Multimodal Approach to Affective Human-Robot Interaction Design with Children

SANDRA Y. OKITA, Teachers College, Columbia University
VICTOR NG-THOW-HING and RAVI K. SARVADEVABHATLA, Honda Research Institute

Two studies examined the different features of humanoid robots and the influence on children's affective behavior. The first study looked at interaction styles and general features of robots. The second study looked at how the robot's attention influences children's behavior and engagement. Through activities familiar to young children (e.g., table setting, story telling), the first study found that cooperative interaction style elicited more oculesic behavior and social engagement. The second study found that quality of attention, type of attention, and length of interaction influences affective behavior and engagement. In the quality of attention, Wizard-of-Oz (woz) elicited the most affective behavior, but automatic attention worked as well as woz when the interaction was short. The type of attention going from nonverbal to verbal attention increased children's oculesic behavior, utterance, and physiological response. Affective interactions did not seem to depend on a single mechanism, but a well-chosen confluence of technical features.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation/methodology; theory and methods*; H.1.2 [**Models and Principles**]: User/Machine Systems—*Software psychology; human factors*; I.2.9 [**Computing Methodologies**]: Artificial Intelligence—*Robotics*; K.3.1 [**Computers and Education**]: Computer Uses in Education

General Terms: Design, Experimentation, Human Factors

Additional Key Words and Phrases: Human-robot interaction, human-robot communication, affective interaction, young children, social robots, collaboration

ACM Reference Format:
Okita, S. Y., Ng-Thow-Hing, V., and Sarvadevabhatla, R. K. 2011. Multimodal approach to affective human-robot interaction design with children. ACM Trans. Interact. Intell. Syst. 1, 1, Article 5 (October 2011), 29 pages.
DOI = 10.1145/2030365.2030370 http://doi.acm.org/10.1145/2030365.2030370

1. INTRODUCTION

Humanoid robots consist of biologically inspired features [Bar-Cohen and Breazeal 2003] and intelligent behaviors that make complex interactions with humans possible. Humanoid robots present an array of interesting design choices when modeling affective interactions with humans. The human-like appearance of robots naturally elicits a social response, revealing new insights to the role of personal robots. Social relations with robots may be hard to define, but the codependency between human and technology is clear.

In developing models to promote affective interactions, a two-way approach is considered. One approach is from the robot's perspective. Advancement in sensors and

This work is supported by Honda Research Institute, Inc., USA.

Authors' addresses: S. Y. Okita (corresponding author), Department of Mathematics, Science and Technology, Teachers College, Columbia University, New York; email: okita@tc.columbia.edu; V. Ng-Thow-Hing and R. K. Sarvadevabhatla, Honda Research Institute, Inc., USA.

DOI 10.1145/2030365.2030370 http://doi.acm.org/10.1145/2030365.2030370

audio-visual tools help "detect" human behavior, while automation and expressive tools help robots "respond" affectively toward humans. Identifying a delicate balance between "detection" and "response" has improved the overall quality of human-robot interaction. Not only have the standards for engagement increased, but more familiarity with robots makes children's expectations higher, and tolerance lower. As Sheridan [2002] states, "design engineers have to be taught that an object is not design of a thing, but design of a relationship between human and a thing."

Another approach is from the human perspective. Much attention has been given in making machines more responsive and sophisticated. Little research has been done to examine which features elicit affective responses in children. Technology as a tool can examine how humans interpret, understand, and behave around technology. Learning about humans through the use of technology allows human users to become part of the system and part of the design. Models for affective interaction cannot be developed in isolation from one another. Experimental studies help isolate features that contribute to human response and behavior. These approaches involve the background of both cognitive science and Human-Computer Interaction (HCI) to develop technological tools and activities that test theories about human-robot interaction and human behavior.

This research is ongoing, and introduces work from two recent studies. The first study looks at interaction styles (lecture versus cooperative versus self-directed) and general features (human-like versus monotone robot-like voice) in robots. The findings are incorporated into the design of experimental tools and interaction models. The tool and model is then implemented into the robot and tested in the second study. The second study looks at the robot's attention feature, and the influence on affective engagement and behavior. As future work, the article explores physiological sensors as a way to measure the influence of attention on affection.

2. RELATED WORK

Research has examined the future potential of robots as a social companion for adults [Mutlu et al. 2006], children [Okita et al. 2009; Kanda et al. 2004], elderly assistance [Wada et al. 2002], special needs [Robins et al. 2004], and within the home [Bartneck and Forlizzi 2004]. Often these biologically inspired life-sized robots are designed to resemble a human and/or animal in appearance, form, and function. This often elicits a social response tapping into people's prior knowledge and experiences. The challenge has been to understand the reach of these robots for children who often are quite expressive when it comes to affective responses. Evolution has provided some answers on how children make attributions toward artifacts [Gelman and Gottfried 1996], but humanoid robots present a new array of interesting design choices when it comes to modeling affective interactions [Arsenio 2004]. In a study with entertainment robots [Okita and Schwartz 2006], children applied "scripts" of familiar play routines when interacting with robotic animals (e.g., playing house, doctor) through which they engaged in affective social interactions. The study examined how children make animistic intuitions based on the robots' appearance and behaviors. Children observed robotic animals (e.g., Sony's Aibo, Omron's NeCoRo, AIST's Paro) exhibit different behaviors (e.g., find and kicks ball, dances to music, no behavior/off) in response to the environment. One implication of the results was that children bring a syncretic set of beliefs to robots and develop the category "robot" slowly in a piecemeal fashion. Rather than replacing one theory with another, they change discrete beliefs based on facts they may have acquired about technology differing from living things. Children's attributions were only mildly connected to the robot's behavior. For instance, children inferred that robots need food, but couldn't grow with the hard exterior shell. Some felt robots needed a remote control to move, could not be associated with any negative feelings (e.g., become sad, or angry), or be capable of any "bad" behavior. However, some felt that the robot dog could still be "bad" and jump on the couch when no one was looking. The study also found that for older children, improving the realism of the robots did not have much effect on the child's conceptual beliefs. The effects of robot realism seemed to be in the child's beliefs and ideas than in appearance. Enabling the child to activate a familiar schema, through which s/he can sustain a social relationship with the robot, seemed far more important. Children were more likely to share positive emotions and empathy with the robots as part of familiar schema.

3. ROBOTIC FEATURES AND ASSESSING AFFECTIVE INTERACTION IN YOUNG CHILDREN

Using experimental and observational methods, two studies were conducted. In examining what features in robots contribute to affective human-robot interaction, we focused on how the robot's voice, interaction style, and attention influence children's behavior and engagement.

3.1. Young Children's Affective Interactions with Humanoid Robots

Proxemics, facial expressions (i.e., smile), and physical contact (i.e., touch) are often used as measures for affection. However, in these studies, participants were young, and robots were about the same size if not taller than the child. As a safety precaution, affective measures that did not involve physical contact were considered.

As an affective measure, the two studies looked at the child's oculesic behavior (i.e., direct eye contact) for important social and emotional information. Eye contact is a good sign of attentiveness and involvement, and widely recognized as an "invitation to communicate" [Anderson 2008]. Increased eye contact over time is often a sign of affection and attraction, essential in establishing emotional closeness [Ray and Floyd 2006]. During the pilot stage of our studies, we found that children make direct eye contact with the robot when they show interest, seek attention, have questions, want approval, and express emotions (e.g., excitement, boredom). Since nonverbal communication is often interpreted within a broader social context, a series of body movements (or kinesics) and nonverbal behaviors (e.g., oculesics) were looked at as a whole, and coded (e.g., boredom, frustration).

No in-depth conversation analysis was conducted. The difference in cognitive and language development across age groups made comparison difficult. For example, the quality of interaction differed greatly with age. For 4 to 5 year olds, content did not matter, conversation was more about timing and turn taking. As children grew older (6 to 8 year olds) they talked less, but listened and answered more carefully. Older children (9 to 10 year olds) only made remarks that they thought were within the capability of the robot. In general, one can say that the child's conversation was constrained by what the robot could do in response. Rather than focusing on a specific content or conversation topic, the two studies looked at using familiar schemas applicable across a wide age range of children (e.g., table setting, turn taking, story telling). Until robots have the intelligence to flexibly respond to young children's interactive bids, following a well-known schema/script helped set a controlled direction for interaction. For this reason, instead of content analysis, the child's level of engagement was measured in terms of the number of direct responses made toward the robot (including the number of remarks initiated by child), and the total length of the conversation.

3.2. Influence of Specific Features on Child-Robot Interaction

The two studies attempt to see whether specific design features in robots influence children's affective behavior toward robots. The first study looks at two features. One is the general feature of "robot voice" (e.g., human-like voice, monotone robot-like voice). Scherer and Oshinsky [1977] mention how emotional attributions are often conveyed more through sensory associations (e.g., sound, voice, gesture) than verbal descriptions.

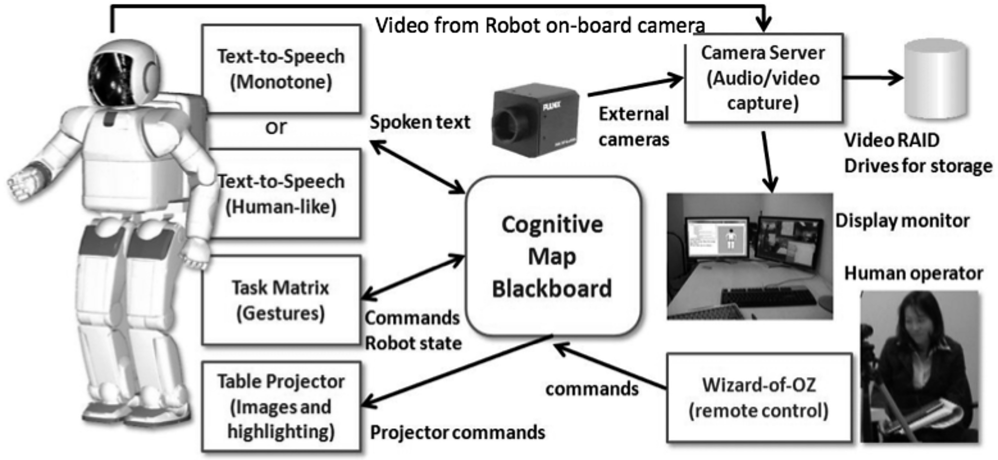

Video from Robot on-board camera
Text-to-Speech (Monotone)
or
Text-to-Speech (Human-like)
Spoken text
External cameras
Camera Server (Audio/video capture)
Video RAID Drives for storage
Display monitor
Human operator
Cognitive Map Blackboard
Task Matrix (Gestures)
Commands Robot state
Table Projector (Images and highlighting)
Projector commands
commands
Wizard-of-OZ (remote control)

Fig. 1. Cognitive map system architecture.

Sound seems to have an immediate and cognitively unmediated effect on emotions that could lead to positive and/or negative associations. The study wants to see if a human child-like voice with natural gestures, or a monotone robot-like voice with minimal and stiff gestures would influence the human-robot interaction. Interaction styles (e.g., lecture style, cooperative, self-directed) are examined to see if different "responses" influence affective behavior.

