Using robot animation to promote gestural skills in children with autism spectrum disorders

W.-C. So,* M.K.-Y. Wong,* J.-J. Cabibihan,† C.K.-Y. Lam,* R.Y.-Y. Chan‡ & H.-H. Qian§

*Department of Educational Psychology, The Chinese University of Hong Kong, Hong Kong, China
†Department of Mechanical and Industrial Engineering, Qatar University, Doha, Qatar
‡Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, China
§The Chinese University of Hong Kong, Shenzhen, China

Abstract

School-aged children with autism spectrum disorders (ASDs) have delayed gestural development, in comparison with age-matched typically developing children. In this study, an intervention program taught children with low-functioning ASD gestural comprehension and production using video modelling (VM) by a computer-generated robot animation. Six to 12-year-old children with ASD (N = 20; IQ < 70) were taught to recognize 20 gestures produced by the robot animation (phase I), to imitate these gestures (phase II) and to produce them in appropriate social contexts (phase III). Across the three phases, significant differences were found between the results of the pretest and the immediate and follow-up posttests; the results of both posttests were comparable, after controlling for the children’s motor and visual memory skills. The children generalized their acquired gestural skills to a novel setting with a human researcher. These results suggest that VM by a robot animation is effective in teaching children with low-functioning ASD to recognize and produce gestures.

Keywords animated robot, autism spectrum disorders, gesture, video modelling.

Autism spectrum disorders (ASDs) affect as many as 1 in 68 children in USA (Centers for Disease Control and Prevention, 2014). Children with ASD are characterized by impairments in communication and social interaction (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; American Psychiatric Association, 2013). It is well known that children with ASD have difficulties with nonverbal communication (e.g. Asperger, 1944; Bartak, Rutter, & Cox, 1975; Wetherby & Prutting, 1984). They have difficulties in recognizing and producing gestures. The present study was aimed at promoting their gestural communication skills.

When we talk, we gesture. Gestures are spontaneous hand movements produced while talking (McNeill, 1992, 2005). Early studies reported reduced gesture rates in children with ASD (Bartak, Rutter, & Cox, 1975; Wetherby & Prutting, 1984). Recent studies have also found gesture delay in such children (Charman et al., 2003; Luyster, Lopez, & Lord, 2007). Previous findings have shown that, overall, young children with ASD gesture less often than both typically developing (TD) and developmentally delayed children (e.g. Bono, Daley, & Sigman, 2004; Mastrogiuseppe et al., 2015). Regarding different types of gestures, researchers agree that children with ASD exhibit delay in producing protodeclarative gestures (gestures that elicit joint attention and shared interests – e.g. a child points to a toy car in order to direct his mother’s attention to it; Baron-Cohen, 1989; Carpenter, Pennington, & Rogers, 2002).

However, most autism research on gestural production has been conducted among children under 6 years old. Very little is known about the ability to produce and comprehend different types of gestures among school-aged children. Two recent studies by So et al. showed that gesture deficit is found even among school-aged children (So et al., 2015a; So et al., 2015b). Compared

Accepted: 26 June 2016
Correspondence: Wing-Chee So, Department of Educational Psychology, The Chinese University of Hong Kong, Hong Kong, China. Email: wingchee@cuhk.edu.hk

with their TD counterparts, 6 to 12-year-old children with ASD gesture less often and use fewer types of gestures, especially conventional interactive gestures (So et al., 2015b). They also have difficulties in producing iconic gestures at specific locations to identify referents (So et al., 2015a). Children with low-functioning ASD (IQs < 70) have more severe impairments in social interactions and communication (both verbal and nonverbal) than their high-functioning peers (IQs ≥ 70). Therefore, these children encounter substantial problems with basic communication, such as partaking in social interactions with others (e.g. Curcio & Paccia, 1987; Wetherby, Prizant, & Hutchinson, 1998). In light of this, the purpose of this study was to provide an intervention for school-aged children with ASD (especially ones with low-functioning ASD) in order to promote their gestural communication skills. At least 30% of school-aged children with low-functioning ASD are not able to use spoken words to communicate with others (Volkmar & Wiesner, 2009). Teaching them to recognize and produce gestures would significantly improve their communication skills.

There has been little research on intervention techniques designed to teach children with ASD the usage of various types of gestures. In one of the few related studies,Buffington et al. (1998) taught four 4 to 6-year-old children with ASD nine gestures – which were attention-directing (e.g. point to an object), affective (e.g. shake the head) and descriptive (e.g. indicate something is huge) – using a structured, behavioural approach. In the training trial, the therapist presented a stimulus followed by a modelling of the correct gestural and verbal response. The child was expected to imitate both the gesture and its corresponding verbal response. In the probing trial, the therapist presented a novel stimulus and waited for the child to produce gestural and verbal responses. The results showed that the children were able to produce more appropriate gestures and verbal responses after the treatment sessions. They were also able to generalize the acquired gestural and verbal responses to novel stimuli in a novel setting. In another study, Ingersoll, Lewis, and Kroman (2007) used a naturalistic intervention approach to teach five 2 to 4-year-old children with low-functioning ASD (IQs < 70) gestural imitation within ongoing play interactions. All the children were found to imitate gestures more often after training, in both the treatment and novel settings. The improvement was maintained even 1 month after the training. However, the children were not taught the same number or even the same kind of gestures during training because the presentation of gestures was based on whatever kinds of play activities spontaneously occurred.

