Towards the Development of Personalized Learning Companion Robots for Early Speech and Language Assessment

Gary Yeung[1], *Alison L. Bailey*[2], *Amber Afshan*[1], *Marlen Q. Pérez*[2], *Alejandra Martin*[2], *Samuel Spaulding*[3], *Hae Won Park*[3], *Abeer Alwan*[1], *Cynthia Breazeal*[3]

[1]Dept. of Electrical and Computer Engineering, University of California, Los Angeles, USA
[2]Dept. of Education, University of California, Los Angeles, USA
[3]Media Lab, Massachusetts Institute of Technology, Cambridge, USA

Abstract

This pilot study investigated the feasibility of implementing child-friendly robots for administering clinical and educational assessments with young children. JIBO, a social robot, was used as a new interface to administer a letter and number naming task and the 3rd Goldman Fristoe Test of Articulation (GFTA-3). The reason for using these assessment materials is to develop robust automatic speech recognition (ASR) and automated social interaction systems that can aid in administering such assessments more efficiently. The voice of JIBO simulates interaction with a peer, and images and playful transitions are displayed on JIBO's face/screen. Several preliminary observations with 15 pre-kindergarten and 18 kindergarten students included the rate of task completion and strategies to increase student participation. Changes to the length and prompt delivery of the assessment protocol were considered based on these observations, and further observations are planned for future work with an additional cohort of 43 pre-kindergarten and 50 kindergarten students. Recommendations are given to inform future implementations and analyses.

Objectives

This study is part of a larger project, supported by the National Science Foundation (NSF), investigating the usage of social robots for the early education of young children. This project will include the investigation of automatic speech recognition (ASR) for young children, multi-modal assessment algorithms for language and reading skills, automatic personalization of story content and questions, and effective methods for developing expressive, empathetic, and socially responsive robots. The presented study is the first phase of this project, exploring the feasibility of using social robots with young children in a school setting. Pre-kindergarten (pre-K) and kindergarten children were chosen because this age group could benefit the most from additional exposure to spoken vocabulary and dialogue [1-5]. The major questions explored in this study are as follows:

1. How willing were the pre-K and kindergarten students to interact with the robot to complete oral language and speech assessments?
2. In what ways can the child-robot interaction be improved through additional dialogue?
3. What changes to the technology will be needed to make administrations autonomous?

Perspectives

The teaching of language and literacy skills is one of the most important goals of elementary education. However, only 35% of 4th grade students reach age-appropriate reading proficiency according to the National Assessment of Educational Progress [6]. Pre-literacy skills such as phonological and alphabetic knowledge developed at the pre-K and kindergarten level can support the development of literacy skills [1-5]. Additionally, there is a close relationship between the development of oral language skills and the development of reading skills [7]. An essential factor to the development of language skills is sufficient exposure to a rich variety of oral vocabulary [8]. Another critical component of such exposure is the social context of learning: children need to be actively engaged physically and emotionally to maximize learning potential [9]. This need also holds in effective assessment situations [10, 11].

The use of social robot companions for clinical and educational applications has great potential in aiding the efforts of teachers, clinicians, and parents. Intelligently designed robots can be capable of autonomously engaging with children through playful educational activities without the need for much human interaction, thus freeing up educators to focus on other educational needs. Activities can be designed to accomplish specific goals such as the evaluation of speech and phonetic acquisition or the teaching of pronunciation. Meanwhile, individual interaction between a child and robot allows for the system to adapt to children’s individual needs, enhancing their strengths or focusing on their challenges. Such adaptive systems can also be used to better engage children by adapting to children’s speaking styles and behaving like an engaged listener [12].

JIBO, the social robot medium used for this study, was developed by Cynthia Breazeal to be a home personal assistant robot. Each JIBO is fitted with a three-axis motor system with 360-degree rotation, allowing it to make several expressive actions, such as tilting its head, looking up and down, and dancing. Additionally, with a small screen on its face, two cameras for face tracking, and microphone array for identifying locations of sounds, JIBO is a good fit for examining how robot companions can both interact with children and administer educational assessments.

Intelligent Tutoring Systems (ITS) that can adapt to a student’s skill level have been shown to positively impact student learning gains in students aged 10 years and older [13-15].

However, the development of long-term personalized interactions with younger children requires new age-appropriate and autonomous implementations [16]. The development of such systems for pre-K or kindergarten children, a critical time for early language and literacy intervention, remains a substantial challenge.