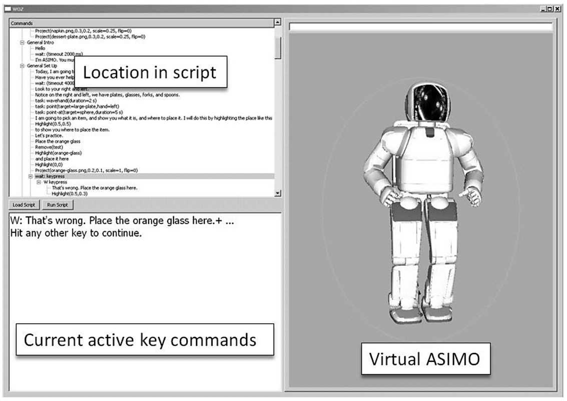
From previous research, we realized that children often show attention seeking behaviors toward unfamiliar situations/objects. This behavior begins to appear when children become bored, disengaged, or detect technical limitations. To entertain themselves, they often "test the tolerance level of the robot" rather than "communicate with the robot." If attention level increases, would children stop testing and start to communicate with the robot? Also, the study was interested in making attention a positive experience (e.g., robot pays attention and listens to child) rather than negative (e.g., testing the robot/child or checking on the child). The second study looks at the quality of attention (e.g., nothing/no attention, automatic attention, and Wizard-of-Oz) to see if different levels of detection influence affective behavior. The study also looks at the type of attention, nonverbal attention (e.g., nodding, looking at) and verbal attention (e.g., "ok", "wow") to see if different responses influence affective behavior.

4. ENVIRONMENT SETUP AND SYSTEM ARCHITECTURE FOR HUMAN-ROBOT INTERACTION STUDY

The interactive task in the first study was implemented using the cognitive map system architecture [Ng-Thow-Hing et al. 2009]. The architecture allows specialized modules for perception, decision-making, and a blackboard for communication (see Figure 1). This architecture was used to manipulate the general feature, such as human-like voice of text-to-speech module with the monotone, mechanical voice. The task matrix used for robot expression consisted of modules for gesture creation and speech synthesis. The table projector module highlighted the card selection and moved card images onto the table.

4.1. Wizard-of-Oz

For the decision-making module, the Wizard-of-Oz (woz) approach was used. Human operator(s) could direct the robot's behavior through a remote console (see Figure 2).



```
      Project(napkin.png,0.3,0.2, scale=0.25, flip=0)
      Project(dessert-plate.png,0.3,0.2, scale=0.25, flip=0)
  ⊟ General Intro
      Hello
      wait: (timeout 2000 ms)
      I'm ASIMO. You mu
  ⊟ General Set Up
      Today, I am going t
      Have you ever help
      wait: (timeout 4000
      Look to your right and left.
      Notice on the right and left, we have plates, glasses, forks, and spoons.
      task: wavehand(duration=2 s)
      task: point(target=large-plate,hand=left)
      task: point-at(target=sphere,duration=5 s)
      I am going to pick an item, and show you what it is, and where to place it. I will do this by highlighting the place like this
      Highlight(0.5,0.5)
      to show you where to place the item.
      Let's practice.
      Place the orange glass
      Remove(test)
      Highlight(orange-glass)
      and place it here
      Highlight(0,0)
      Project(orange-glass.png,0.2,0.1, scale=1, flip=0)
    ⊟ wait: keypress
      ⊟ W keypress
          That's wrong. Place the orange glass here.
          Highlight(0.5,0.3)
```

```
W: That's wrong. Place the orange glass here.+ ...
Hit any other key to continue.
```

Fig. 2. Wizard-of-Oz application.

The children were led to believe that the robot was autonomous. Researchers experienced the challenges in developing robust speech recognition for children, and the accurate detection/recognition of objects on the table. To successfully automate the decision-making module, there was a need to gather information on how children behave around the robots. The woz interface gave that flexibility to examine the range of behaviors and engage in free conversation.

The woz module allowed an XML-based script to be specified (e.g., describing sequences of commands for speech utterances, gestures, and projector operations). Conditional sequences could 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 could monitor several camera views simultaneously as well as a virtual model of the robot's current body configuration.

4.2. Recording Setup

To capture a complete range of behavior, seven cameras were arranged in the observation room (see Figure 3). 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 on the child to record where the children are looking during the interaction [Yoshida and Smith 2008]. Camera-5 is the view from the robot's eyes. Camera-6 and camera-7 are high-definition cameras providing a wide field of view. Audio was captured separately using a wide-array receiver microphone and lapel microphones attached to the child. Parents were able to observe their child through a window looking into the room.

4.3. Physical Safety and Ethical Safeguards

Physical safeguards were placed in the experimental setting to protect children during robot interaction. The table for the card activity acted as a safety barrier between the child and robot in case the robot lost balance. The table was a marker for the child to stay behind the table and not get too close (e.g., looking under the robot's elbow). The robot had a stable balance, and was free standing, but a "hanger" was put in place in case the robot needed support or if the child pushed the robot over. A panic button was in place to stop all processes if needed.

All studies have IRB approval. Participation was voluntary, and participant rights were clearly noted in the parental consent forms. Parents had the right to withdraw their consent or discontinue participation at any time without penalty. The child could

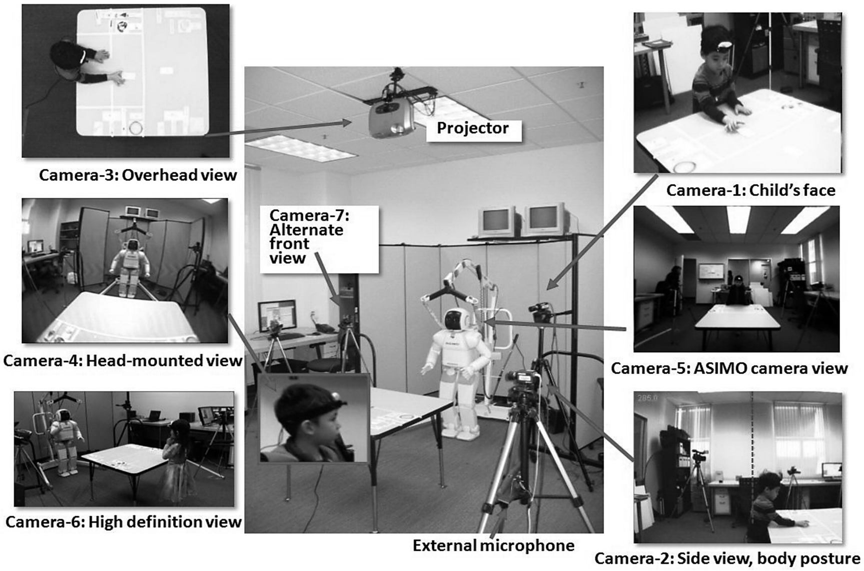

Camera-3: Overhead view
Projector
Camera-1: Child's face
Camera-7: Alternate front view
Camera-4: Head-mounted view
Camera-5: ASIMO camera view
Camera-6: High definition view
External microphone
Camera-2: Side view, body posture

Fig. 3. Camera setup.

discontinue participation at any time, and could refuse to answer any questions. Children were asked on several occasions if they would like to continue, quit, or take a break. Parents also had the option to monitor the child in real time from a distance (e.g., through a glass window, or through a video monitor). The video recordings of the human-robot interaction primarily served as data (e.g., to observe oculesic behavior). Consent was obtained before recording, only research staff would have access to the data, and analysis required no personal identifiers. Findings from the study will only be used for educational research purposes (e.g., conferences, research meetings, journal publications, and articles).

5. INFLUENCE OF ROBOT FEATURES AND INTERACTION STYLE ON AFFECTIVE ENGAGEMENT

The work on entertainment animals [Okita and Schwartz 2006] inspired many design choices for this study with humanoid robots. The preferable age group seemed to be between 4 to 7 years as these children are quite expressive with their feelings (e.g., bored, happy, scared). Robot appearance and behavior may not have shown any difference with robotic animals, but life-sized humanoid robots (i.e., same height as child) may trigger a new array of affective responses (e.g., excitement, interest, fear). Interaction based on a familiar yet communicative activity (e.g., table setting, story telling) is likely to increase engagement. Implementing a human-like feature that allows the robot to display a range of responses (both verbal and nonverbal) can suggest different communication modalities (e.g., gesture/speech) to children and generate affective behavior and engagement.

The first study investigated the following: (1) how different interaction styles may affect behavior in long, turn-taking activities, (2) how general features (e.g., monotone robot-like voice versus human-like voice) influence children's behavior, and (3) identify important perceptual cues that need to be recognized and handled. Examining these characteristics can be potentially useful when incorporating the findings into the design of affective interaction scenarios.

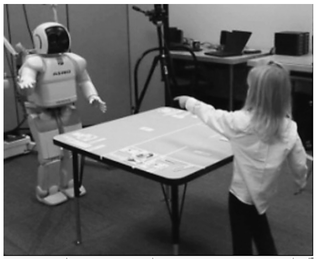


Fig. 4. Interactive environment with robot.

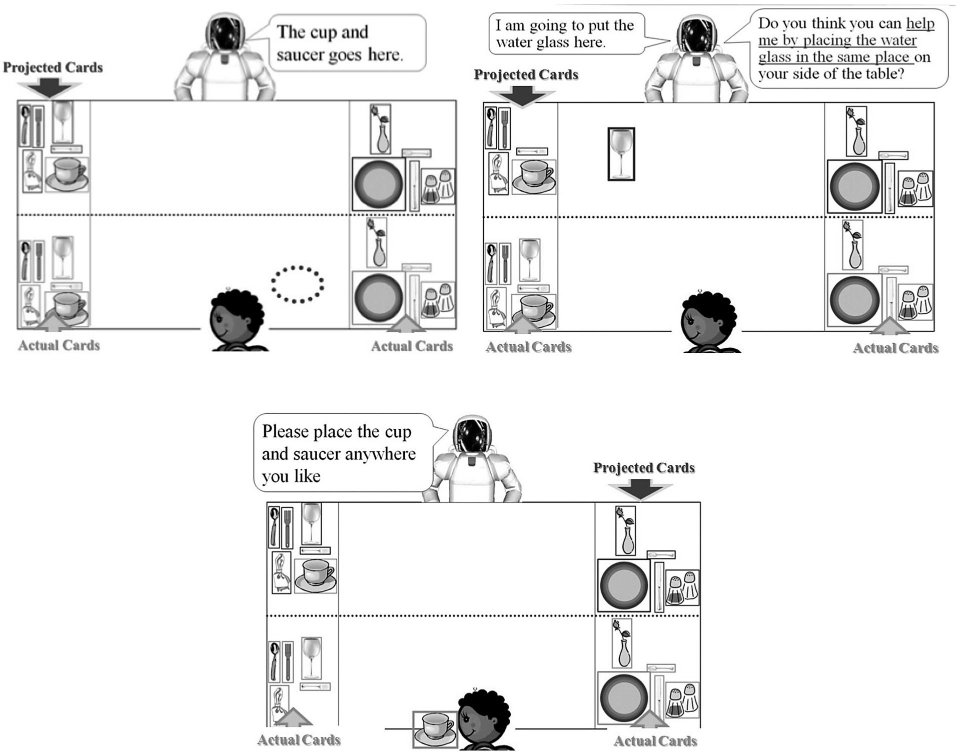

Projected Cards
The cup and saucer goes here.
Actual Cards
Actual Cards
I am going to put the water glass here.
Do you think you can help me by placing the water glass in the same place on your side of the table?
Projected Cards
Actual Cards
Actual Cards
Please place the cup and saucer anywhere you like
Projected Cards
Actual Cards
Actual Cards

Fig. 5. Lecture style (top left), cooperative style (top right), self-directed style (bottom).

