Learning Together: ASIM0 Developing an Interactive Learning Partnership with Children

Sandra Y. Okita, Teachers College, Columbia University, Victor Ng-Thow-Hing, Honda Research Institute USA, Ravi Sarvadevabhatla, Honda Research Institute USA

Abstract—Humanoid robots consist of biologically inspired features, human-like appearance, and intelligent behavior that naturally elicit social responses. Complex interactions are now possible, where children interact and learn from robots. A pilot study attempted to determine which features in robots led to changes in learning and behavior. Three common learning styles, lecture, cooperative, and self-directed, were implemented into ASIM0 to see if children can learn from robots. General features such as monotone robot-like voice and human-like voice were compared. Thirty-seven children between the ages 4- to 10-years participated in the study. Each child engaged in a table-setting task with ASIM0 that exhibited different learning styles and general features. Children answered questions in relation to a table-setting task with a learning measure. Promissory evidence shows that learning styles and general features matter especially for younger children.

I. INTRODUCTION

Robot technologies are now moving from industry into homes as personal companions capable of complex interactions. Humanoid robots can perform human-like motor tasks like pushing [1] and walking [2]. As more biologically inspired robots replicate human characteristics [3], the future role of robots are examined. Robots as a social companion for adults [4], children [5], elderly [6], and as an assistant for those with special needs [7] are much needed areas for application.

Humanoid robots present an array of interesting design choices when modeling interaction with children. Their bodies are designed to resemble a human in both form and function. Humanoids invite interesting social responses from children. A refined robot such as Honda Motor Corporation’s ASIM0 [2] is sure to reveal important preconceived notions in children, and present interesting behavioral responses. Humanoid robots have the intelligence and capacity as a powerful communication tool and educational learning partner for children. Children differ from adults in many ways, and relatively little research exists on extended interaction between children and robots.

In pursuing the possibility of ASIM0 as a learning partner, the study focuses on three basic interests. The first interest is to investigate learning practices that structure interaction in moderately long, turn-taking scenarios with robots. The second interest is to see how children respond to various behavioral traits in ASIM0, ranging from the type of voice, to the frequency of gestures. The third interest is to identify important perceptual cues and responses for robots to recognize and handle respectively. The set of characteristics found may be potentially useful when incorporating the findings into interactive scenarios. Reinforcing natural communication should help increase the child’s subjective comfort toward robots. For practical use, these findings may contribute to future development in perceptual detection algorithms.

Young children lack in experience, patience, and have a relatively low attention span compared to adults. When interaction with the robot becomes moderately long, scenarios that are both engaging and familiar become important. Research has found that for young children, designers should explore building interaction patterns that yield educational and psychological effects, than developing intelligent, realistic, life-like robots. This is because young children may place little value on realistic appearance, and abstract concepts such as robot intelligence [8].

The study involves children taking part in a table-setting task. The materials used are often familiar objects such as forks and spoons are usually associated with a common script of setting the dinner table. The content can also provide a potential learning opportunity. For example, children may be familiar with forks and plates, but may not have noticed the features or the reasoning behind the design. The three learning styles of lecture, cooperative, and self-directed, are well known protocols applied in the classroom that is familiar script.

II. RELATED WORK

People learn in various forms of computerized instruction. Computerized instruction in the past was more machine-based demanding interaction with no direct social exchange with the human learner. The system is “socially indifferent” because there are no social functions, social interests, or social abilities designed as part of the system. An early example is Sidney Pressey’s Testing and Teaching Machine developed in the 1920’s[17]. The interaction involved a student turning to a machine, and taking a multiple-choice test.

Recently, people have started to learn more from computerized devices that blend properties of people and objects. For example, people can learn from virtual people in
virtual environments. Learning can even occur with life-sized humanoid robots. On the other hand, more researchers are now developing and incorporating social interests and social abilities into machines for human learning.

One is “socially implicit” systems that tacitly draw on social schemas for interaction, but usually do not include a real social presence or metaphor. Anderson, Boyle, and Reiser[16] developed an intelligent tutor called "Cognitive Tutor" which was a computational model, representing student thinking and cognition. Cognitive Tutors draw on the idea of tutors helping and correcting students through the learning process. The tutor is usually a disembodied text on a screen. There is no visual character maximizing the social metaphor of a tutor. Students have no discernable relationship with the computer as they would with a human tutor. A number of cognitive models on human thinking has been successfully implemented for human learning. In many cases, the study of machine learning and development has often involved drawing insight from children’s developmental stages. Arthur Arsenio[9] proposes a learning framework for humanoid robots that was inspired by cognitive development in children. However, many of the systems in this category use a command line for interaction and no visual representation of a tutor character with social responses.