One of the major hurdles of this study is the poor performance of current ASR techniques to automate the scoring of children's responses. Child ASR is often riddled with errors, and current tools for performing child speech recognition are insufficient for human-robot interaction applications [17]. A recent study on kindergarten-aged children revealed further that kindergarten children (and presumably younger children) need to be targeted specifically when developing ASR systems [18]. The further development of ASR for young children can provide great benefits to the development of fully autonomous, peer-like social robot systems. The data that we have collected, and will continue to collect, will also be used for this purpose of developing robust ASR systems for children.

Methods and Data Sources

Participants

Participants attended a university demonstration elementary school in the southwestern United States. Social robots were introduced to teachers and students as part of their early science and technology inquiry-based curriculum. Approximately 40% of the school is enrolled in Spanish-English dual language immersion classrooms. Initially, 15 pre-K (5 boys, 10 girls) and 18 kindergarten (7 boys, 11 girls) students, enrolled in the 2017-2018 school year, were recruited to test the first implementation. Approximately one-third of both the pre-K and kindergarten participants were biracial. Additionally, approximately one-half of the pre-K participants and one-third of the kindergarten participants had some Latino ethnicity. Other ethnicities represented included Caucasian, Black, Asian, Pacific Islander, and Indian.

An additional cohort of 43 pre-K (23 boys, 20 girls) and 50 kindergarten (23 boys, 27 girls) students, enrolled in the 2018-2019 school year, will be recruited to evaluate the changes made to improve children's engagement with the robot and assessment protocol as a result of the current cohort findings. Additionally, the current cohort will be followed longitudinally to document their reactions to the social robot JIBO over time. Finally, parents of both cohorts were asked to complete a questionnaire that included student familiarity with digital devices.

Procedure

JIBO was programmed to administer the 3rd Goldman Fristoe Test of Articulation[1] (GFTA-3) Sounds in Sentences (SIS) and Sounds in Words (SIW) and a letter and number naming task. All images for words, letters, and numbers were displayed on JIBO's face/screen. Instructions, prompts, and friendly interactions were recorded by a female researcher, with recordings pitch-shifted to sound like a young child's voice.

Each student individually interacted with JIBO alongside two researchers. A Logitech C390e webcam was used as a microphone and placed at a 45-degree angle to the child, approximately 1-2 feet away. One researcher sat next to the child as an "instructor" and played with JIBO alongside the child. The other researcher sat behind JIBO as an "operator" controlling the display of items with a computer. JIBO first introduced itself and asked warm-up questions (e.g., "What is your name?", "What is your favorite color?") to put the child at ease and serve as an example to the child of the question-and-answer style of dialogue that would take place while the assessment was being administered. Additionally, throughout the session, JIBO would occasionally praise the child for answering a question correctly (e.g., "You're doing an awesome

[1] The examples of the GFTA-3 used throughout this paper are strictly notional.

job!”, “That’s so cool!”). At the end of the session, JIBO would thank the child for playing, say goodbye, and laugh in response to being petted by the researcher and/or child. An example setup is depicted in Figure 1. Sessions were initially designed to be approximately 30-minutes long.

Sounds in Sentences

The GFTA-3 SIS task for ages 2-7 years involved JIBO telling a story about children walking home. Five pictures were shown in chronological order, one at a time. Each picture had 3-5 associated sentences from the story. After the first telling of the story, JIBO told the story again, as well as displayed the associated pictures, and instructed the student to repeat each sentence.

Sounds in Words

The GFTA-3 SIW task prompted children to say 58 different words by showing a picture on JIBO’s screen along with JIBO prompting the child with various questions or sentences. When asking the child to identify an object (e.g., “bird,” “tree”), JIBO would display a picture and ask, “What is this?” In some cases, a single image could correspond to multiple objects (e.g., a picture of a “ball” and “bat”). In these cases, the item to be identified would be highlighted. Additionally, some words were prompted by a complete-the-sentence prompt (e.g., “She is short, and she is…” for the word “tall”). If the child did not respond or said an incorrect word, the operator would have JIBO give the desired word as well as a secondary prompting for the word.

Letters and Numbers

A sequence of letters and numbers was randomly generated and individually displayed to the student on JIBO’s screen with the prompt “What letter is this?” or “What number is this?” This task had no secondary prompt.

Parent Survey

The parents of the children were asked to complete a survey about their child's languages spoken, reading habits, and familiarity with technology. The parents of all the past and future participants of the study were targeted. Approximately 90% of the participants live in a household that speaks English most of the time. Similarly, approximately 95% of the participants speak English most of the time. Over 70% of pre-K participants and 80% of kindergarten participants have some exposure to computers, smartphones, or tablet devices. Full statistics of the responses are summarized in Table 1.

Findings and Implications

In this pilot study, our main goal was the successful completion of the GFTA-3 and the letters and numbers tasks using JIBO to administer the task. This section documents the interactions between JIBO and the initially recruited students and notes possible improvements to these interactions to increase successful completion.