5.1. Interaction Style Study

In this study, Honda's humanoid robot carried out one of three interaction styles (lecture, cooperative, and self-directed) in a table setting task with children. Both the humanoid robot and the child had their own set of utensil cards to set up each respective side of the table. Figure 4 shows the experimental setting between the humanoid robot and the child. The study compared three interaction styles and observed if interactive patterns influenc children's affective behavior. See Figure 5.

5.1.1. Lecture Style. A teacher-like humanoid robot lectured to the child on where to put the table setting items. See Figure 5 (top left). For each item, the robot demonstrated with his projected cards, and told the child where to place the item on his/her side. For example, the robot would say, "Please place the napkin here," and then pause to let

Table I. Interaction Style Experimental Study Design

	Monotone Feature	Human-Like Feature
Lecture *Show and Tell*	Robot-like mechanical voice Minimal Gestures	Human-like child voice Index and arm gestures
Cooperative *Copy you, Copy me*	Robot-like mechanical voice Minimal Gestures	Human-like child voice Index and arm gestures
Self-Directed *You do, I do*	Robot-like mechanical voice Minimal Gestures	Human-like child voice Index and arm gestures

the child carry out the task on his side of the table, then respond by "Great! Now place the butter knife here". At the end of the task, the robot and child had the same table setting.

5.1.2. Cooperative Style. A peer-like humanoid robot cooperated with the child in deciding and setting up each side of the table. The robot and child cooperated by taking turns deciding where to put each utensil. See Figure 5 (top right). The child was asked to engage in a (I) *copy you* (if you) *copy me* cooperative task. They would take turns copying one another's card placement. For example, the robot would say, "I am going to put my napkin here. Can you please help me by placing your napkin in the same place, but on your side of the table?" In return, the robot would also cooperate by placing the same item in the same place (as the child) on the robot's side of the table. The robot responded 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, the robot and child had the same table setting.

5.1.3. Self-Directed Style. Humanoid robot and child shared the same environment, but engaged in the task separately. This *you do, I do* interaction was similar to parallel play. See Figure 5 (bottom). Both the robot and the child organized their own side of the table one item at a time. For example, the robot 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, the robot says, "Great! I think I'm going to put my card over here." At the end of the task, the robot and child had different table settings.

5.2. Research Method and Design

An experimental study was conducted where 37 children between the ages 4 to 10 years old engaged in a 20 to 25 minute table setting activity with Honda's humanoid robot [Honda Motor Co. 2000]. This study was a between-subjects 3 × 2 study design where interaction style (lecture versus cooperative versus self-directed) and general features (human-like voice versus monotone robot-like voice) were independent variables (see Table I). Human-like referrs to a computer-generated young child's voice, with natural gestures that included nodding, pointing, and a stepping gesture going forward and back when placing the (projected image) card. Monotone robot-like voice referrs to a monotone mechanical voice, with stiff robot-like gestures of nodding, and stepping forward and back. Children were randomly assigned to one of six conditions. No counterbalancing took place because the study consisted of one common activity, with no within-subject variables.

5.3. Procedure

Participants engaged in a 20 to 25 minute table setting activity with Honda's humanoid robot. The table setting activity (see Figure 4) involved taking turns placing pictures of plates, napkins, and forks on the table. Placing small objects accurately and quickly onto the table can be challenging for robots. For this reason, only the child used physical cards, and the robot used images of cards projected on the table. During the interaction, the robot shared four facts about the utensil (see Table II). Researchers were stationed

Table II. Information for Each Item

Type of facts	Object from Table Setting: Water Glass
Name	"This is called a Water Glass"
Purpose	"The water glass is used for drinking water during dinner"
Feature	"If you look closely, you will notice that it has a large base so it doesn't tip over"
Application	"When you drink from the glass, be sure to hold the water glass by the stem to keep the water cold."

behind a partition and monitored the interaction (see Figure 1, operator). After the table setting task, a post-test was conducted to ask the child about each of the utensils placed on the table.

5.4. Materials and Measures

Since the conversation was constrained by what the robot could do in response, the interaction was structured around a familiar schema (e.g., turn taking, table setting), to help control the direction of the conversation around familiar tasks (e.g., help setting up for dinner) and objects (e.g., fork, spoon). The robot prompted the conversation with the child by sharing both familiar and unfamiliar information about the utensil (e.g., name, purpose, distinct feature of the item, and how to use it). See Table II for examples. The twelve utensils were water glass, breadbasket, salt and pepper, dinner knife, dinner fork, spoon, dinner plate, butter knife, soupspoon, cup and saucer, centerpiece, dessert fork, and napkin.

The behavioral measures included eye contact and the number of times direct comments were made toward the robot. The study was also interested in examining what important perceptual cues and responses the robot needed to recognize and handle respectively. There were three behavior measures in the study: (1) the number of times the child made eye contact with the robot, (2) the number of times the child made comments toward the robot, and (3) qualitative patterns seen in the different affective responses from children.

5.5. Results

Originally 37 participants were recruited, but one participant did not complete the session. Interaction style (lecture versus cooperative versus self-directed) and general features (human-like voice versus monotone robot-like voice) were both added in spite of the small sample size of N = 36 that averaged to about n = 6 participants in each condition. For this reason, the study closely observed whether there were any main effects between the independent variables rather than running rigorous statistical analysis for interaction effects. Focusing on the main effects allowed us to work with a larger sample size to study interaction style (lecture n = 11, cooperative n = 12, self-directed n = 13), and general features (human-like voice n = 19, monotone robot-like voice n = 17).

5.5.1. Eye Contact. Measuring the child's eye contact with the robot was important because eye contact indicated comprehension, uncertainty, interest, and whether the child was trying to communicate with the robot. Running a one-way ANOVA showed that a significant difference was found between interaction styles $F\ (2,\ 33) = 20.36$, MSE = 974.63, $p < 0.001$. Post-hoc Tukey's HSD tests showed that cooperative style was significantly higher when compared to both the lecture style (p < .01, Cohen's d = 1.30, effect-size $r = 0.54$) and self-directed style (p < .001, Cohen's d = 2.76, effect-size $r = 0.81$). The difference between lecture style and self-directed style was approaching significance (p = .056, Cohen's d = 1.09, effect-size r = .048). Younger children (4 to 6 years) made more direct eye contact than older children (7 to 10 years), but the

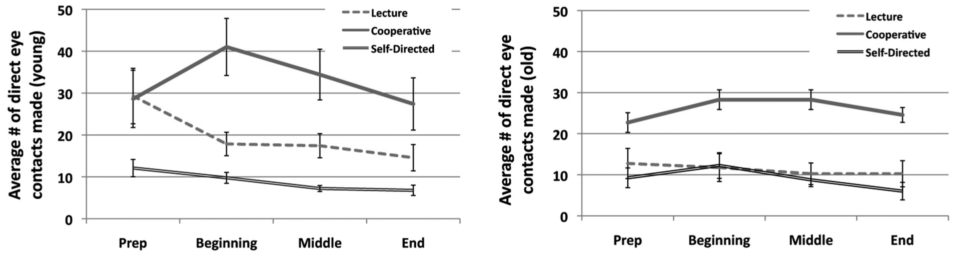


Fig. 6. Average number of direct eye contacts with the robot by interaction style younger children 4 to 6 years old (left), older children 7 to 10 years old (right).

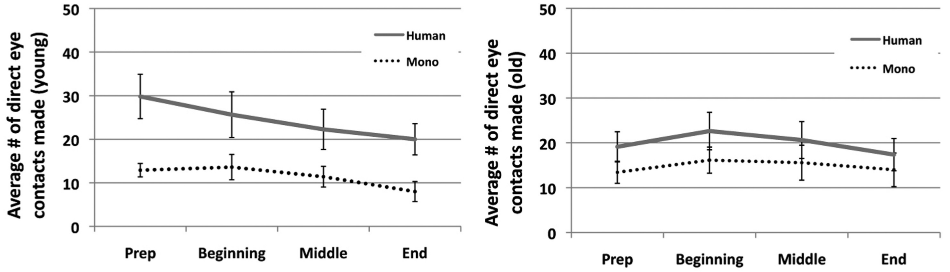


Fig. 7. Average number of direct eye contacts made toward robot by general feature, younger children 4 to 6 years old (left), older children 7 to 10 years old (middle).

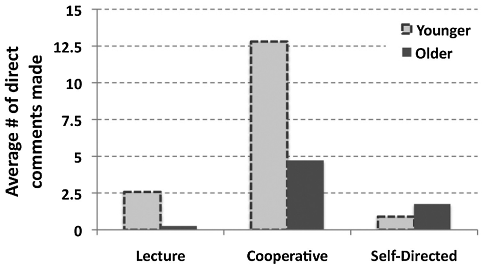


Fig. 8. Average number of direct comments made to robot by interaction style.

patterns were the same (see Figure 6). Over the course of interaction, the number of cases decreased and leveled out. However, in the cooperative interaction style, younger children showed a more gradual decrease. The self-directed style was flat over the course of interaction for both young and older children. For general features (human-like versus monotone robot-like voice), an independent t-test showed a significant difference in the number of direct eye contacts made toward the robot F (1, 34) = 7.87, MSE = 1715.84, $p < .01$, Cohen's $d = 0.95$, effect-size $r = 0.43$). Looking more closely, general features seemed to matter more for younger children than for older children (see Figure 7).

5.5.2. Direct Comments and Utterances. There was no difference in interaction style when it came to the number of times children made direct comments to the robot F (2, 33) = 2.14, MSE = 83.72, $p = 0.134$. However, direct comments were made more in the cooperative style compared to lecture style or self-directed style (see Figure 8). Possibly, the opportunity for the child to contribute her knowledge and choices to the robot in the cooperative style made the interaction more personal and motivating. There was no significant difference in the number of times direct comments were made by age (young: M = 4.29, SD = 11.04, older: M = 2.73, SD = 6.88).

5.5.3. Social Responses. There were three reoccurring social responses that stood out during the interaction: (1) children's expectations and their responses when expectations were not met, (2) different attention seeking behaviors made to the robot, and (3) different quality of interactions by age. The study revealed that the robot behaviors children found unpredictable. Children were repeatedly warned prior to the study that the robot would take a step, but some children would leap back every time. Various impressions were made toward the monotone robot-like voice as "unfriendly," "scary," and "not the right voice for the robot." For the human-like voice, children would say, "he's friendly," "funny and fun," and would ask about the age and gender, "Is he a kid too?" "Is she is the same age as me?" Others showed confusion in that the robot can talk without a visible moving mouth. Unnatural pauses in the robot triggered children to troubleshoot by talking louder, waving their hands in front of the robot's eyes, or repeating the conversation. The comments children made toward the robot seemed to imply that children see the robot as somewhere in-between a live entity and an active machine. For example, a child would say, "Have you tested him out before?" "I'm afraid it's going to want to come over here."

At times, slight delays were seen between a potential trigger and emotion. For example, a child seemed happy interacting with the robot, when suddenly the child expressed frustration and fear. Initially there was no visible incident that could have the triggered such emotion. The reason turned out to be that the child was getting frustrated little by little every time the robot "didn't smile back at me when I smiled." This response showed a need for more emphasis on robot gesture to cover for the lack of facial expressivity, and the need for physiological measures (e.g., skin conductance sensor) to detect subtle emotional build-up.