More recently, Charlop et al. (2010) used video modelling (VM) to teach three 7 to 11-year-old children with ASD some gestures as well as verbal comments, intonation and facial expressions. In their study, the children observed videotapes of target behaviours relating to specific discriminating stimuli. The videos were tailor-made for each child. In terms of gesture production, all the children produced appropriate gestures in the posttests (none of them were able to do so in the pretests), and they were able to apply the acquired gestures when interacting with an unfamiliar person in a novel setting. However, as in the study by Ingersoll, Lewis, and Kroman (2007), these children were not all taught the same gestures because the choice of target behaviours was determined by the preferences of the individual children. Taken together, previous research has shown that children with ASD can, through behavioural interventions, learn to produce meaningful gestures (Buffington et al., 1998; Charlop et al., 2010; Ingersoll, Lewis, & Kroman, 2007).

However, there are methodological concerns with previous studies. First, the sample sizes were small (ranging from three to five). Therefore, previous findings might be subject to the influence of individual variations in gestural learning (Ingersoll, Lewis, & Kroman, 2007). In addition, some of the previous studies did not control for children’s motor and memory skills, which might have influenced their ability to learn gestures (Buffington et al., 1998; Charlop et al., 2010).

More importantly, it is not certain whether children in these studies understood the meanings of the gestures. The children in these studies imitated the gestures modelled by researchers or familiar people during the treatment sessions. Buffington et al. (1998) and Charlop et al. (2010) measured the number of appropriate gestural responses produced in the posttests and generalization probes; these might provide indirect evidence of the children’s understanding of the gestures (i.e. they understood the meanings of the gestures and for that reason could produce them appropriately). Yet, Ingersoll, Lewis and Kroman (2007) did not evaluate the appropriateness of the gestures produced; rather, they counted only the number of gestures imitated, regardless of whether these gestures were the same. Indeed, none of the three studies tested whether the children could identify the meanings of the gestures in a communicative context.
For social communication and interaction, understanding the meanings of gestures is as important as producing gestures. Actually, we should teach children with ASD gesture meanings even before asking them to imitate the gestures. Imitating meaningful actions increases children’s natural motivation to complete the actions (Ingersoll, 2008). Previous findings have also shown that individuals with ASD imitate meaningful gestures better than nonmeaningful ones (e.g. Cossu et al., 2012; Wild et al., 2012).

In the present study, we designed a new intervention protocol to address these methodological issues. We implemented a multiphase therapeutic intervention program in order to teach children with ASD gestural recognition (phase I), followed by gestural imitation (phase II) and gestural production (phase III) in appropriate social contexts. We also developed a pretest and posttests in each phase – in order to assess children’s gestural recognition and production skills – and a generalization test to examine whether they could generalize the acquired skills to human-to-human interactions. In addition, we enlarged the sample size (to 20 children with ASD) and examined the effect of our gestural intervention on gesture recognition and production after controlling for the children’s visual memory and motor skills.

In the present study, we used VM to teach children with ASD to understand and produce gestures. VM usually involves a child watching videotapes featuring adults, peers or him/herself performing a target behaviour; the child subsequently imitates the behaviour (Charlop-Christy, Le, & Freeman, 2000). The majority of VM studies have focused on teaching children with ASD social and communication skills (Mason et al., 2012) and have found that VM can significantly improve social interactions and initiations (e.g. Buggey, 2012; Boudreau, 2013; Maione & Mirenda, 2006), conversation skills (Charlop & Milstein, 1989; Charlop-Christy, Le, & Freeman, 2000), communication (Buggey et al., 1999) and imitation (Cardon, 2012; Kleeberger & Mirenda, 2010).

There are several variations of VM, such as using self-modelling or using another as the model; VM with another as the model is as effective as VM with self-modelling (Mason, 2013). In the present study, we used an animated figure – specifically, a humanoid robot animation shown on a computer screen – as the model. Recent studies have shown that individuals with ASD are more responsive and respond faster to feedback given by a technological object than a human being (e.g. Pierno et al., 2008). Of the different kinds of technological objects, they prefer robot-like toys to nonrobotic toys and human beings (Daunhahn & Werry, 2004; Robins, Daunhahn, & Dubowski, 2006). In the past decade, social robots have been widely used in therapy for individuals with ASD (Fong, Nourbakhsh, & Daunhahn, 2003; Li, Cabibihan, & Tan, 2011). This is partly because a social robot does not have all the facial features and expressions of human beings; thus, using one avoids sensory overstimulation and distraction in children with ASD. Abundant research has shown that social robots can attract the attention of children with ASD, who treat these robots as social agents (Kozima, Michalowski, & Nakagawa, 2009; Miyamoto et al., 2005). Social robots are also found to arouse the interest of children with ASD, eliciting their positive and productive responses (Ho & Cabibihan, 2012; Scassellati, Admoni, & Mataric, 2012); this, in turn, helps them to develop joint attention behaviours, self-initiated interactions, nonverbal communication skills and an ability to make eye contact (e.g. Ricks & Colton, 2010; Werry et al., 2001).

