated treatment for children with ASD, where usually three to four children were tested (see [20, 50, 51] for reviews). The small number of participants in social robotics studies of children with ASD is caused by several reasons. For instance, research groups that perform such studies typically have fewer resources than clinical groups investigating other ASD treatment methods. Moreover, as these research groups generally have access to only one robot, it is difficult to conduct large-scale evaluations [50]. However, we emphasize that it is important to address the small-sample-size problem in social robotics studies of children with ASD in the interest of the validity of the findings and because of the heterogeneity in individuals with ASD. In contrast to other social robotics studies that are mostly qualitative by nature, we tried to circumvent the limitations of a small sample size by obtaining a relatively large amount of quantitative data and using statistical analysis to explore the effect of the experimental factors on the treatment outcomes. However, we should point out that while statistical testing gives an indication to the probability of genuine findings, caution should be taken when drawing conclusions from data with small sample sizes. Future research is needed to

validate our findings. Our recommendations for future studies are (1) involving a larger number of participants and increasing the number of trials per participant to reduce between-participant variance, (2) testing participants without previous experience with the robot such that familiarity effects can be ruled out, and (3) testing participants who have no or limited experience with PRT therapy sessions to minimize the floor effect observed in the average percentage of required prompts. In addition, longitudinal studies within clinical facilities and in the child's natural environment are needed to examine, among other possible factors of effectiveness, the dynamics in the effects of intonation and bodily appearance of robots used in therapy sessions over time and the generalizability of obtained effects in children's communication in everyday life.

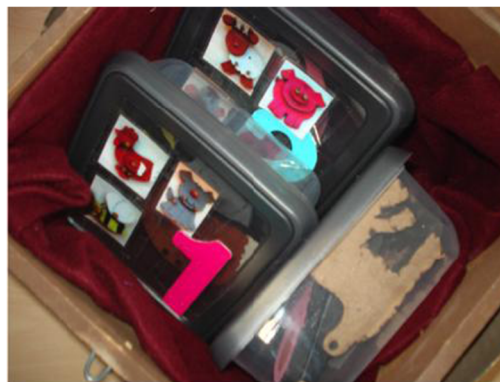
**Acknowledgements** The authors are grateful to ZonMw, the Netherlands Organization for Health Research and Development (project number 95103010, ZonMW Programma Translatie Onderzoek), for sponsoring the PicASSo project. AC was supported by a VIDI grant from the Netherlands Organization for Scientific Research (NWO-276-76889-001). We also wish to thank Martine van Dongen, Wouter Staal, and Iris Oosterling (PicASSo project) for their input in the course of this project, Tino Lourens for his help with the TiViPE programming, Chris Janssen for feedback on an earlier version of the manuscript, and Karakter expert center for child and adolescent psychiatry for making this cooperation possible. Finally, a big thank-you goes to all participating children and their parents.