This pilot study helped to see what kind of attention seeking behavior children show. For example, as the children became familiar with the setting, they started to engage in attention seeking behavior that tested the robot's tolerance level (e.g., making faces, hiding under the table, deliberately dropping things on the floor). Awareness of such "attention seeking" behaviors can improve the robot's robustness and responsiveness. Children's attention seeking behavior was more "testing the robot" than "communicating with the robot". Much work needs to be done to bring the quality of interaction to the next level (e.g., exchanging information, ideas) so a more meaningful relationship may develop through affective interactions.

Another interesting behavior across age was the "quality of interaction." We noticed that younger children continued to initiate conversation with the robot even if they received irrelevant responses. Young children around the age 4 to 5 years carried on multiple conversations with the robot. In other words, what the robot was saying did not matter. As long as there was a response, young children continued to talk and solicit responses from the robot. For example, if the child said, "What is a cup and saucer?" and if the robot replied with, "See how the cup fits perfectly on the saucer?" the child would respond anyway with "I'm not supposed to drink coffee." Older children made fewer attempts, and were less flexible. Children 6 to 7 year olds would listen carefully to the robot, but had very limited responses. For example, if the robot said, "There is an interesting feature about the water glass, do you notice it?" the child will listen carefully, and simply respond with a yes or no. Older children (i.e., between 8 to 10 year old) based their responses on assumptions they had about robots. For example, if the robot asked "have you ever set a table?" The older children would respond to the experimenter with comments like, " Um. . .should I answer?" or "Is it going to understand what I say?” In other words, age seemed to influence the “quality of interaction” where for 4 to 5 year olds response timing was more important than content, 6 to 8 year olds were more concerned with obedience and “doing things right,” while 9 to 10 year olds tried to stay within the repertoire available to the robot.

On several occasions, an interesting observation was that children attempt to communicate with the robot through the operator, or their parents. This showed the importance of having the operator out of sight to lessen the distraction, and to deter the robot from becoming the “third wheel.”

6. MULTIMODAL APPROACH TO GENERATING AFFECTIVE INTERACTIONS

This section explores some important design features in robots that may promote affective interaction. In human-human interactions, both verbal and nonverbal cues are important to fully understand people’s intentions, or grasp the context of the conversation. Similarly, in human-robot interaction, there is a need to explore multimodal communication from the perspectives of both humans and robots.

6.1. Robots Detecting Humans and Objects

The robot software architecture consists of a collection of interconnected, parallel modules, which reports perceptual information to the decision-making module that generates observable behaviors in robots. Important perception modules include face detection to identify the presence of people (e.g., face pose and gaze detection), turn-taking cues, and sound localization for multiple speakers in the room. An active visual attention system is implemented to focus on areas of motion, high intensity, and skin color.

6.2. Robot Response to Humans

The attention system presented important location-based stimuli so that the robot can show natural reactions. An example may be glancing at someone who is speaking, or facing toward the source of a novel sound. Through face detection the robot is able to look at the human during interaction, and engage him in conversation. In places where current perception algorithms cannot be applied, the wizard-of-oz control interface is used to ensure that interaction can proceed successfully [Hauser et al. 2007; Sarvadevabhatla and Ng-Thow-Hing 2009]. The tools are designed so that the user can operate remotely. The operator is out of sight from the participant to lessen distraction. This is important in maintaining the belief that the robot is autonomous. The woz interface allows the operator to directly click on the live video display (a video feed from robot camera) to identify a target for looking and pointing. To manually produce smooth movement in head and hand coordination, more than one operator is needed. Multiple instances can run simultaneously in the woz software, and control tasks can easily be divided between multiple operators (e.g., one operator controls the robot’s verbal conversation, while a second operator controls hand motions and looking behavior). See Figures 9 and 10.

6.3. Gesture Generation

To produce a more natural, lively communication, a multimodal synchronized gesture-speech model [Ng-Thow-Hing et al. 2010] was developed. The model automatically selects gestures to perform, based on the content of speech. Speech is analyzed for word patterns, parts-of-speech, and other semantic tags to produce continuous arm trajectories that are synchronized as the words are spoken. Gestures are especially important for robots without facial expressions, as movements act as emotional cues. Subsequent evaluations of the gesture model show that the addition of gesture modality can produce different emotional impressions and hence shape the tone of interaction

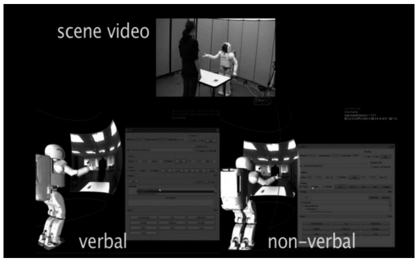


Fig. 9. Woz tool enables multiple operators control over the robot.

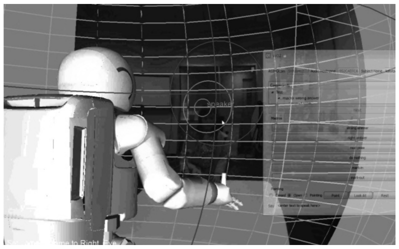


Fig. 10. Woz interface with panoramic attention.

(see Figure 11). Faster motions tend to have positive associations such as excitement and happiness, while slower motions depict sadness and fatigue. Figure 12 illustrates several gesture sequences that are produced automatically by the robot. The continuous expression of gestures with an active attention system helps maintain a smooth flow of communication. Such features are important to avoid the phenomenon of "punctuated stillness" where the robot suddenly stops movement after finishing a sentence, leading people to wonder if the robot is still working.

7. QUALITY OF ATTENTION IN AFFECTIVE ROBOT-HUMAN INTERACTIONS

The first study showed promising interaction styles for engagement, and interesting patterns that elicit affective behavior. More specifically: (1) cooperative peer-like interaction style and human-like features led to more oculesic behavior and social engagement, (2) quality of interaction differs with age, revealing children's preconceived notions about robots, (3) helped identify important behavioral cues and affective responses that need to be implemented into the model, (4) changes to the environmental setup to deter the robot from becoming a "third wheel," and (5) the possible use of physiological measures to detect subtle emotional changes in children. These findings revealed several design implications when automating the decision-making module. For example, children showed several attention seeking behaviors toward the robot, but many involved testing the robot rather than communicating with the robot. The attention system was developed to examine the quality of attention, and the type of attention. Small-size case studies are ideal for this study to closely observe how the different attention features in robots influence children's behavior and engagement.

7.1. Quality of Attention

The second study looked at the quality of attention (e.g., nothing, automatic, and Wizard-of-Oz) to see if different levels of detection influenced affective behavior (see Table III). The study also looked at the type of attention, either nonverbal attention

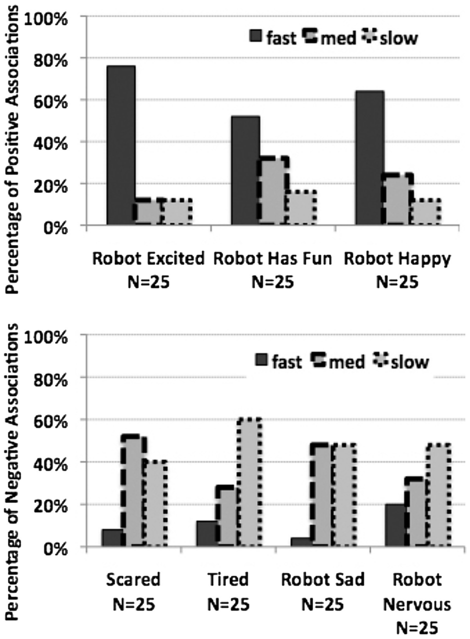
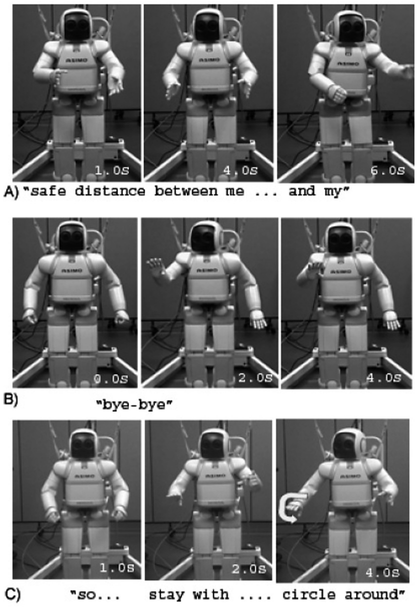


Fig. 11. Positive and negative associations in relation to gesture speed.



A) "safe distance between me ... and my"

B) "bye-bye"

C) "so... stay with .... circle around"

Fig. 12. Examples of gestures generated automatically by robot during speech.

Table III. Study Design on the Quality of Attention

	Non-verbal Attention only (Duck Story)	Non-verbal and Verbal Attention (Pig Story)
Nothing	Basic (pre-scripted comments, gestures, movement) same for both duck and pig	
Automatic	Basic +Non-verbal automatic attention (Look in direction where sound comes, Look, leans in, points when card is shown)	Basic +Non-verbal automatic attention +Verbal automatic attention (Detects when child stops talking. Gives generic feedback e.g., Oh, uh huh, I see, wow)
Woz	Basic +Non-verbal woz attention (Look in direction where sound comes, Look, leans or point when card is shown)	Basic +Non-verbal woz attention +Verbal woz attention (Gives accurate response, and generic feedback when child talks or asks questions)

(e.g., nodding, looking) or verbal attention (e.g., “ok”, “uh huh”, “I see”) to see if different responses influenced affective behavior.

7.1.1. Nothing/No Attention. The robot only displays basic prescripted comments and gestures. There is no active visual attention systems or perceptional information reporting to the decision-making module.

7.1.2. Automatic Attention. In addition to the basic pre-scripted comments and gestures, the visual attention system is active. The robot is autonomous. As the robot interacts with the child, perceptional information is detected and reported to the decision-making module, displaying automated verbal attention responses and/or nonverbal attention responses. Because the detection and responses are automated, the robot’s behavior and timing are not always accurate.

7.1.3. Wizard-of-Oz Attention. In addition to the basic prescripted comments and gestures, the visual attention system is active. However, human operators control the robot remotely. As the robot interacts with the child, human operators detect and respond to the child’s behavior without using the decision-making module. The human operators display accurate verbal attention and/or nonverbal attention responses. The robot’s attention, accuracy, and response rates are higher.

7.2. Research Method and Design

This small-scale case study on the quality of attention was a repeated-measure, randomized crossover study. This 1 × 3 study design consisted of a between-subjects variable on quality of attention (nothing versus automatic versus Wizard-of-Oz), and a within-subjects variable on the type of attention (nonverbal attention, verbal attention). Children were 6 years old, a little older than the first study for longer attention span and higher concentration level. Children were randomly assigned to one of three conditions. Within-subjects measures were used to see how participant behavior evolved as more sophisticated attention features were added-on. For example, in phase 1 of the session, the robot only displayed nonverbal attention. In phase 2 of the same session the robot displayed both nonverbal attention and verbal attention. The add-on approach was taken due to our findings from study 1 where behavior measures decreased over time (even with familiarity and practice). Study 2 examined simple attention features that may increase behavior and engagement. Both studies were lengthy interactions compared to average length in HRI with children. A fair amount of time was setaside for each phase (introduction, phase 1, and phase 2) to see behavior change over time (e.g., after familiarity and practice).