The VM intervention in the present study lasted for 12 weeks over three phases, with each phase lasting for 4 weeks. We argued that learning the meanings of the gestures produced by the robot animation would facilitate imitation of these gestures and even production of gestures in appropriate social contexts. Ultimately, the gestural communication skills taught by the robot animation would be generalized to human-to-human interactions. Figure 1 summarizes the stages of the acquisition of gestural communication skills. Therefore, all the participating children with low-functioning ASD were trained to recognize 20 gestures modelled by the robot animation in phase I, imitate them in phase II and produce them in appropriate social contexts in phase III. In each phase, there were four training sessions (two per week). The effectiveness of the program was evaluated by using standardized tests in each phase; these were administered before and immediately after the training (the pretest and posttest 1) and then 2 weeks later (posttest 2). All the pretests and posttests employed the robot animation as the model. We also tested whether the children would be able to generalize the acquired gestural skills to a novel setting that involved human-to-human interactions.

We hypothesized that the participating children would be able to recognize more gestures in the immediate posttest in phase I, imitate more gestures in the immediate posttest in phase II and produce appropriate gestures...
more often in specific social contexts in the immediate posttest in phase III. We also hypothesized that the positive learning outcomes would be maintained in the follow-up posttests for all the phases. In addition, we expected that, following the training, the children would be able to produce appropriate gestures when interacting with human beings.

Method

Participants

Twenty Chinese-speaking (Cantonese-speaking) children aged 6 to 12 participated in this study (five were female; mean age 9.12 years, ±1.32 SD; age range 6.94 to 11.73 years). They had been diagnosed with autism or other autistic disorder between the ages of 18 and 60 months ($M=32.40; SD=13.26$) by paediatricians at the Child Assessment Centres for the Department of Health in Hong Kong. All the children were attending Hong Chi Morninghill School, Tsui Lam, Tseung Kwan O; this is a special school in Hong Kong for children diagnosed with ASD and having mild to moderate intellectual disabilities. Their ASD diagnoses were further confirmed by clinical psychologists and paediatricians from the Pamela Youde Child Assessment Centre, Hong Kong, through standard clinical interviews with their parents and on the basis of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (American Psychiatric Association, 2000). All the procedures were...
approved by the institutional review board of the university of the first author, in compliance with the Declaration of Helsinki. We obtained the parents’ informed consent prior to the study. Children also gave their assent to participate in this study.

The children’s IQs were assessed by qualified clinical psychologists from the Pamela Youde Child Assessment Centre. Twelve children had their IQs assessed with the Wechsler Intelligence Scale for Children®, Fourth Edition (Hong Kong; WISC IV-HK); their IQs ranged from 51 to 72 (M = 61.52; SD = 7.87). The other eight children had their IQs assessed with the Stanford–Binet Intelligence Scale (Fourth Edition) because they were not capable of completing the subtests in the WISC IV-Hong Kong; their IQs ranged from 49 to 62 (M = 51.18; SD = 4.42).

Stimuli

Twenty gestures commonly used in daily life were taught in this intervention program. Eight of them were iconic gestures (or pantomimic enactments), and the rest were markers (Table 1). The findings of a study by Cabibihan, So, and Pramanik (2012) showed that these gestures are well recognized by speakers in Chinese society. We used the Choreographe software program (Aldebaran Robotics, Paris, France) to create video clips, each featuring a robot animation producing a gesture (Figure 2; the animated clips can be viewed using the link: http://bit.ly/20gestures). Each clip lasted for 3 to 4 s.

Twenty audio clips were made; each contained a vocal token recorded in Cantonese, conveying the same meaning as the accompanying gesture. For example, a vocal token of ‘awesome’ (‘Hou1ye’) that matched the gesture in the video clip for AWESOME (both hands clapping) was recorded. These clips were used in the first two phases of the intervention program. Sixty audio clips in Cantonese that contained scenarios for appropriate gestures were also made. For instance, a scenario describing the use of the AWESOME gesture (e.g. “I won a competition, so I said, ‘Awesome’”) was recorded. The mean length of each scenario was 7.63 s (ranging from 6 to 11 s). Scenarios were used in phase III. All the audio clips were recorded by a female Cantonese speaker. In order to make the audio clips sound like the speech produced by a robot, robotic effects were added and the speech rate was reduced using an audio editor (Audacity, v.2.1.0, The Audacity Team, US state).

In each of the clips, the gesture and the corresponding speech started at the same time. Thus, the participating children watched the gesture videos while listening to the audio clips. All the video clips were displayed on a 17-in computer screen. In each phase, the clips were