#### Compliance with ethical standards

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

## Appendix

### Game illustrations

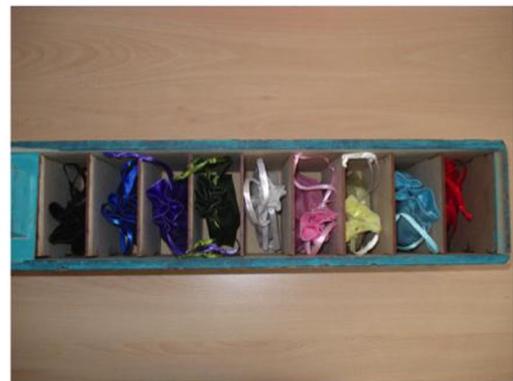




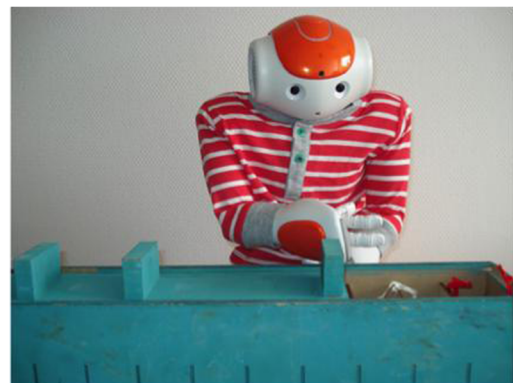
1. At the start of the game, the child sits in front of the robot; a big brown trunk is located on his left side. After appropriate child-robot interaction takes place (e.g., greeting, noticing that the box is tied up, asking for scissors), the child gains access to the trunk. There are three smaller boxes in the trunk, each containing one puzzle piece of each animal puzzle depicted on its cover.



2. The child takes the puzzle boxes and the trunk is removed by the therapist. The robot then asks the child to pick one of the puzzle boxes. The therapist removes the two remaining ones.



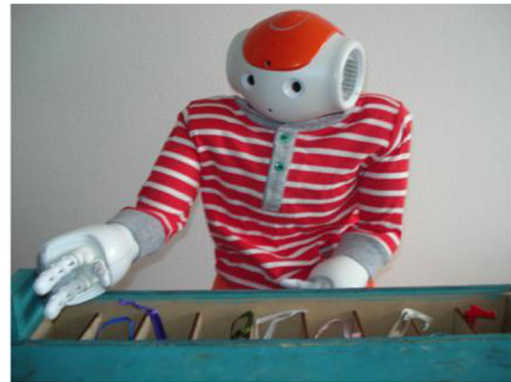
3. Then, the therapist puts a blue trunk with a sliding cover on the top in front of the child. There are nine colored bags in the trunk, which contain the missing pieces of all animal puzzles. For some game levels, the bags also contain pins that should be used to attach the various pieces to each other. For others game levels, the child needs to ask the robot for the pins in an appropriate way. The pins are provided only after appropriate child-robot interaction.



4. Upon starting the first puzzle, the child notices that the materials provided are insufficient. Following appropriate interaction (i.e., asking for the remaining pieces), the robot opens the cover of the trunk a bit such that three colored bags are in sight and tells the child which bag contains the right puzzle pieces.



5. The child then takes the correct bag from the blue box, thus gaining access to the remaining materials. He can now construct the first puzzle with these materials. During this process, functional as well as social learning opportunities are provided (e.g., asking for help, showing interest). Once the puzzle is completed, the robot makes the sound of the animal in question if the child asks for it appropriately.



6. This procedure repeats itself two more times, such that the remaining two animal puzzles can be completed by the child. After all three puzzles are put together, some additional learning opportunities are presented to the child that involve social interaction and tidying up. Finally, the robot and child say goodbye to each other and the game ends.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, 5th edn. American Psychiatric Association, Arlington
2. Aresti-Bartolome N, Garcia-Zapirain B (2014) Technologies as support tools for persons with autistic spectrum disorder:

- a systematic review. *Int J Environ Res Public Health* 11: 7767–7802
3. Arnhold A, Chen A, Järvikivi J (2016) Acquiring complex focus-marking: Finnish four- to five-year-olds use prosody and word order in interaction. *Frontiers in Psychology* 7:1886
  4. Baker-Ericzén MJ, Stahmer AC, Burns A (2007) Child demographics associated with outcomes in a community-based pivotal response training program. *J Posit Behav Interv* 9(1): 52–60
  5. Barakova EI, Bajracharya P, Willemsen M, Lourens T, Huskens B (2015) Long-term LEGO therapy with humanoid robot for children with ASD. *Expert Syst* 32(6):698–709
  6. Barakova EI, Gillessen J, Huskens B, Lourens T (2013) End-user programming architecture facilitates the uptake of robots in social therapies. *Robot Auton Syst* 61(7):704–713
  7. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 57:289–300
  8. De Bildt A, Greaves-Lord K, de Jonge M (2013) ADOS-2: Autisme Diagnostisch Observatie Schema. Hogrefe, Handleiding Amsterdam
  9. Billard A (2003) Robota: clever toy and educational tool. *Robot Auton Syst* 42:259–269
  10. Billard A, Robins B, Nadel J, Dautenhahn K (2007) Building Robota, a mini-humanoid robot for the rehabilitation of children with autism. *Assist Technol* 19:37–49
  11. Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer [Computer program] Version 6.0.05, retrieved 12 November 2015 from <http://www.praat.org/>
  12. Cabibihan JJ, Javed H, Ang M (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
  13. Charlop-Christy MH, Le L, Freeman KA (2000) A comparison of video modeling with in-vivo modeling for teaching children with autism. *J Autism Dev Disord* 30(6):537–552
  14. Cicchetti DV (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 6(4):284–290
  15. Crumpton J, Bethel CL (2016) A survey of using vocal prosody to convey emotion in robot speech. *Int J Soc Robot* 8(2):271–285
  16. Cruttenden A (1997) Intonation. Cambridge University Press, Cambridge
  17. Dautenhahn K, Werry I (2004) Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmat Cogn* 12(1):1–35
  18. DePape AR, Chen A, Hall GBC, Trainor LJ (2012) Use of prosody and information structure in high functioning adults with autism in relation to language ability. *Front Psychol* 3:1–13
  19. Diehl JJ, Friedberg C, Paul R, Snedeker J (2015) The use of prosody during syntactic processing in children and adolescents with autism spectrum disorders. *Dev Psychopathol* 27:867–884
  20. Diehl JJ, Schmitt LM, Villano M, Crowell CR (2012) The clinical use of robots for individuals with autism spectrum disorders: a critical review. *Res Autism Spect Dis* 6:249–262
  21. Feil-Seifer D, Matorić MJ (2009) Towards socially assistive robotics for augmenting interventions for children with autism spectrum disorders. In: Khatib O, Kumar V, Pappas J (eds) *Experimental robotics*. Springer, Berlin, pp 201–210
  22. Fisher WW, Piazza CC, Roane HS (eds) (2011) *Handbook of applied behavior analysis*. The Guilford Press, New York
  23. Flannery KB, Horner RH (1994) The relationship between predictability and problem behavior for students with severe disabilities. *J Behav Educ* 4:157–176
  24. Gelman A, Hill J (2007) *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, Cambridge
  25. Hendriksen J, Hurks P (2009) *Wechsler preschool and primary scale of intelligence - third edition—Nederlandse bewerking (WPPSI-III-NL)*. Pearson, Amsterdam
  26. Huskens B, Palmén A, van der Werff M, Lourens T, Barakova EI (2015) Improving collaborative play between children with autism spectrum disorders and their siblings: the effectiveness of a robot-mediated intervention based on Lego therapy. *J Autism Dev Disord* 45:3746–3755
  27. Huskens B, Verschuur R, Gillessen J, Didden R, Barakova EI (2013) Promoting question-asking in school-aged children with autism spectrum disorders: effectiveness of a robot intervention compared to a human-trainer intervention. *J Dev Neurorehabil* 16(5):345–356
  28. Järvinen-Pasley A, Wallace GL, Ramus F, Happé F, Heaton P (2008) Enhanced perceptual processing of speech in autism. *Dev Sci* 11(1):109–121
  29. Jeste SS, Geschwind DH (2014) Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol* 10(2):74–81
  30. Johnson CP, Myers SM (2007) Identification and evaluation of children with autism spectrum disorders. *Pediatrics* 120(5):183–215
  31. Kanner L (1943) Autistic disturbances of affective contact. *Nerv Child* 2:217–250
  32. Koegel RL, Egel AL (1979) Motivating autistic children. *J Abnorm Psychol* 88(4):418–426
  33. Koegel RL, Koegel LK (2006) *Pivotal response treatment for autism: communication, social, & academic development*. Paul H. Brookes Publishing Co., Baltimore
  34. Koegel LK, Singh AK, Koegel RL (2010) Improving motivation for academics in children with autism. *J Autism Dev Disord* 40:1057–1066
  35. Kort W, Schittekatte M, Dekker PH, Verhaeghe P, Compaan EL, Bosmans M, Vermeir G (2005) *Wechsler Intelligence Scale for Children—third edition—Nederlandse bewerking (WISC-III-NL)*. Handleiding en verantwoording. Harcourt Test, Amsterdam
  36. Kozima H, Nakagawa C (2006) Interactive robots as facilitators of children’s social development. In: Lazinica A (ed) *Mobile robots: towards new applications*. I-Tech, Vienna, pp 169–186
  37. Ladd DR (2009) *Intonational phonology*, 2nd edn. Cambridge University Press, Cambridge
  38. Lourens T, Barakova EI (2011) User-friendly robot environment for creation of social scenarios. In: Ferrández JM, Álvarez JR, de la Paz F, Toledo FJ (eds) *Foundations on natural and artificial computation*. Springer, Berlin, pp 212–221
  39. Macintosh K, Dissanayake C (2006) Social skills and problem behaviours in school aged children with high functioning autism and Asperger’s disorder. *J Autism Dev Disord* 36(8):1065–1076
  40. McCann J, Peppé S (2003) Prosody in autism spectrum disorders: a critical review. *Int J Lang Commun Dis* 38(4):325–350
  41. McDonald, J. (2016). A spreadsheet to do the Benjamin-Hochberg procedure on up to 1000 *P* values Retrieved from <http://www.biostathandbook.com/multiplecomparisons.html>
  42. Nalin, M., Bergamini, L., Giusti, A., Baroni, I., & Sanna, A. (2011, March). Children’s perception of a robotic companion in a mildly constrained setting. In *proceedings of the human-robot interaction conference* (pp. 26–29)
  43. O’Riordan M, Passeti F (2006) Discrimination in autism within different sensory modalities. *J Autism Dev Disord* 36:665–675
  44. Peca A, Simut R, Pintea S, Costescu C, Vanderborght B (2014) How do typically developing children and children with autism perceive different social robots? *Comput Hum Behav* 41:268–277
  45. Robins B, Dautenhahn K, Dubowski J (2006) Does appearance matter in the interaction of children with autism with a humanoid robot? *Interact Stud* 7(3):479–512
  46. Russo, N. (2007). *Barking frogs and chirping frogs: a behavioral and brain EEG study of multisensory matching among persons with autism spectrum disorder (doctoral dissertation)*. Retrieved from <http://>