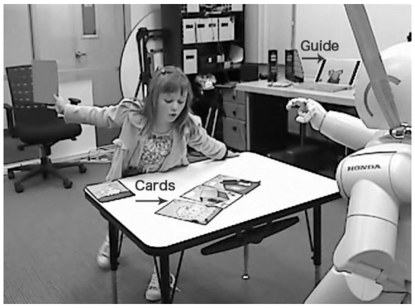


Fig. 13. Wizard-of-Oz condition child describing the story of the Three little pigs and the big bad wolf saying, "I'll huff and I'll puff and I'll blow the house down. . .".

7.3. Procedure

There were 9 children that engaged in a one-to-robot 30 to 35 minute activity with Honda's humanoid robot (e.g., fact exchange, story telling). The cover story was preparing the robot to visit the zoo. In preparing the robot, the child exchanged facts about two animals (e.g., duck and pig). Each animal involved two activities: (1) exchanging information they knew about the animal (e.g., the size, what they eat, how they communicate), and (2) telling the robot a well-known story about the animal (i.e., "Ugly Duckling", "Three Little Pigs and the Big Bad Wolf"). Talking about animals, and story telling were familiar activities for children. Attention needed to be a positive experience (e.g., robot pays attention, and listens to child) than negative (e.g., robot monitoring or checking on the child). To deter the robot from becoming a "third wheel," the human operator(s) were taken out of the picture. Instead, a cartoon character Alyssa appeared on a nearby computer screen to help guide the child. Picture cards of ducks and pigs were set aside so the child could use them to aid her conversation (see Figure 13).

Children were first introduced to Alyssa, a cartoon character on the computer screen. Alyssa gave a cover story and introduced the robot to the child. Each child engaged with the robot for two sessions (the first session on the duck, and the second session on the pig). Each session consisted of two parts (see Table IV). In the first part the child and robot exchanged information (e.g., the size of the duck, what they eat, how they communicate, color). Alyssa guided the interaction by making suggestions such as ("why don't you share with the robot the size of the duck," "why don't you go tell the robot what you think ducks eat"). The interaction was designed so the child could offer facts from her prior knowledge about familiar animals. Alyssa suggested different ways that the child could describe the duck, "You can tell the robot, show the robot with your hands, or use the duck card and explain"). Alyssa also prepared the child on what behavior he could expect to see from the robot ("Remember, when the robot likes what it hears, he may look, move, or ask you some questions"). The robot also provided input about the animal ("Alyssa told me that ducks are brown"), and occasionally asked questions ("Are all ducks brown?"). In the second part of the session the child described a well-known story about the animal to the robot. Alyssa asked the child if she could tell the story ("I think it would really help the robot understand about ducks if you tell him a story. Do you know the story of the "Ugly Duckling"? Do you think you can tell the story to the robot?"). After story telling, the child had the option to take a short bathroom break. Children then repeated the procedure with the second session on the pig. When the two sessions (duck and pig) were completed, the robot and Alyssa thanked the child for coming.

Table IV. Procedure of Attention Study and Questions Asked in Each Question

Procedure		
Introduction		Cover Story (Help prepare robot to go to the zoo) Child introduces self to robot
Session 1: Duck		
Part1	Behavioral Non-verbal Attention Question set1 (BQ1)	Alyssa: Share something you know about a duck? Robot: What are you showing me? Robot: What are you going to tell me about the duck?
	Behavioral Non-verbal Attention Question set2 (BQ2)	Alyssa: Describe the size of the duck to robot Robot: The size of the duck? Robot: What was the color of the duck in the picture?
	Behavioral Non-verbal Attention Question set3 (BQ3)	Alyssa: Tell the robot what ducks like to eat. Robot: What do ducks eat? Robot: How do ducks talk?
Part2	Behavioral Non-verbal Attention Story Telling (BS)	Alyssa: Tell the robot the story of the ugly duckling? Robot: Do you want to be the ugly ducking?
Session 2: Pig		
Part1	Behavioral Verbal Attention Question set1 (BVQ1)	Alyssa: Share something you know about a pig? Robot: What are you showing me? Robot: What are you going to tell me about the pig?
	Behavioral Verbal Attention Question set2 (BVQ2)	Alyssa: Tell how pigs talk to each other? Robot: Is that how pigs talk to each other? Robot: What else do pigs eat?
	Behavioral Verbal Attention Question set3 (BVQ3)	Alyssa: Share the size of the pig? Robot: Is that the size of the pig? Robot: Are pigs always running around?
Part2	Behavioral Verbal Attention Story Telling (BVS)	Alyssa: Tell story of the three little pigs and big bad wolf? Robot: What kind of house would you build?

7.4. Materials and Measures

There were four measures in the quality of attention study: (1) oculesic behavior measuring the number of times the child made direct eye contact with the robot during the activities, (2) engagement, meaning the number of incidents where the child makes direct comments toward the robot and length of communication, (3) children's behavioral response to story-telling activities, and (4) frequent behaviors seen during the interaction, including gestures, actions, and motions made to the robot (directly or indirectly), or others (e.g., parents).

7.5. Results

This small case study consisted of 9 children. The findings will be explained by examining promising trends rather than running rigorous statistical analysis.

7.5.1. Effect of Robotic Features on Children's Eye-Gaze. A trend was seen in the amount of eye contact made across the three conditions (woz versus automatic versus nothing). The woz condition was highest compared to automatic and nothing (see Figure 14). In part 1 (duck session with nonverbal attention BQ1-BQ3), the woz condition showed a stable count while the number of cases decreased for the automatic condition. An increase in direct eye contact was observed as the verbal attention feature was added in part 2 of the session (pig session with verbal attention BVQ1-BVQ3). The nothing condition was somewhat flat over the course of part 1 and 2 of the session. Overall, more eye contact was made in the woz condition, and verbal attention seemed to increase the engagement level from BQ3 to BVQ1 for both woz and automatic (see Figure 14).

7.5.2. Effect of Robotic Features on Children's Engagement. Overall length of conversation was longer for children in the woz condition compared to automatic and nothing (see Figure 15). The difference increased over time with a sudden jump in part 2 when

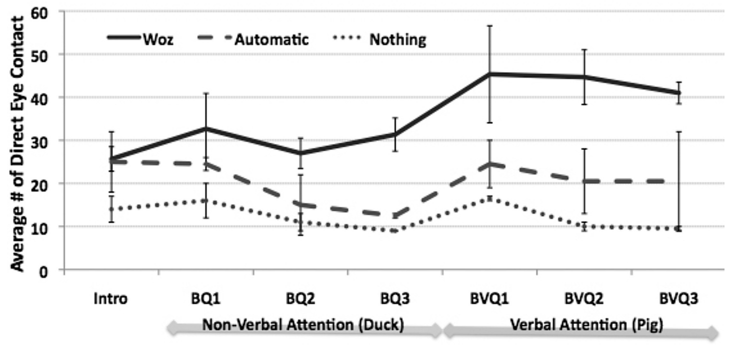


Fig. 14. Average number of direct eye contacts by child to robot.

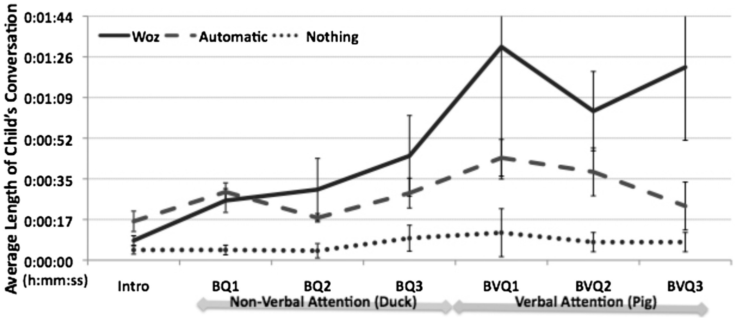


Fig. 15. Average length of child's conversation with robot.

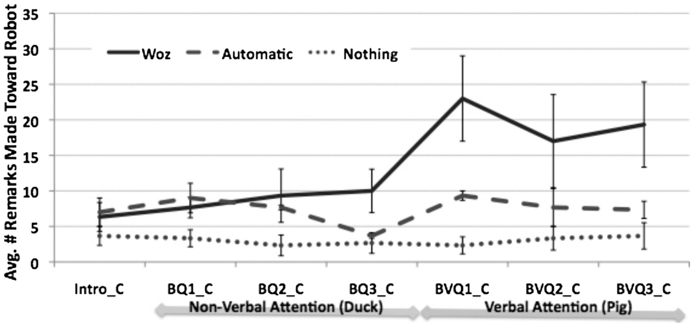


Fig. 16. Average number of times child made direct remarks to robot.

verbal attention was added on (pig session). The automatic condition decreased over time while the nothing condition was low throughout the session.

For the number of times direct remarks were made toward the robot (see Figure 16), part 1 of the session (duck, nonverbal attention) showed a gradual increase for woz, and a decrease for automatic. The number of remarks showed an increase in both woz and automatic as the verbal attention feature was added in part 2. This increase was short, lived as both conditions showed a decrease. Overall, the woz condition was highest in the number of times remarks were made toward the robot.

7.5.3. Effect of Robotic Features on Children's Behavior. In part 1, there was very little difference in the number of behavioral incidents across the three conditions

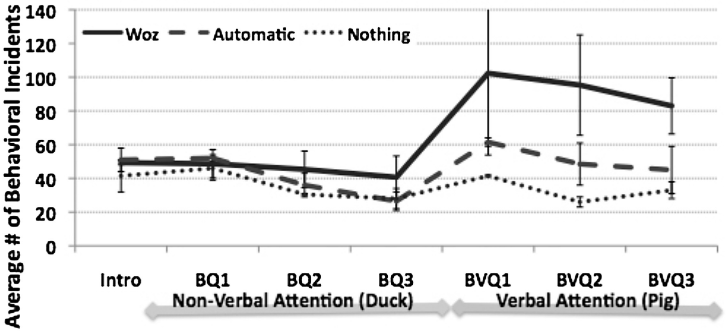


Fig. 17. Average number of behavioral incidents made by child during session.

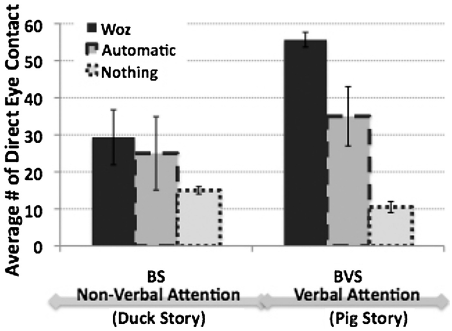


Fig. 18. Number of direct eye contact made by child to robot during story telling.

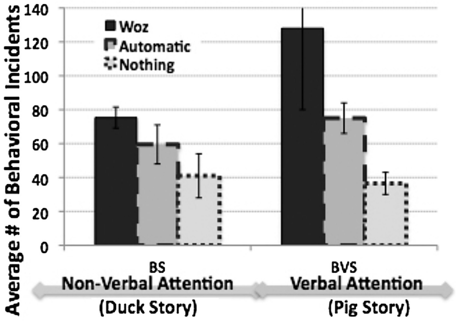


Fig. 19. Number of behavioral incidents made by child during story telling.