<table>
<thead>
<tr>
<th>Iconic gestures</th>
<th>Meaning</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both palms open; both arms are crooked and placed on top of each other, moving to left and right</td>
<td>BABY&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Both arms crook with both fists at waist level</td>
</tr>
<tr>
<td>Both hands make fists and move up and down alternately</td>
<td>DRIVE&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Both hands clap</td>
</tr>
<tr>
<td>With palms open, both arms extend horizontally and flap</td>
<td>BIRD&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Right hand taps the chest</td>
</tr>
<tr>
<td>Both arms crook and move towards self</td>
<td>HUG&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Head nods</td>
</tr>
<tr>
<td>Right palm crook and move towards self and the left arm extends</td>
<td>COME&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Right arm crooks with right palm moving outward from the chest</td>
</tr>
<tr>
<td>Right and left hands down and both arms take turns to move back and forth</td>
<td>WALK&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Right palm facing self moves up and down in the lower chest</td>
</tr>
<tr>
<td>Right hand moves to the mouth</td>
<td>EATING</td>
<td>Both arms crook and form a cross at the chest level</td>
</tr>
<tr>
<td>Right hand moves up and down at the lower chest level</td>
<td>HUNGRY&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Both hands cover eyes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Markers</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANGRY&lt;sup&gt;a&lt;/sup&gt;</td>
<td>AWESOME&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>MYSELF&lt;sup&gt;a&lt;/sup&gt;</td>
<td>YES&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>WELCOME&lt;sup&gt;a&lt;/sup&gt;</td>
<td>HUNGRY&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>WRONG</td>
<td>ANNOYED</td>
</tr>
<tr>
<td>NOT ALLOWED</td>
<td>WAIT</td>
</tr>
<tr>
<td>GOODBYE</td>
<td>WHERE</td>
</tr>
</tbody>
</table>

Note:
<sup>a</sup>Gestures also examined in the study by Cabibihan et al. (2012).
shown in a randomized order. A researcher recorded the children’s responses (e.g. correct or incorrect answers), and the robot animation would ‘act’ appropriately (e.g. giving feedback), according to the children’s responses.

Procedures

The experiment was conducted in a treatment room at Hong Chi Morninghill School, Tsui Lam, Tseung Kwan O, in Hong Kong. The treatment room was often used by the children for their activities. A camera was placed in front of the child being assessed in order to capture his/her responses and hand movements. The intervention program lasted for 12 weeks over three phases (each phase lasting for 4 weeks). Each phase contained a pretest, four training sessions (with two sessions per week), an immediate posttest and a follow-up posttest after 2 weeks. There was a generalization test after completion of phase III.

For each session, the child was accompanied by a teacher. The assessment and training sessions were administered by a researcher, who was either the assistant or the author. A small reward by way of reinforcement

Figure 2 Iconic Gestures (Left) and Markers (Right) Produced by the Nao Animation
(snacks or access to toys) was offered by the teacher at the end of each pretest/posttest and training session. All the sessions in all the phases were videotaped. Each session lasted for approximately 30 min. Details of the intervention program are provided in the succeeding texts.

Pre-intervention assessment
Before the intervention, the children took neuropsychological tests because our assessments of the use of gestures required visual memory and motor skills, and these might influence their gesture learning. Their visual-motor coordination was assessed by the Beery Visual Motor Integration (VMI) test and the Beery Visual Perceptual (VP) subtest, which required the children to reproduce and match geometric shapes, respectively (Beery & Beery, 2004).

Phase I
The robot animation first greeted the child and then gave the instructions for the pretest. Then, the robot animation asked the child whether he/she understood the instructions. The ‘understand’ and ‘not understand’ buttons were displayed on the screen. If the child indicated that he/she did not understand the instructions, the researcher would click the not understand button and the robot animation would repeat the instructions. The purpose of the phase I pretest was to examine whether the children were able to recognize the gestures produced by the robot animation, which demonstrated the 20 gestures (e.g. CLAPPING HANDS) to the child (one at a time, in a randomized order) and then asked the child to indicate the meaning of the gestures. The robot animation verbally provided three predetermined choices (awesome, ‘myself’ and ‘welcome’) that were also visually displayed in separate buttons on the computer screen. The child was given 10 s to respond, either by pointing to the correct answer on the screen (e.g. awesome) or by verbally responding to the researcher. The robot animation prompted the child if he/she gave no response and gave the child ten more seconds to respond. Upon receiving the child’s response, the researcher clicked the corresponding button displayed on the computer screen and instructed the robot animation to proceed to the next gesture. The pretest lasted for approximately 30 min and was completed after all 20 gestures had been covered. A small reward by way of reinforcement was provided after the pretest.

Then, the children proceeded to the gestural training. There were four training sessions, with two 30-min training sessions per week. In each training session, a child watched the robot animation produce the 20 gestures (e.g. CLAPPING HANDS) – one at a time in a randomized order – while listening to the robot animation saying what the gesture meant (awesome); this meaning was also visually displayed on the computer screen. Each gesture was presented twice. Then, the researcher clicked the ‘next’ button, and the robot animation proceeded to demonstrate the next gesture. The training was complete after all 20 gestures had been covered, and a small reward by way of reinforcement was then given.

Immediately after completing all four training sessions (after approximately 2 weeks), the children took posttest 1, which was identical to the pretest. Two weeks after the training, the children took posttest 2, which was the same as posttest 1.

Phase II
As in phase I, the children had a pretest, four training sessions and two posttests. The procedures in the phase II pretest were similar to those in the phase I pretest, except that the children were now required to imitate the isolated gestures. The robot animation asked the child to demonstrate the gestures, one at a time in a randomized order (e.g. “What is the gesture for ‘awesome’?”). Words that represented the gestures’ meanings (e.g. awesome) were also displayed on the computer screen. The pretest was complete after the child had been asked to demonstrate all 20 gestures. A camera located behind the computer captured each child’s gestures for coding. The researcher judged the accuracy of the gesture production on the spot and clicked either the ‘correct’ or ‘incorrect’ button displayed on the computer screen. After the pretest, the children received training. The robot animation produced each of the 20 gestures twice, simultaneously saying its meaning (e.g. it said awesome while clapping its hands). The gesture’s meaning was also visually displayed on the computer screen. Each time, the child was asked to imitate the gesture. Two posttests that were identical to the pretest were conducted: one immediately after the training and the other 2 weeks later.