Another is “socially explicit” systems that build on explicit social metaphors of interaction and appearances to invite social interaction. Systems in this category consist of features that maximize social metaphors and social presence so that an effective social interaction can take place. Honda’s ASIMO, not only has human-like movement and appearance, but also includes implicit features from cognitive models that invite social interaction[10].

There are several important reasons why humanoid robots are preferable social partners to learning. One reason is that humanoids have human-like appearance and behavior. Human-like appearance and behavior elicits strong social responses that invite active engagement. Social interaction plays an important role in learning, and has proven to be quite effective in collaborative learning, peer learning[11], reciprocal teaching[12], and behavior modeling[13]. Psychological reasons are that humanoid features and gestures can potentially present a more natural communication interface for people. This is especially important for children and the elderly, who may be unfamiliar with the latest gadgets, or physically impaired to operate traditional mouse and keyboard interfaces. Some practical reasons are the humanoid body form promises versatile navigation around human work and living spaces, such as climbing stairs, workspace accessibility and stepping around or over obstacles on the floor[14]. Personal robots interact with humans at a more individual level, offering a wider range of services in the household as servants, educational partners or social companions.

Kanda et. al [5], conducted a study with Robovie, a humanoid robot in a school setting for 18 days. Robovie helped practice English with Japanese children. A large drop in engagement was found by the 2nd week. The findings suggested the need to improve personal relationships between the students.

On a more general note, developing a new learning partnership between people and robots, is an exciting area of research that overlaps various fields. There is much to be learned about the human partner in relation to the technical partner. This pilot study is 1) a learning effectiveness study that determines if social interactions with robots lead to effective learning, and 2) a causal study that try to determine which properties in robots or in the social interaction lead to changes in learning and behavior.

![Fig. 1. Experimental Setting](image)

III. ASIMO AS A LEARNING PARTNER FOR CHILDREN

In this study, ASIMO will be carrying out one of three Learning Styles (Lecture, Cooperative, and Self-directed) when engaging in a table setting task with a child. Both ASIMO and the child have their own set of utensil cards to set up their respective side of the table. Figure 1 shows the experimental setting between ASIMO and the child. For natural interaction, ASIMO will show low-level micro-behaviors.

The study will compare three different Learning Styles (Lecture, Cooperative and Self-Directed) to see if the interaction pattern with the robot has an effect on the child’s learning and behavior.

A. Lecture Style

ASIMO will take on a teacher-like role that shows and tells the child where to put the table setting items. For each item, ASIMO will show by demonstration, or by telling where to place an item. For example, ASIMO will say, “Please place the napkin here”, and then pause to let the child carry the task out, then respond by “Great! Now place the butter knife here”. ASIMO is showing the child how to set the table. At the end of the task, ASIMO and the child should have the same table-setting.

B. Cooperative Style

ASIMO will act as a classmate to the child and participate in peer learning. ASIMO and the child set each side of the table together, by taking turns deciding where the utensil goes. The child is asked to engage in a copy you, copy me cooperative task. They will take turns copying one another’s card placement. The child is asked to cooperate by placing the same item in the same place as ASIMO. For example, ASIMO will say, “I am going to put the napkin here. Can you please help me by placing your napkin in the same place, but on that side of the table?” In return, ASIMO will also cooperate by placing the same item in the same place as the child. ASIMO responds to the child, “ I see you placed the butter knife there. I will help you by placing the butter knife
in the same place but on this side of the table.” At the end of the task, ASIMO and the child should have the same table-setting.

C. Self-Directed Style

ASIMO and the child share the same environment, but engage in the task separately. This you do, I do interaction is similar to parallel play. Both ASIMO and the child set up their own side of the table one item at a time. For example, ASIMO would say, “I am going to put the napkin here. Where do you want to place your napkin? Please place it anywhere you like.” After the child places the card, ASIMO says, “Great!, I think I’m going to put my card over here.” At the end of the task, ASIMO and the child should arrive at different table-settings.