(see Figure 17). As verbal attention was added in part 2, the number of incidents increased. The woz and automatic condition showed a sharp increase, and a slow gradual decrease over time. The nothing condition was flat throughout the session.

7.5.4. Children's Response Story-Telling Activities. For the story-telling task the child was asked to tell the story of the "Ugly Duckling" to the robot in part 1 (nonverbal attention) and the story of the "Three Little Pigs and the Big Bad Wolf" in part 2 (verbal attention) of the session. The nature of the task was different compared to the information sharing activity because the child would be taking the role of the storyteller, initiating most of the conversation. There was an increase in incidents from part 1 to part 2 across all four measures: eye contact (see Figure 18), behavioral incidents (see Figure 19),

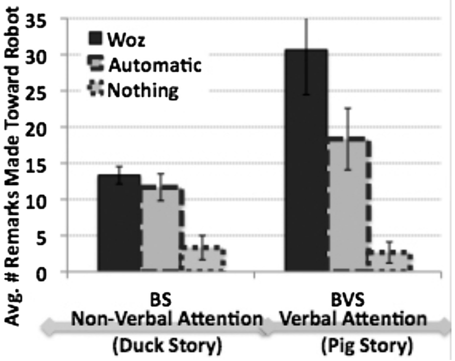


Fig. 20. Average number of remarks made by child to robot during story telling.

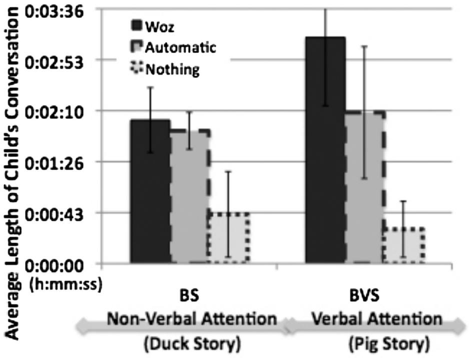


Fig. 21. Average length of child's conversation during story telling.

number of remarks made (see Figure 20), and the average length of conversations (see Figure 21). Overall, woz had the highest number of incidents across the two sessions (part 1 and part 2). The incidents increased for woz and automatic (more so for woz) when verbal attention was added. There was no change (if not a slight decrease) for the nothing condition across all four measures.

7.5.5. Frequent Behaviors Seen During Engagement. In general, there was more direct and indirect behaviors made toward the robot, with very little behavior made toward others (i.e., experimenter and/or parents). By placing Alyssa as a guide, and taking the experimenter/operator out of sight, children talked more directly to the robot. Figure 22 shows: (1) the total number of behavioral incidents across conditions, (2) which number of behavioral incidents were targeted directly at the robot, indirectly to self or robot, or addressed to others, and (3) the most frequent behaviors seen. Table V gives the description/example of each of the codes. The "X" in the figure and table refers to all other nonfrequent behaviors. Incidents targeted directly at the robot mostly involved "checking behaviors (CH)" and "explain (XP)" behaviors. This reflects the task involved in the session "the child checking to see if the robot is listening," "checking to see how robot will respond to the child's information/fact," and explaining to the robot what he/she knows about the duck/pig (e.g., size of the pig)". For incidents made indirectly to self/robot, frequent behaviors involved "Looking Around (LA)", "Following Instruction (FI)", "Full Attention (FA)".

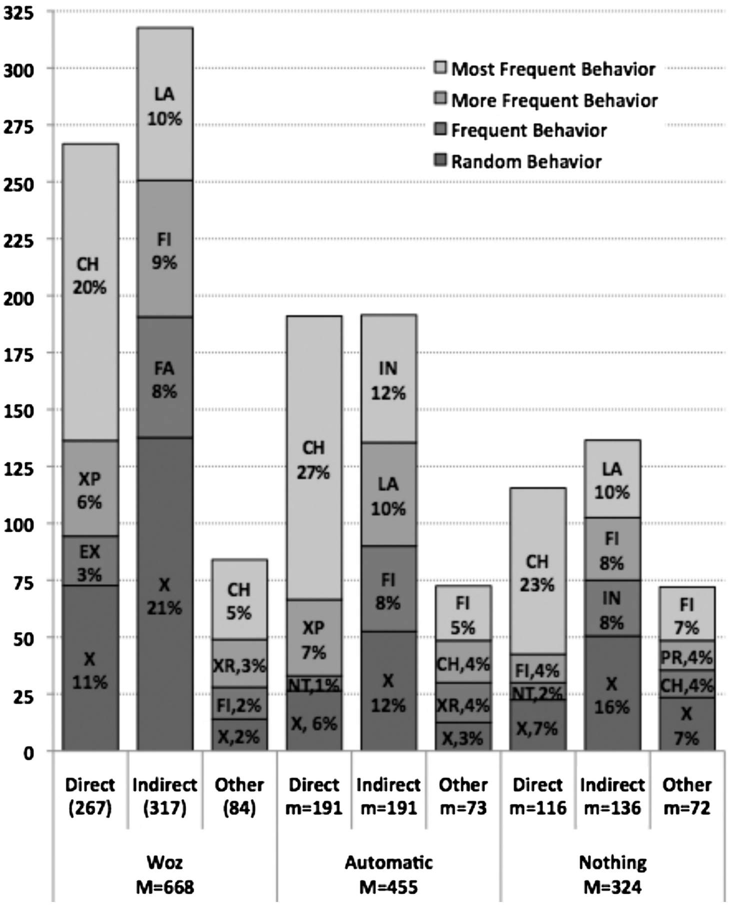
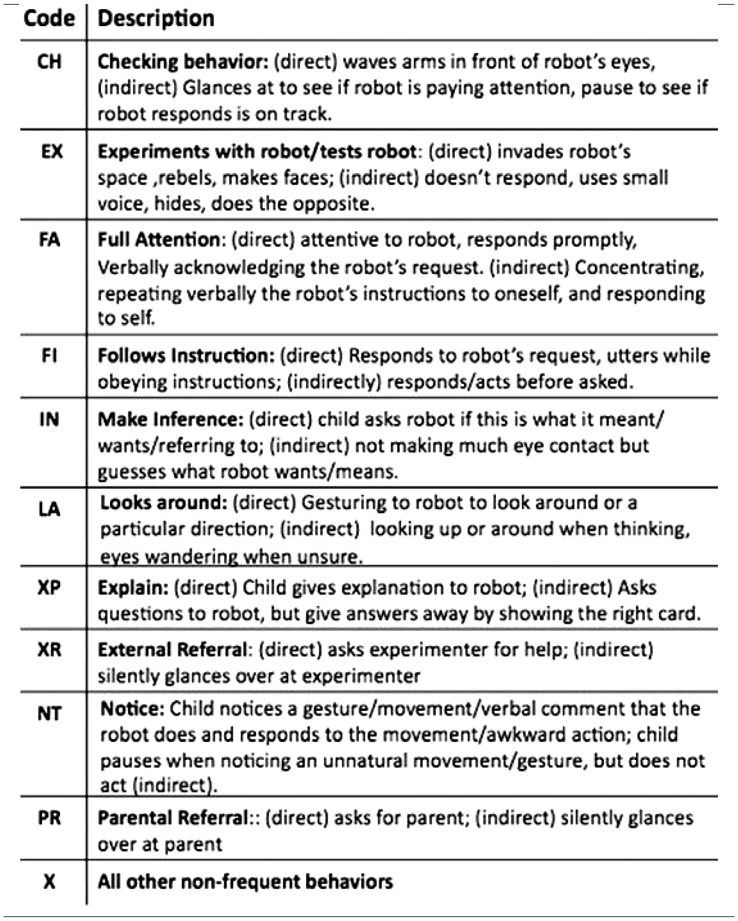


Fig. 22. Most frequent behavior seen by child during the session.

A slight variation was seen across conditions. Looking Around (LA) was most frequent in the woz condition, and Make Inference (IN) the most frequent in the automatic condition. This may be due to the wide range of responses and interpretation that can be made from the robot's gesture/movement in the automatic condition. Many of the indirect response involved the child thinking or self-reflecting following a response from the robot. Often children's eyes wandered when thinking how to respond, or explain. Children also quietly mouthed or repeated what the robot said to think over the information exchanged. There were fewer behavioral incidents made toward "others." Most involved Checking (CH), Following Instructions (FI), External Referrals (ER), or Parental Referrals (PR) where children relied on external approval/confirmation/clarification from parents or experimenters.

Looking further into the data, Figures 23, 24, and 25 show whether the number of behavioral incidents and their targets (direct, indirect, and other) differ over time (or as the session progresses) and across conditions. There was little difference in part 1 of the session (duck, nonverbal attention), but the number of incidents increased in the latter half when part 2 (pig, verbal attention) was added on as a feature. Across all three conditions, woz had the highest number of behavioral incidents. Overall, there were slightly more indirect than direct behavioral incidents. This may be due to the nature

Table V. Description of Frequent Behavior



Code	Description
CH	**Checking behavior:** (direct) waves arms in front of robot's eyes, (indirect) Glances at to see if robot is paying attention, pause to see if robot responds is on track.
EX	**Experiments with robot/tests robot**: (direct) invades robot's space ,rebels, makes faces; (indirect) doesn't respond, uses small voice, hides, does the opposite.
FA	**Full Attention**: (direct) attentive to robot, responds promptly, Verbally acknowledging the robot's request. (indirect) Concentrating, repeating verbally the robot's instructions to oneself, and responding to self.
FI	**Follows Instruction:** (direct) Responds to robot's request, utters while obeying instructions; (indirectly) responds/acts before asked.
IN	**Make Inference:** (direct) child asks robot if this is what it meant/ wants/referring to; (indirect) not making much eye contact but guesses what robot wants/means.
LA	**Looks around:** (direct) Gesturing to robot to look around or a particular direction; (indirect) looking up or around when thinking, eyes wandering when unsure.
XP	**Explain:** (direct) Child gives explanation to robot; (indirect) Asks questions to robot, but give answers away by showing the right card.
XR	**External Referral**: (direct) asks experimenter for help; (indirect) silently glances over at experimenter
NT	**Notice:** Child notices a gesture/movement/verbal comment that the robot does and responds to the movement/awkward action; child pauses when noticing an unnatural movement/gesture, but does not act (indirect).
PR	**Parental Referral**:: (direct) asks for parent; (indirect) silently glances over at parent
X	**All other non-frequent behaviors**

of the task that involved children sharing their own knowledge of a familiar animal, and explaining/teaching facts to the robot. Initially the hypothesis was that "other" referrals would be more frequent in the beginning of the session than later. It turns out that the number of behaviors targeted at others did not differ much throughout the session as well as across conditions.

Most results across the five measures showed that the woz condition elicited more affective behavior and engagement, followed by automatic and nothing. The difference was most apparent when the verbal attention feature was added on to the robot (part 2, BVQ1-BVQ3). In some cases, there was little difference across conditions until this additional feature was added. Nonverbal attention may be difficult for young children to notice. A similar but greater increase was seen for the story-telling activity, when verbal attention was added. However, its important to note that automatic nonverbal attention did similarly well as woz nonverbal attention when the interaction was short. The benefit of woz over automatic was seen over time. A simplified full-scale study would definitely be the next step. A subtracting features study could be run to balance with the large-scale add-on study.