- [digitool.library.mcgill.ca/webclient/StreamGate?folder\\_id=0&dvs=1463492302636-127](http://digitool.library.mcgill.ca/webclient/StreamGate?folder_id=0&dvs=1463492302636-127)
- 47 Sauermann A, Höhle B, Chen A, Järvikivi J (2011) Intonational marking of focus in different word orders in German children. In: Washburn MB, McKinney-Bock K, Varis E, Sawyer A (eds) Proceedings of the 28th west coast conference on formal linguistics. Cascadilla Proceedings Project, Somerville, pp 313–322
  - 48 Simes RJ (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73:751–754
  - 49 Scassellati B (2007) How social robots will help us diagnose, treat, and understand autism. *Robot Res* 28:552–563
  - 50 Scassellati B, Admoni H, Matarić M (2012) Robots for use in autism research. *Annu Rev Biomed Eng* 14:275–294
  - 51 Verschuur R, Didden R, Lang R, Sigafos J, Huskens B (2014) Pivotal response treatment for children with autism spectrum disorders: a systematic review. *Rev J Aut Dev Dis* 1(1):34–61
  - 52 Wong C, Odom SL, Hume K, Cox AW, Fettig A, Kucharczyk S et al (2013) Evidence-based practices for children, youth, and young adults with autism spectrum disorder. The University of North Carolina, Frank Porter Graham Child Development Institute, Autism Evidence-Based Practice Review Group, Chapel Hill