IV. RESEARCH METHOD AND DESIGN

A. Participants

Thirty-seven children from a local private school between the ages 4- to 10-years-old participated in a one-to-one 20-25 minute interaction with Honda’s humanoid robot ASIMO.

<table>
<thead>
<tr>
<th>Monotone Feature</th>
<th>Human-Like Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture Show and Tell</td>
<td>Robot-like mechanical voice</td>
</tr>
<tr>
<td>Cooperative Copy you, Copy me</td>
<td>Robot-like mechanical voice</td>
</tr>
<tr>
<td>Self-Directed You do, I do</td>
<td>Robot-like mechanical voice</td>
</tr>
</tbody>
</table>

Fig. 2: Experimental Design

B. Design

This study is a 3 x 2 design (Figure 2) looking at the Learning Style (Lecture vs. Cooperative vs. Self-directed), by General Feature (monotone voice-gesture vs. human-like voice-gesture) to see if the different features and learning styles will have an effect on the child’s behavior toward the robot and learning gained from the interaction. Human-like refers to ASIMO having a young child’s voice, and gestures include nodding, pointing, arm movement, and taking one-step forward and back. The monotone refers to ASIMO speaking with a monotone robot-like mechanical voice, and minimal gestures limited to nodding and one-step forward and back with no arm movements.

C. Procedure

Children are randomly assigned to one of 6 conditions. Participants will engage in a one-to-one (robot) interaction with ASIMO. The study involves a table-setting task using card pictures of plates, napkins, forks, and spoons on the child’s side of the table. On ASIMO’s side, the same card images are projected and animated on the table instead of being physically present because of ASIMO’s current restriction in manipulating small objects like forks and spoons. During the interaction, ASIMO will be sharing four facts about each utensil (See Table 1). Researchers will be stationed behind a partition (to monitor) while the child interacts with ASIMO. When the table-setting task is done, the experimenter will conduct a post-test that asks the child questions about each of the utensil items placed on the table.

V. MATERIALS AND MEASURES

ASIMO gives four facts for each picture item placed on the table: name of the item, the purpose, a feature of the item, and how to apply it. Table 1 below gives an example.

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Object from Table Setting: Water Glass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>“This is called a Water Glass”</td>
</tr>
<tr>
<td>Purpose</td>
<td>“The water glass is used for drinking water during dinner”</td>
</tr>
<tr>
<td>Feature</td>
<td>“If you look closely, you will notice that it has a large base so it doesn’t tip over”</td>
</tr>
<tr>
<td>Application</td>
<td>“When you drink from the glass, be sure to hold the water glass by the stem to keep the water cold.”</td>
</tr>
</tbody>
</table>

A. Learning Measures

The post-test will measure learning on how much information the child can recall from the interaction with ASIMO. There were twelve items with four facts for each item. The utensil items included are water glass, breadbasket, salt and pepper, dinner knife, dinner fork, spoon, dinner plate, butter knife, soup spoon, cup and saucer, centerpiece, dessert fork, and napkin.

B. Behavioral Measures

The study attempts to measure how children respond to various behavioral traits in ASIMO. Due to the extensive amount of data, this paper will only cover the initial preliminary findings. The behavioral measures include facial expression, attention, voice, eye gaze direction, and the type of engagement seen with ASIMO. For example, if the child initiates conversation, shows body language such as leans forward, and whether children copy robot’s behavior. The study is also interested in identifying what important perceptual cues and responses should the robot recognize and handle respectively.

VI. EXPERIMENTAL ENVIRONMENT SET UP

A. System Architecture

All interactive task scenarios used the Cognitive Map robot architecture [14]. The architecture allowed specialized modules for perception, decision-making, and a common information blackboard for communication (see Figure 3). This architecture was used to manipulate the general feature, such as human-sounding text-to-speech module with the monotone, mechanical voice. The task matrix was used for robot expression, modules for gesture creation, and speech synthesis. The table projector included the projector module that highlighted the card selection and moved card images onto the table.
B. Wizard of Oz

For the decision-making module, we decided to use a Wizard-of-Oz (WOZ) approach where a human operator can direct the robot’s behavior through a remote console without the child knowing. The children are lead to believe that the robot is autonomous. The challenges in creating robust speech recognition for children, correct detection, and recognition of all cards on the table were the developers realize that there is a need to gather more information on how children behave, around ASIMO before automating the decision making module. This led to the design of the Wizard-of-Oz. Through this pilot investigation, the goal was to discover the range of possible behaviors a child may have. To do so, little restriction should be made on their speech. The WOZ approach circumvents this problem by taking advantage of a real human listener.