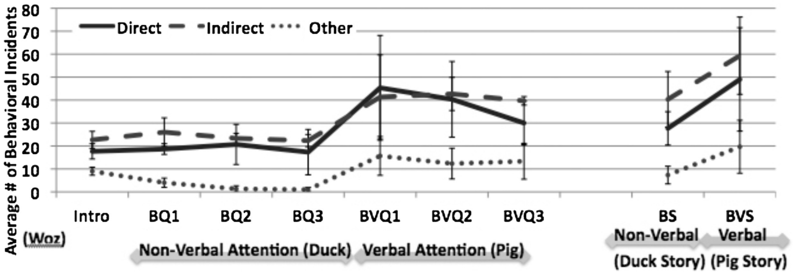


Fig. 23. Average number of behavioral incidents by child in woz condition.

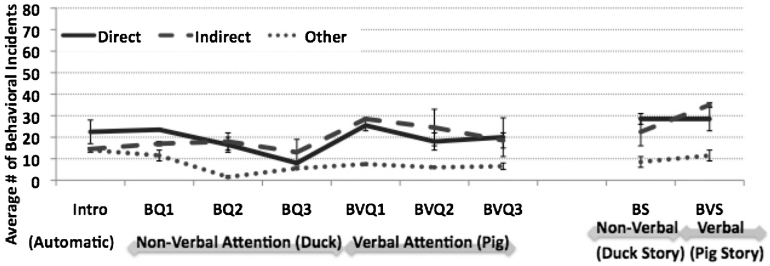


Fig. 24. Average number of behavioral incidents by child in automatic condition.

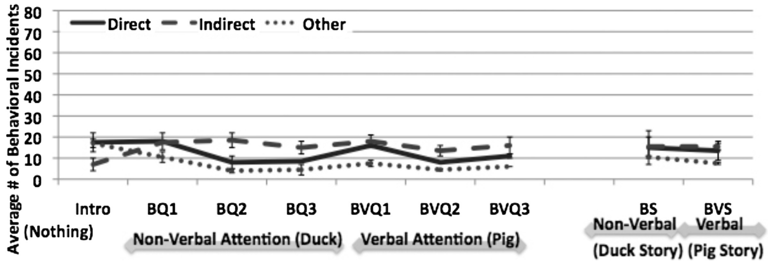
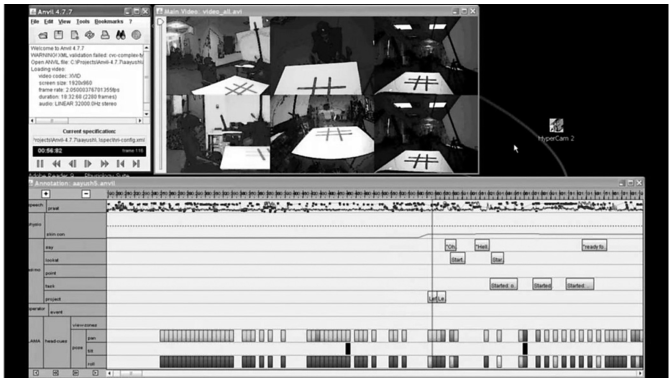


Fig. 25. Average number of behavioral incidents by child in nothing condition.

8. TOOL DEVELOPMENT FOR MULTIMODAL ANALYSIS

The next two sections introduce tools that were developed to automate analysis of head motion and body movement. Recordings of extended-duration interactions produce a large amount of data. Such analysis usually involves manual annotation (coding) of relevant events in the recording. Typical annotations are done frame-by-frame for video and short segments of time for audio. Often several passes over the same data have to be made by multiple coders for inter-rater reliability. The annotation process is usually extremely labor intensive and error prone. If the observer is not explicitly looking at the right time frame or the right point of view, crucial interactions can be missed.

Multiple video streams from different viewpoints were recorded in the two studies for this reason. There was a need to create tools that enable multimodal analysis, eliminate the laborious coding tasks from each camera angle, and a tool that applies computer vision algorithms to automate the detection and documentation of microbehavior occurrences. To make sure that no false positives or false negatives occur, performance comparisons of these automatic methods were needed [Dautenhahn and Werry 2002]. The SAMA (Subject Automated Monitoring and Analysis) system was designed to automate analysis of head motion and body movement. In SAMA, the multiview camera



```
Welcome to Anvil 4.7.7
WARNING! XML validation failed: cvc-complex-t
Open ANVIL file: C:\Projects\Anvil-4.7.7\aayush\
Loading video:
    video codec: XVID
    screen size: 1920x960
    frame rate: 2.05000376701355fps
    duration: 18:32:68 (2280 frames)
    audio: LINEAR 32000.0Hz stereo
```

Current specification:
?rojects\Anvil-4.7.7\aayush\ \spec\hri-config.xml

Fig. 26. Screenshot of Anvil synchronizing multiple viewpoint videos, physiological annotations, and automatically generated SAMA annotations.

data could be processed to obtain head pose and gaze-related annotations. The basic design goal of using SAMA along with Anvil (free video annotation tool for multilayered annotation based on a user-defined coding scheme) was to analyze the sensor data (in particular, camera information) for cues and patterns in the human-robot interaction (see Figure 26). For example, when looking for instances where the subject turned away from the humanoid, or instances where the robot and subject were speaking simultaneously, SAMA analyzes the multiview video data and outputs a semantic annotation tag set for each time slice. For each time instance, it simultaneously processes the corresponding frame from each of the 5 cameras (not including the back-up camera-6 and -7). For each frame, various pose-related face properties such as head-roll, tilt, and pan were estimated. The relative positions of the camera viewpoints (front, profile, side, etc.) were also extracted. The face properties from all the viewpoints along with viewpoint information were combined and processed using a rule base, which determined the final semantic tag set for that time instant. In this way, the entire set of videos associated with an interaction episode can be coded with information on the current gaze location of the subject [Sarvadevabhatla et al. 2010]. The analysis from SAMA provided useful cues of where to focus on in the video through the indication of low-level gaze cues and their transitions. For this reason, Anvil was an ideal environment for analysis. By combining the analysis from SAMA on video data with data from other sensors (e.g., audio, physiological measures), there was an opportunity to examine unobserved long-range relationships between interaction elements. By combining information from multiple viewpoints, SAMA could provide informative tagging that was far more informative than from a single viewpoint.

9. EXPLORATIONS IN PHYSIOLOGICAL MEASUREMENTS IN MULTIMODAL ANALYSIS

Although video and audio analysis of interactions between children and robots could reveal emotional responses, children often found it hard to express their emotions in unfamiliar situations. Some children could choose to hide their emotional displays out of fear or embarrassment. Sometimes direct physiological measurements could help reveal subtle triggers from interaction or indicate a general level of arousal. Sensors were tested to examine the potential for future large-scale usage.

One particular challenge of using skin conductance sensors with children was their tendency to move around during interaction. The more engaging the interaction, the more active the participant was. This ruled out wired skin conductance sensors where electrodes around fingers suffer loss of contact with movement. The wireless Q sensors

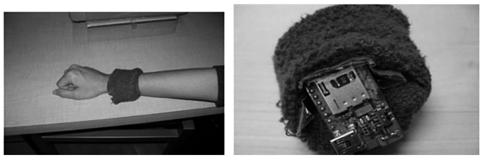


Fig. 27. Affectiva Q sensor for wireless skin conductance measurement.

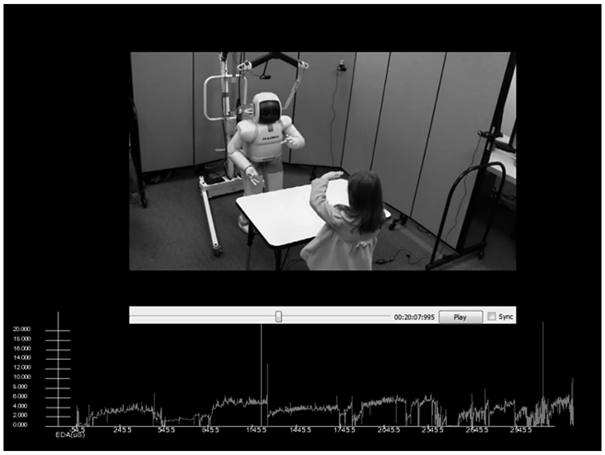


Fig. 28. Software synchronizes skin conductance signals with video recordings of sessions.

(see Figure 27) comprised an ideal device. Recently introduced by Affectiva, Inc. [Poh et al. 2010], the work was based on research by Picard [1997] on affective computing. These sensors were attached to palm straps or wristbands that allowed data recording from the child's arm during interaction. This allowed children to move freely, gesture, and manipulate objects during interactions. The device also included temperature sensors and accelerometers to measure hand motions. Skin conductance was sampled at 32 Hz and stored in a small micro SD memory card (within the device) during the session.

9.1. Calibration and Postprocessing

In order to synchronize the physiological signal with recorded video, subjects were asked to perform an exercise where they raised their hands up and down as fast as they could. The accelerometer in the sensor helped locate the position in the physiological data and was synchronized with the video (see Figure 28). In a later version of the sensor, we were able to directly synchronize the sensor's time with a host computer. The skin conductance data was measured in microsiemens and filtered to remove noise and motion artifacts (that appeared as high-frequency spikes). A fourth-order butterworth low-pass filter was applied with a cutoff frequency of .00625 Hz. The filtered signal was applied to the Matlab detrend filter which removed the mean value from the sampled vector to remove DC drift in the sensors. The signal was more reliable once the sensors had warmed to body temperature, and perspiration appeared at contact interface. For baseline, it was important to obtain a signal that was recorded after the electrodes had been warmed to body temperature. A portion of the signal with relatively little activity prior to the session was selected as the baseline whose mean was removed from the signal. Final range normalization was done to map the minimum and maximum conductance values to 0 and 1, respectively (see Figure 29).

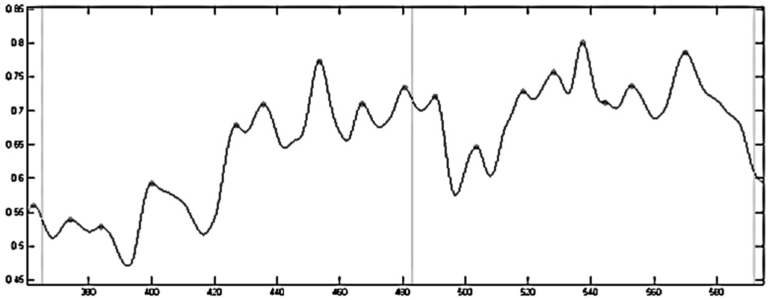


Fig. 29. Skin conductance response peaks denoted with red circles.

9.2. SCR Detection

From the normalized signal, several measures were obtained. All Skin Conductance Response (SCR) peaks and valleys in the filtered signal were found with time stamps. The distance of the rise between a peak and its previous valley was also measured. From these measurements, we partitioned the SCR peaks according to when they occurred during the phase of interaction: introduction, BQ1, BQ2, BQ3, BS, BVQ1, BVQ2, BVQ3, BVS. The SCR peaks were tallied to compute the mean and standard error of the mean for each phase.

9.3. GSR Analysis

Since the skin conductance data for each subject was normalized, performing intersubject comparisons with physiological skin conductance data was problematic because of inherent individual differences in baseline arousal. Even if the baseline was subtracted from the raw data, individual's skin conductance changes may occur at different rates. Normalizing the data to a specific range may help accentuate the SCR peaks for an individual, but removes relative magnitude information between subjects, making direct comparisons impossible. Consequently, the challenge lies in how to compare or aggregate physiological information between subjects.