Phase III
The procedures in the phase III pretest were similar to those in the phase II pretest, except that the children were
now asked to produce appropriate gestures in social contexts. In the pretest, the robot animation asked the child to demonstrate an appropriate gesture in different scenarios, one at a time in a randomized order (e.g. “I won a competition, so I said ‘Awesome’; what gesture should I produce?’”). Words that represented the gestures’ meanings (e.g. awesome) were displayed on a computer screen. The pretest was complete after each child had been asked to demonstrate the gestures in 20 different scenarios. After the pretest, the children proceeded to training in which the robot animation narrated the same 20 scenarios (each one twice) while producing appropriate gestures (e.g. “I won a competition, so I said ‘Awesome’ [CLAPPING HANDS]”). Each time, the child was asked to imitate the gesture. The training was complete after the 20 scenarios had been presented. The children then took posttest 1 immediately after the training and posttest 2 2 weeks after the training. The procedures in both posttests were the same as those in the pretest, except that the scenarios in the posttests had contents different from those in the pretests (e.g. “I finished my homework and I felt awesome; what gesture should I produce?”). As in phase II, the researcher judged the accuracy of the gesture production and clicked the corresponding button (correct or incorrect).

**Generalization**
The generalization test was conducted after completion of phase III. Its purpose was to examine whether the children could produce appropriate gestures in different scenarios narrated by a human researcher (instead of the robot animation). The scenarios used in the generalization test had different contents from those in the pretest and posttests in phase III. Words that represented the gestures’ meanings (e.g. awesome) were also displayed on a computer screen.

Most of the children were able to pay attention during the training and assessments. A short break was given to children who were inattentive. None of them left the sessions. One child did not complete the posttests in phase III. Two children (including the one absent from the posttests in phase III) did not complete the generalization test because they were out of town.

**Coding and scoring**
We evaluated the children’s visual-motor coordination skills in accordance with the scoring manuals of the VMI and VP (Martin, 2006). The maximum possible scores are 78 and 16 for the VMI and VP respectively. In the pretest and posttests in phase I, we counted the total number of times the children correctly identified the meaning of the gestures; we then calculated the average number. For the corresponding tests in phases II and III and the generalization test, we watched the videos of the children and counted the number of times they produced gestures correctly according to four parameters: use of hand/hands (e.g. placing right/left hand against the head vs. using both hands), hand shape (e.g. open palm vs. curled palm vs. fist), direction of movement (e.g. head nods vs. head shakes; moving hand from left to right vs. moving it up and down) and placement (e.g. hand placed at the head vs. at the chest). See the description of the gestures in Table 1. The following gestures were considered incorrect: using the left hand only to produce the BIRD gesture (reason – incorrect use of hands), opening the palm of both hands when producing the DRIVE gesture (reason – incorrect hand shape with both hands), moving the right hand downward when producing the EAT gesture (reason – incorrect direction of movement) and tapping the head with the right/left hand when producing the MYSELF gesture (reason – incorrect placement). We also examined the proportion of the number of times the children committed errors in each of the four parameters when producing the gestures in both phases II and III.

The inter-observer agreement for evaluating the gesture production was 0.95 ($N=80$; Cohen’s kappa $=0.90$, $p<0.001$) in the phase II pretest, 0.94 ($N=80$; Cohen’s kappa $=0.87$, $p<0.001$) in phase II posttest 1, 0.96 ($N=80$; Cohen’s kappa $=0.91$, $p<0.001$) in phase II posttest 2, 0.93 ($N=80$; Cohen’s kappa $=0.84$, $p<0.001$) in the phase III pretest, 0.95 ($N=80$; Cohen’s kappa $=0.88$, $p<0.001$) in phase III posttest 1, 0.95 ($N=80$; Cohen’s kappa $=0.81$, $p<0.001$) in phase III posttest 2 and 0.93 ($N=72$; Cohen’s kappa $=0.84$, $p<0.001$) in the generalization test.

**Results**
The children’s performance in the pretest and posttests in each phase was correlated to their visual-motor coordination skills. Table 2 shows the correlation between the mean number of times the children answered correctly in the pretests and both posttests in all three phases and their scores in the VMI and VP. The children
Table 2. Correlation Coefficients for the Relationships Between Scores for the Intervention Assessments and Visual Motor Coordination Skills