The WOZ module allows an XML-based script to be specified that describes sequences of commands for speech utterances, gestures and projector operations. Conditional sequences can be triggered by key presses to provide the operator a variety of different robot responses in reaction to the child’s behavior. Each experimental condition was modeled with its own script. The operator can monitor several camera views simultaneously as well as a virtual model of the robot’s current body configuration (Figure 4).

C. Recording Setup

To capture a complete range of behavior, seven cameras were arranged in the observation room (Figure 5). Windows on one wall allowed parents to observe the experiments. Camera-1 is to capture facial expressions. Camera-2 is a side camera for body posture. Camera-3 is an overhead view of the table and child. Camera-4 features a head-mounted camera. Camera-5 is view from ASIMO. Camera-6 is a high-definition camera providing a wide field of view. Camera-7 is another view of the face from a different angle.

Audio was captured separately using a wide-array receiver microphone and lapel microphones attached on the child. In future versions, the plan is to apply computer vision algorithms to help automate this task as well as making the evaluation of body lean and head gaze more objective.

VII. RESULT AND DISCUSSION

Preliminary analysis of the pilot study showed several promising trends in both learning and behavior. Because this was an exploratory study, several variables were added in spite of the small sample size of N= 37 participants (n= 6-7 participants in each condition). For this reason, the analysis did not involve rigorous statistical analysis.

A. Learning Measure

When comparing the Learning Styles, children who interacted with the cooperative ASIMO learned more than the other conditions (See Figure 6). The Self-Directed group learned the least.

When comparing General Features, there was a trend observed where children learned more when interacting with the human-like voice and index gesture ASIMO than the robot-like voice (monotone voice) minimal gesture ASIMO. A similar pattern was seen across all three Learning Styles.

B. Effect of Age

There was a trend seen by age on accuracy level. See Figure 7. In the figure, “Younger” referred to children between the ages of 4 to 6 years, and “Older” referred to children from 7 to 10 years. Overall, older children had higher accuracy than younger children. The interesting finding was that younger participants in the Cooperative condition scored as high as the older participants. This was not observed in the other two conditions. Possibly the Cooperative interaction with ASIMO engaged the children more, leading to higher accuracy. It was interesting to note that the cooperative condition did not result in a similar improvement in learning performance in the older children. Possibly, the choice of task had an effect. The table-setting task may have been too easy for older children, so they were
able to learn quickly regardless of learning style.

Fig. 6. Effect of Learning Style and General Feature on Percent Accuracy

Fig. 7. Effect of Learning Style by Age Group

Fig. 8. Effect of General Feature by Ages 4- 5- 6- years old on Accuracy

C. General Feature

The General Feature (Human voice and gesture vs. Monotone Robot like voice and limited gesture) also showed a difference where participants scored higher when interacting with Human voice and gesture ASIMO. This was consistent across the different Learning Types (Figure 6). Originally, we felt that the biggest effect of General Feature would be with younger children. Our hypothesis was somewhat correct. Figure 8 shows the accuracy level by ages 4-, 5-, 6- years. You can see that the feature has an effect on 4- and 5- year olds where they did better with the Human-like voice and gesture. As children get older, the difference decreases, but overall, participants do better with the Human voice and gesture. The results from this study do not identify which modality (voice or gesture) had a significant influence in the learning performance. Further experiments to test between these two factors are necessary. Observation during the study showed several cases where the young participants felt uneasy when interacting with ASIMO when hearing the monotone robotic voice. Possibly the younger children found the robot-like mechanical voice unusual and difficult to understand. The intonation of words in the human-like voice may have been more familiar and easier to understand. Consequently, the voice may have been the more important of the two factors, but further experiments are needed.

D. Behavior Measures

The behavior data involves extensive coding and is currently being analyzed. This paper will not cover the entire analysis, but qualitative observations seen repeatedly in the interactions with ASIMO will be shared.