Since the length of interaction (differing by subjects) may influence the number of SCR events, the average number of SCR events per minute was computed to normalize the effects of duration. Since all participants experienced the same experimental procedure, the time series were partitioned according to interaction event intervals (i.e., phase of interaction: introduction, BQ1, BQ2, BQ3, BS, BVQ1, BVQ2, BVQ3, BVS). This allowed us to compare the arousal activity across all subjects during the same situation. Arousal activity was estimated by counting the number of SCR (skin conductance response) events that occurred during the time interval. We found that in more responsive conditions (i.e., woz, automatic) interactions were prolonged with the child, requiring the need to normalize the results to eliminate the time bias. For each interactive interval, the total number of SCR events within that interval was divided by the total duration of the interval in minutes to produce the average number of SCR events per minute for each interval. Due to technical difficulties, physiological data could only be retrieved from 6 children out of the total 9 in the study. Preliminary findings showed there was a potential trend that woz and automatic conditions produced a higher number of SCR incidents than the nothing condition (see Figure 30). In this early trial measurement we did see a promising trend of increased average of SCR events ranging from 3.87 events for nothing conditions to 4.56 events for automatic, and 6.00 events for the woz condition. As a small sample size, no claims could be made without a larger sample. Initial indications showed that skin conductance could potentially be useful for characterizing different levels of arousal in subjects

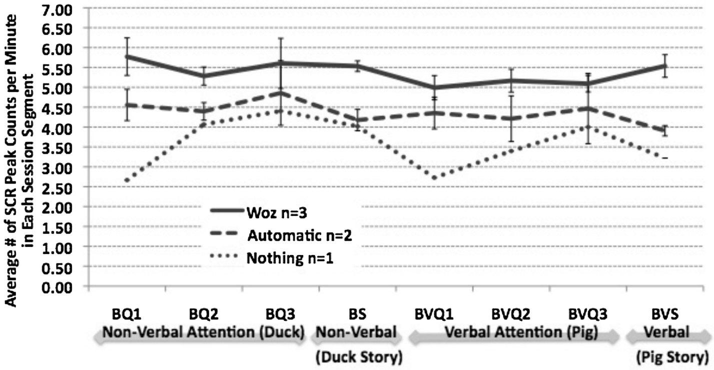


Fig. 30. Average number of SCR peaks per minute by session.

during interaction. Further analysis of the video and audio with skin conductance is currently being processed to identify potential triggers in producing affective response. The study plans to make changes based on the findings, and run a larger full-scale study to confirm the overall trend.

10. GENERAL OVERALL DISCUSSION

The two studies (interaction style study and quality of attention study) involved children's interaction with life-sized humanoid robots. One challenge in our research involved exploring how humans can become part of the design, and part of the system. Two approaches (perspective of robot and perspective of humans) seemed critical in designing and developing specific features in robots. Studies were used to test what features contributed to affective behavior and prolonged engagement. More specifically, we needed to know what robots need to detect from humans, and how robots should respond to humans. The article focused on measuring the influence of these features on affective behavior and engagement.

The strategy had been to produce a suite of hardware and software tools to fit a natural environment. By enabling comprehensive recordings of multiple modalities of interaction, and understanding how humans respond to different behavioral stimuli, the study hoped to achieve childerns' affective behavior and engagement. Our software framework allowed both automatic and manual Wizard-of-Oz interaction models to be created, and provided the researcher with a wide selection of controls for designing different interactive conditions.

The development of automated comprehensive data logging and multimodal analysis assisted researchers in the study of extended human-robot interaction, and revealed long-term trends and patterns that were not observable in short exchanges. One large challenge was the subsequent analysis of all the collected data. Although some automatic computer vision methods could be used to automatically categorize face poses and facial expressions, these methods were susceptible to large error rates, especially when facial features were small in the camera images. Still there was much to be done manually by humans for video coding verbal and nonverbal cues.

11. FUTURE WORK

The studies showed that even though the automatic condition did not perform as well as the woz condition, the attention feature still influenced children's affective behavior and engagement. This implies that we can design more natural and affective interactions with robots, especially in the construction of autonomous behavior models with better human detection and response mechanisms. Another important feature is creating robust speech recognitions for young children, and engaging the robot in a timely manner. There is definitely a need to run a larger study with more participants to design and evaluate an affective decision-making module. For now, interactions may be most affective when interactions involve a peer-like robot designed around a familiar script, and an attention feature that includes both nonverbal and verbal attention. Until robots have the intelligence to flexibly respond to young children's interactive bids, following a well-known script or protocol (e.g., turn taking, story telling, playing house) may help elicit affection through prior knowledge and familiarity. The set of findings from our small-scale exploratory study has brought about useful design implications.

As future work, we would also like to analyze the sequence of events in the human-robot interactions (both children and adult users), to produce better predictive behavior models. This may help anticipate the user's response or change in emotions, and produce more pleasant interactions. A good prediction model can allow a robot to steer a conversation toward a desirable goal or help assist users when navigating through an interactive task. It is important for the robot to have early indications of whether the user is engaged, afraid, or bored. A robot can then decide to dynamically modify its own style of communication while still attempting to fulfill the task goals. A possible application is designing robots for educational purposes as a peer-learner. A robot with good prediction, detection, and feedback models can provide the learner with good scaffolding and motivation.

12. SUMMARY

The two studies attempted to see if specific design features, that is, in robots influence children's affective behavior and social engagement. The first study looked at two features, that is, interaction styles (e.g., lecture style, cooperative, self-directed) and general features (e.g., human-like voice, monotone robot-like voice) to see if different responses influence affective behavior. Results found that cooperative interaction style and human-like voice invited more oculesic behavior and elicited more comments from children during interaction. The second study looked at the quality of attention (e.g., nothing/no attention, automatic attention, and Wizard-of-Oz) to see if different levels of detection influence affective behavior. The study also looked at the type of attention, nonverbal attention, and verbal attention, to see if different responses influence affective behavior. The findings revealed that woz and automatic detection and verbal attention responses increased children's oculesic behavior, length of conversation, and physiological responses. The findings will be applied to design better behavior and interaction models. There is much to be learned about the human partner in relation to the technical partner. In order to reveal new insights to the role robots play in human relationships, robots must accurately detect human behavior and the experience humans bring with them. This combined with careful selection of technology may unfold successful engagement and affective human-robot interaction.

ACKNOWLEDGMENTS

Many thanks to Research Assistants from Teachers College, Columbia University, Amanda Rosenberg, Mihwa Kim, and Yumiko Murai, and Maya Fujimura from University of California, Irvine who helped with transcribing and initial coding of the data.

REFERENCES

Anderson, P. A. 2008. *Nonverbal Communication: Forms and Functions*. Waveland Press, Prospect Heights, IL.

Arsenio, A. 2004. Children, humanoid robots, and caregivers. In *Proceedings of the 4th International Workshop on Epigenetic Robotics*, L. Berthouze, H. Kozima, C. G. Prince, G. Sandini, G. Stojanov, G. Metta, and C. Balkenius, Eds., 117, 19–26.

BAR-COHEN, Y. AND BREAZEAL, C. L. 2003. *Biologically Inspired Intelligent Robots.* SPIE Press, Bellingham, WA.

BARTNECK, C. AND FORLIZZI, J. 2004. Shaping human-robot interaction: Understanding the social aspects of intelligent robot products. In *Proceedings of the Computer Human Interaction Workshop*. ACM, New York, 1731–1732.

DAUTENHAHN, K. AND WERRY, I. 2002. A quantitative technique for-analyzing robot-human interactions. In *Proceedings of the International Conference on Intelligent Robots and Systems*. IEEE/RSJ, 1132–1138.

GELMAN, S. A. AND GOTTFRIED, G. M. 1996. Children's causal explanations of animate and inanimate motion. *Child Devel. 67*, 1970–1987.

HAUSER, K., NG-THOW-HING, V., AND GONZÁLEZ-BAÑOS, H. 2007. Multimodal motion planning for a humanoid robot manipulation task. In *Proceedings of the 13th International Symposium on Robotics Research*.

HONDA MOTOR CO., LTD. 2000. ASIMO year 2000 model. http://world.honda.com/ASIMO/technology/spec.html.

KANDA, T., HIRANO, T., EATON, D., AND ISHIGURO, H. 2004. Interactive robots as social partners and peer tutors for children: A field trial. *Hum.-Robot Interact. 19*, 61–84.

MUTLU, B., OSMAN, S., FORLIZZI, J., HODGINS J., AND KIESLER, S. 2006. Task structure and user attributes as elements of human-robot interaction design. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication.* IEEE,74–79.

NG-THOW-HING, V., LUO, P., AND OKITA, S. 2010. Synchronized gesture and speech production for humanoid robots. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS'10)*. IEEE/RSJ, 4617–4624.

NG-THOW-HING, V., THÓRISSON, K. R., KIRAN SARVADEVABHATLA, R., AND WORMER, J. 2009. Cognitive map architecture: Facilitation of human-robot interaction in humanoid robots. *IEEE Robot. Autom. Mag. 16*, 1, 55–66.

OKITA, S. AND SCHWARTZ, D. L. 2006. Young children's understanding of animacy and entertainment robots. *Int. J. Hum. Robot. 3*, 3, 393–412.

OKITA, S. Y., NG-THOW-HING, V., AND SARVADEVABHATLA, R. K. 2009. Learning together: ASIMO developing an interactive learning partnership with children. In *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 1125–1130.

PICARD, R. 1997. *Affective Computing*. MIT Press, Cambridge, MA.

POH, M. Z., SWENSON, N. C., AND PICARD, R. W. 2010. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Trans. Biomed. Engin. 57*. 5, 1243–1252.

RAY, G. B. AND FLOYD, K. 2006. Nonverbal expressions of liking and disliking in initial interaction: Encoding and decoding perspectives. *Southern Comm. J. 71*, 45–64.

ROBINS, B., DAUTENHAHN, K., TE BOEKHORST, R., AND BILLARD, A. 2004. Effects of repeated exposure to a humanoid robot on children with autism. In *Designing a More Inclusive World,* S. Keates, J. Clarkson, P. Langdon, and P. Robinson, Eds. Springer, 225–236.

SARVADEVABHATLA, R., NG-THOW-HING, V., AND OKITA, S. Y. 2010. Extended duration human-robot interaction: Tools and analysis. In *Proceedings of the 19th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 7–14.

SARVADEVABHATLA, R. AND NG-THOW-HING, V. 2009. Panoramic attention for humanoid robots. In *Proceedings of the 9th IEEE/RAS International Conference on Humanoid Robots*. IEEE, 215–222.

SCHERER, K. R. AND OSHINSKY, J. 1977. Cue utilization in emotion attribution from auditory stimuli. *Motiv. Emot. 1*, 331–46.

SHERIDAN, T. B. 2002. *Humans and Automation: System Design and Research Issues*. John Wiley & Sons, New York, 162–165.

WADA, D., SHIBATA, T., SAITO, T., AND TANIE, K. 2002. Analysis of factors that bring mental effects to elderly people in robot assisted activity. In *Proceedings of the International Conference on Intelligent Robots and Systems*. IEEE, 1152–1157.

YOSHIDA, H. AND SMITH, L. B. 2008. What's in view for toddlers? Using a head camera to study visual experience. *Infancy 13*, 3, 229–248.

Received December 2010; revised August 2011; accepted August 2011