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Phase I pretest</th>
<th>Phase I posttest 1</th>
<th>Phase I posttest 2</th>
<th>Phase II pretest</th>
<th>Phase II posttest 1</th>
<th>Phase II posttest 2</th>
<th>Phase III pretest</th>
<th>Phase III posttest 1</th>
<th>Phase III posttest 2</th>
<th>VMI</th>
<th>VP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase I pretest</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Phase I posttest 1</td>
<td>0.73***</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Phase I posttest 2</td>
<td>0.63**</td>
<td>0.87***</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Phase II pretest</td>
<td>0.66**</td>
<td>0.70***</td>
<td>0.82***</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Phase II posttest 1</td>
<td>0.61**</td>
<td>0.63**</td>
<td>0.71***</td>
<td>0.74***</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Phase II posttest 2</td>
<td>0.62**</td>
<td>0.58**</td>
<td>0.72***</td>
<td>0.77***</td>
<td>0.95***</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Phase III pretest</td>
<td>0.64**</td>
<td>0.67**</td>
<td>0.74***</td>
<td>0.72***</td>
<td>0.94***</td>
<td>0.93***</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Phase III posttest 1</td>
<td>0.63**</td>
<td>0.62**</td>
<td>0.71**</td>
<td>0.75***</td>
<td>0.94***</td>
<td>0.96***</td>
<td>0.94***</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Phase III posttest 2</td>
<td>0.63**</td>
<td>0.64**</td>
<td>0.67**</td>
<td>0.88***</td>
<td>0.90***</td>
<td>0.88***</td>
<td>0.89***</td>
<td>0.95***</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>VMI</td>
<td>0.49*</td>
<td>0.42</td>
<td>0.45*</td>
<td>0.49*</td>
<td>0.50*</td>
<td>0.55*</td>
<td>0.51*</td>
<td>0.49*</td>
<td>0.53*</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>VP</td>
<td>0.33</td>
<td>0.46*</td>
<td>0.53*</td>
<td>0.56**</td>
<td>0.57**</td>
<td>0.71***</td>
<td>0.57**</td>
<td>0.67**</td>
<td>0.66**</td>
<td>0.69**</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: VMI, Beery Visual Motor Integration Test; VP, Beery Visual Perceptual Subtest.

* p < 0.05.
** p < 0.01.
*** p < 0.001.
with better visual memory and motor skills tended to score higher in most of the pretests and posttests across the three phases.

Figure 3 shows the mean number of times the children answered correctly in each phase and in the generalization test. We first examined whether there was a significant improvement in the children’s gestural learning after the interventions in the three phases. Because we found significant correlations between the children’s performance in the pretest and posttests and their performance in the VMI and VP, we conducted a repeated measures ANCOVA, with the phase (phase I, phase II, phase III) and test (the pretest, posttest 1, posttest 2) as the within-subject variables, the score as the dependent variable and the scores in the VMI and VP as the covariates. We found a significant main effect of the test: $F(2, 32) = 6.02, p < 0.006, \eta^2 = 0.27$. The main effect of the phase was not significant: $F(2, 32) = 1.48, p < 0.24$. The effects of motor skills and visual memory were also not significant: for motor skills, $F(1, 16) = 0.30, p < 0.59$; for visual memory, $F(1, 16) = 3.56, p < 0.08$. None of the interaction effects was significant. A Bonferroni pairwise comparison showed that in posttest 1 and posttest 2, the scores were generally higher than in the pretest ($p < 0.001$). There was no significant difference between the two posttests ($p < 1.00$). These results suggest that the children were more likely to recognize and imitate the gestures and produce them appropriately, after the interventions. More importantly, the learning outcome seemed to be maintained in the follow-up posttest (2 weeks after completion of the training).

Some gestures might be easier to recognize and/or produce than others. Therefore, we looked at the proportions of the participants providing correct answers to each of the gesture item in all three phases (Table 3). We collapsed the data for the pretests and both posttests. On average, above 70% of the participants correctly identified the meanings of the majority of the gestures in phase I (but not AWESOME). However, the participants might have found some gestures difficult to produce in phases II and III. In both these phases, the proportions of the participants who accurately produced the GOODBYE, BIRD, YES and DRIVE gestures were greater than the proportions of the participants who accurately produced the ANNOYED, HUNGRY and WRONG gestures.

In phases II and III, we judged the accuracy of gesture production on the basis of four parameters: use of hand(s), hand shape, movement direction and placement. We also examined the parameter(s) within which the children were likely to commit errors. Figure 4 shows the mean proportion of the number of times an error was found in each of the four parameters in the pretests and posttests in phase II and phase III. Two separate repeated-measures ANOVAs – with the test (pretest, posttest 1, posttest 2) and parameter (use of hand or hands, hand shape, direction of movement, placement) as the within-subject independent variables and the mean proportion of the number of times an error was made as the dependent variables – were conducted for phase II and phase III. In phase II, we found significant main