Children seemed to respond both positively and negatively to ASIMO’s specific features. Positive responses include ASIMO’s greetings and encouragement (e.g. good job). Children seem engaged and attentive when ASIMO shows movement with gestures and questions the child about experiences. For example, children may smile in response to ASIMO, or imitate ASIMO’s gestures as if they are trying to communicate. When encountering unnatural pauses in the robot, children attempt to troubleshoot by talking louder, waving their hands in front of ASIMO’s eyes, or initiate conversations. Developing a robot to be aware of these “corrective” behaviors can help to improve its robustness and responsiveness.

Some children seemed to view ASIMO as somewhere in-between an alive being and an active machine. For instance, on encountering ASIMO, one child said, “Have you tested him out before?” Instead of saying “it”, the child uses the “him”, however the question refers to “testing him out” which is awkward if the child believed that ASIMO was a alive living being.

Negative responses include young children’s reaction to the monotone voice as being unfriendly, and hard to understand. Some children commented confusion and awkwardness that ASIMO can talk without a visible or moving mouth. Some robot motions seem to trigger cautious reactions, especially when ASIMO took one-step forward and back (Taking one-step forward when it was ASIMO’s turn, and one step back when it was the child’s turn). The child was told at the beginning of the study that ASIMO will take a step forward, but will not get any closer. However, it seems as though children forget, and may need a verbal warning. Having ASIMO simply state his intention to walk closer, ask for permission, or create slow-moving anticipatory motions may help alleviate these problems. The study revealed some ASIMO behaviors that children considered unpredictable.

Unpredictable actions seemed to invite eager engagement from older children. Many tried to “test” the perceptual capabilities of ASIMO and often tried to troubleshoot by trying to trigger a response from ASIMO (e.g. making faces, hiding under the table, deliberately dropping things on the floor). This pilot study helped to see what kind of attention
seeking behavior children show to get ASIMO’s attention. There is a need to respond appropriately to these outreach actions to raise the expectations and believability of agency in robots. One possible reason why older children are often skeptical of the robot’s abilities may be due to their experience with low-functioning robot toys currently available in the market.

Some reactions during the study implied that ASIMO should pay more attention to certain reactions. Children often look up when they are done with the task, or cannot keep up with ASIMO’s speed. Depending on the context, looking up can be interpreted as a signal of confusion or a desire for help.

Some children start the task before ASIMO finishes giving instruction, which implies they are not listening. Just as with humans, if ASIMO can detect this premature motion, it can ask the child to pay attention, or even suggest the child to pay attention.

We noticed that younger children continue to initiate conversation with ASIMO even if they receive unsuccessful responses. Older children make fewer attempts, and are less receptive. If there is a parent by their side, young children resort to asking their parents for assistance and not ASIMO. To encourage more dialogues, ASIMO can detect and acknowledge the parent’s presence, and get involved with the conversation. For example, “your mother makes a great point, but can I try answering your question too?”

We are currently exploring the video and audio data more thoroughly, looking at body posture, and facial expressions. The comments that children make are also being transcribed and coded. The number of interactions and attention data will also be included in the behavioral assessment.

VIII. FUTURE WORK AND CONCLUSION

Learning and behavior data will be used to better design behavior models for ASIMO, incorporate more awareness, and develop tools to automate analysis of head motion and body movement. Improving ASIMO’s attention system and better synchronization of gesture and speech, should help create a richer interaction experiences for the child.

Our pilot study examined what features can be built into ASIMO to help children learn. Children engaged in a table-setting task with ASIMO that exhibited different learning styles (lecture, cooperative, self-directed) and general features (voice and gestures). Results showed that children did better when the interaction was cooperative, and when ASIMO’s voice and gestures were more human-like. Overall, selection of Learning Style and General Feature mattered, possibly more so for younger children. Behavioral observations suggested future possible interventions and responses for ASIMO to increase continuous engagement with the child.

There is much to learn about the human partner in relation to the technical partner. Sheridan[15] states that “design engineers have to be taught that the object is not design of a thing, but design of a relationship between a human and a thing.” Possibly the next step in this partnership is to figure out how the learner can become part of the system and part of the design. It is important to note, that successful learning partnership between robots and humans do not depend on a single learning mechanism or innovative technology. Instead, it depends on situations that bring together a well-chosen confluence of effective learning resources and the choice of partnership with a technology that can help unfold the potential knowledge of the user.

ACKNOWLEDGMENT

We appreciate the participants taking time for this study. The research was supported by Honda Research Institute, USA.

REFERENCES