<table>
<thead>
<tr>
<th>Gesture</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOODBYE</td>
<td>0.85</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>YES</td>
<td>0.88</td>
<td>0.80</td>
<td>0.93</td>
</tr>
<tr>
<td>DRIVE</td>
<td>0.82</td>
<td>0.78</td>
<td>0.86</td>
</tr>
<tr>
<td>BIRD</td>
<td>0.75</td>
<td>0.83</td>
<td>0.84</td>
</tr>
<tr>
<td>WAIT</td>
<td>0.70</td>
<td>0.67</td>
<td>0.81</td>
</tr>
<tr>
<td>BABY</td>
<td>0.82</td>
<td>0.55</td>
<td>0.72</td>
</tr>
<tr>
<td>HUG</td>
<td>0.78</td>
<td>0.60</td>
<td>0.70</td>
</tr>
<tr>
<td>WALK</td>
<td>0.77</td>
<td>0.58</td>
<td>0.68</td>
</tr>
<tr>
<td>EAT</td>
<td>0.83</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>MINE</td>
<td>0.72</td>
<td>0.58</td>
<td>0.72</td>
</tr>
<tr>
<td>WELCOME</td>
<td>0.78</td>
<td>0.42</td>
<td>0.67</td>
</tr>
<tr>
<td>ANGRY</td>
<td>0.72</td>
<td>0.58</td>
<td>0.51</td>
</tr>
<tr>
<td>AWESOME</td>
<td>0.63</td>
<td>0.48</td>
<td>0.54</td>
</tr>
<tr>
<td>NOT ALLOWED</td>
<td>0.83</td>
<td>0.35</td>
<td>0.42</td>
</tr>
<tr>
<td>WHERE</td>
<td>0.73</td>
<td>0.43</td>
<td>0.42</td>
</tr>
<tr>
<td>HELLO</td>
<td>0.75</td>
<td>0.33</td>
<td>0.47</td>
</tr>
<tr>
<td>COME</td>
<td>0.77</td>
<td>0.28</td>
<td>0.49</td>
</tr>
<tr>
<td>ANNOYED</td>
<td>0.73</td>
<td>0.27</td>
<td>0.32</td>
</tr>
<tr>
<td>WRONG</td>
<td>0.70</td>
<td>0.27</td>
<td>0.32</td>
</tr>
<tr>
<td>HUNGRY</td>
<td>0.73</td>
<td>0.23</td>
<td>0.26</td>
</tr>
</tbody>
</table>
effects of the test – $F(2, 38) = 13.36, p < 0.001, \eta^2 = 0.41$ – and the parameter – $F(3, 57) = 17.15, p < 0.001, \eta^2 = 0.47$. The interaction effect was not significant: $F(6, 114) = 1.41, p < 0.22$. A Bonferroni pairwise comparison showed that the proportions of the number of times the children made errors across the four parameters in both the posttests were lower than those in the pretests in phase II ($p < 0.003$). There was no significant difference between the two posttests ($p < 0.55$). However, the proportions of the number of times the children made errors in hand shape and movement were higher than those in the number of hands used and placement ($p < 0.012$). There was no significant difference between the number of errors in hand shape and movement or between the number of errors in the number of hands used and placement ($p < 0.17$).

As in phase II, we found significant main effects of the test – $F(2, 36) = 6.66, p < 0.003, \eta^2 = 0.27$ – and the parameter – $F(3, 54) = 23.40, p < 0.001, \eta^2 = 0.57$. The interaction effect was not significant: $F(6, 108) = 0.76, p < 0.60$. A Bonferroni pairwise comparison showed that the proportions of the number of times the children made errors across the four parameters in both posttests were lower than in the pretests in phase II ($p < 0.03$). There was no significant difference between the two posttests ($p < 0.41$). Similarly, in phase III, the proportions of the number of times the children made errors in hand shape and movement were higher than those in the number of hands used ($p < 0.04$). The proportion of the number of times the children made an error in hand shape was also greater than that in placement ($p < 0.001$). Unlike in phase II, however, the mean proportion of the number of times the children made errors in placement was higher than that in the number of hands used ($p < 0.007$). Overall, hand shape and movement were the two most difficult parameters for the children – followed by the placement and the number of hands used – when producing the gestures in phases II and III.

Finally, we examined whether the children could generalize the acquired gestural skills, taught by the robot animation, to a novel setting in which the scenarios were narrated by a human researcher. We compared the children’s performance in posttest 2 in phase III to that in the generalization test (Figure 3). A paired-sample t-test showed that there was no difference between posttest 2 and the generalization test: $t(17) = 1.21, p < 0.24$. The children performed comparably in both of the posttests and the generalization test, suggesting that they were still able to produce appropriate gestures in a novel setting with the human researcher.

Discussion

We found significant differences between the pretests and both posttests in all three phases, after controlling for visual-motor coordination skills, among 20 children with low-functioning ASD. Specifically, the number of gestures recognized and imitated increased after the interventions in phases I and II, respectively. More importantly, the number of gestures accurately produced in
appropriate social contexts also increased after the intervention in phase III. Moreover, we found that the positive learning outcomes were maintained 2 weeks after the training, in all three phases. Even more promisingly, the children were able to generalize their acquired gestural skills to a novel setting with human beings. These results suggest that our VM of a robot animation is effective in teaching children with low-functioning ASD both gesture recognition and gesture production. 

Children and adolescents with ASD learn best through visual means (e.g. Hodgdon, 1995; Mesibov & Shea, 2008). They are often motivated to attend to videos and thus are more likely to imitate the model provided by them (Charlop-Christy, Le, & Freeman, 2000). Previous research has shown that the use of VM (with peer-modelling, adult-modelling or self-modelling) is an effective intervention modality for individuals with ASD (see reviews in Delano, 2007) as it takes advantage of the visual strengths of children with ASD (McCoy & Hermansen, 2007), and it helps these children to focus on the relevant information (Charlop-Christy, & Daneshvar, 2003). Most previous studies on VM have focused on training in social and communicative skills (Mason et al., 2012); only one of them has looked at the effect of VM on gestural learning (Charlop et al., 2010). However, Charlop et al. trained only three children with ASD and did not control for the children’s motor and visual memory skills that might influence gestural learning. Our findings show that visual memory and motor skills are significantly correlated to gestural recognition and production. Therefore, they might have a potential influence on gestural learning generally. In addition, Charlop et al. did not test whether the children understood the meanings of gestures. Therefore, the present study improves the quality of research on VM interventions for gestural learning by providing multiphase training in gestural recognition as well as production and by conducting training with a larger number of participants while controlling for their motor and visual memory abilities.

Our VM intervention started with gestural recognition (phase I), which is considered to be the first step in gestural learning. Children with ASD should learn to identify the meanings of gestures before learning how to produce these gestures. The children participating in our study were able to recognize more gestures after the phase I intervention. After completing this intervention, the children proceeded to phase II, in which they received training in gestural imitation. Previous studies have found that imitating meaningful gestures is challenging for children with ASD (e.g. Stieglitz Ham et al., 2011; Smith & Bryson, 2007). One possible explanation is that, in previous studies, the gestures were produced by human beings. Children with ASD show low interest in interacting with other human beings (e.g. Klin & Jones, 2006), and they find it challenging to pay attention to another’s facial expressions and nonverbal cues (Koegel et al., 1999), which results in difficulty in imitating gestures produced by human beings. However, a social robot does not have or display all the features a human being has or displays, such as facial features and expressions, and thus, it avoids sensory overstimulation of, and distraction in, children with ASD. Our results suggest that a robot (either a real robot or a robot animation) may serve as a good candidate for teaching gestural skills to children with ASD. We found that most of the children with low-functioning ASD paid attention to the VM during the assessments and training sessions. They were also able to imitate more than half of the gestures produced by the robot animation in the posttest in phase II.

The final phase (phase III) taught the children gesture production in appropriate social contexts. Imitating gestures or producing them in isolation is helpful but not sufficient for effective communication in daily life. It is important for children with ASD to learn the circumstances in which suitable gestures should be produced. Producing gestures in appropriate social contexts (e.g. CLAPPING HANDS when praising others’ good work) is crucial for everyday communication. The children who participated in our study were found to be able to produce gestures in appropriate social contexts in over 60% of their responses in the posttest following the intervention in phase III. It is noteworthy that the contents of the scenarios presented in both posttests were different from those in the pretest and training, suggesting that the children with ASD were able to generalize the acquired gestural skills to new social contexts. In addition, the children were also able to produce appropriate gestures in different scenarios that were narrated by the human researcher; this suggests that, when interacting with human beings, they were able to apply the gestural knowledge taught by the robot animation.

That having been said, some gestures were more difficult than others to produce accurately. For example, the scores for the GOODBYE, BIRD, YES and DRIVE
gestures were higher than those for the ANNOYED, HUNGRY and WRONG gestures, in both phases. Like the BIRD and DRIVE gestures, the HUNGRY gesture is an iconic gesture. However, both the BIRD and DRIVE gestures mimic behaviour seen in daily life; they mime the actions of flying and driving a car. The HUNGRY gesture, on the contrary, indicates a physiological need to eat food. Indeed, our findings show that the scores for the EAT gesture, which mimics the action of taking food, were higher than for the HUNGRY gesture. These results probably indicate that children with ASD find it easier to learn gestures that mimic everyday behaviour than to learn gestures that represent physiological or psychological states. However, further research should be conducted to address this issue, as most of the iconic gestures tested in the present study involved mimicry of everyday behaviour.

The children with ASD also found it harder to produce the ANNOYED and WRONG gestures than the GOODBYE and YES gestures. All four of these gestures are markers, but the GOODBYE and YES gestures are less complicated than the ANNOYED and WRONG gestures. The GOODBYE gesture requires only one hand (the left or right hand waves). The YES gesture does not even involve movement of the hand(s); rather, it involves head movement only (the head nods). Unlike the GOODBYE and YES gestures, the ANNOYED and WRONG gestures require both hands to be located at specific parts of the body (the forehead for the ANNOYED gesture and the chest for the WRONG gesture) with correct hand shapes (with closed fists for the ANNOYED gesture as well as for the WRONG gesture).

Regarding the variations in accuracy in producing the gestures, our findings show that the children with ASD made more errors in hand shapes and direction of hand movements than in the number of hand(s) used and their placement. For example, a child had his palms open, instead of joining his fists, when producing the WRONG gesture. Another child just patted her stomach, instead of moving her hand up and down her lower chest, when producing the HUNGRY gesture. Both gestures were produced at the correct locations but with incorrect hand shapes and direction of movement.

This intervention study presents encouraging results showing that VM by a robot animation is useful for promoting gestural recognition and production in a large group of children with low-functioning ASD (after controlling for the children’s motor and visual memory skills). It is also useful for the generalization of their acquired gestural skills to human-to-human interactions. Our intervention protocol can be recommended for clinicians or teachers in special schools to teach children with low-functioning ASD gestural communication skills. Future research should assess whether the positive learning outcomes generated by the intervention can be maintained for more than 2 weeks and whether the kind of VM by a robot animation described in this paper can have an impact on such children’s behaviour outside an experimental setting (e.g. in conversations with their peers and caregivers).

References


Centers for Disease Control and Prevention (2014). CDC estimates 1 in 68 children has been identified with autism spectrum disorder.