Completion of the Initial Protocol

The initial protocol was tested on the students enrolled in the 2017-2018 school year. Approximately 80% of the students completed the initial protocol. Of the pre-K children, 12 (80%) of the students (4 boys, 8 girls) successfully completed the task. Of the kindergarten children, 14 (78%) of the students (4 boys, 10 girls) successfully completed the task. One boy in pre-K refused to play with JIBO. One girl in pre-K chose not to perform the SIS task as she was a Spanish speaker still unfamiliar with English. Another girl in pre-K appeared intimidated by JIBO, and the session was halted by the researchers. Three boys in kindergarten stopped during the SIW task due to fatigue or boredom from the length of the task. One girl in kindergarten only finished the SIS task as she also did not enjoy playing with JIBO.

During the experiment, a few adjustments were considered to compensate for the length of the various tasks and the unease of some students. First, the instructor would interject throughout the SIW task to discuss the pictures (e.g., “What do you think of the picture?”). This was mainly done to reduce the number of times the child had to hear the prompt “What is this?” and to give the student a break from picture naming. Initial testing suggested that this method was effective at keeping the student focused throughout the task. In an autonomous implementation, the robot would be able to identify when the child was becoming fatigued from vocal, facial, or speech cues and administer a different discussion prompt. Another method considered was breaking the SIW, letters, and numbers tasks into shorter tasks over multiple sessions.

Sounds in Sentences

Two main difficulties occurred during the SIS task. First, many of the students had difficulties identifying the cues to begin repeating after JIBO. For example, when JIBO administered a compound sentence, many of the students would start repeating after JIBO completed the first part of the sentence and stop again once they realized the sentence was not yet finished. We are currently considering solutions to this problem. First, we could reduce the pause when JIBO says various clauses and lists so the student does not get confused about when JIBO has ended the sentence. This may require an investigation into how the children react to prosody changes. Another solution is to signal when the child should speak so the child will continue to listen until the signal occurs.

A second problem was that some children were unable to remember the entire sentence. This occurred when the sentence had a compound predicate or when the sentence had lists such as a list of adjectives (e.g., “The cat was fat, fuzzy, and orange”). During our pilot study, the instructor would follow up and break the sentence into parts if the student could not remember the entire sentence. One possible solution to this is to have JIBO be prepared to break the sentence into parts if necessary. Visual or acoustic signals for JIBO to identify when a child needs help must be investigated if the system is to be implemented in an autonomous way.

Sounds in Words

Besides the length of the SIW task, there was some difficulty in recording the initial phoneme of the article “the” or demonstrative pronouns “this” and “that.” This sound tends to occur in such words that are conceptually difficult for the child to identify, even after multiple prompts. In the future, prompts might be adjusted to record this phoneme while adhering to the GFTA-3 protocol.

Letters and Numbers

While the kindergarten students performed well on the letter and number naming task, some of the pre-K students had not mastered their letters and numbers. In these cases, the instructor would help the child identify the letter or number (e.g., “This letter is Q. Now you tell me the name of the letter.”). JIBO currently uses a single prompt, either “What letter is this?” or “What number is this?” In the future, a secondary prompt should be considered for this task to allow JIBO to assist the youngest students autonomously.

Scholarly Significance

This exploratory social robot study with young children has the potential to develop effective, scalable, and affordable early childhood literacy and language diagnostic tools. Existing studies have shown that ITS can have positive impacts on student learning [13-15]. While robots used for education in general need not necessarily simulate social interaction, social robots may be critically important for kindergarten students and younger age groups. JIBO interacted with students in a friendly way by introducing itself and providing praise, putting most students at ease for working with a robot assessor. Such a social component of educational robots may allow for more effective usage with young children, and our preliminary findings have begun to illuminate how to accomplish this.

For the assessment to be fully automated, ASR systems for child speech must be able to both recognize a child's speech and evaluate their performance. With the addition of child ASR tools and proper interactions, JIBO can be trained to administer several different tasks and evaluations. This could include the critical teaching of literacy, vocabulary, and social skills [19, 20]. We plan on evaluating such systems in our future work.

References

1. B. Hart and T. R. Risley, *Meaningful differences in the everyday experience of young American children.* Baltimore, MD: Paul H. Brookes Publishing, 1995.

2. M. Fish and B. Pinkerman, "Language skills in low-SES rural Appalachian children: Normative development and individual differences, infancy to preschool," *Journal of Applied Developmental Psychology*, 23(5), pp. 539-565, 2003.

3. M. M. Páez, P. O. Tabors, and L. M. López, "Dual language and literacy development of Spanish-speaking preschool children," *Journal of Applied Developmental Psychology*, 28(2), pp. 85-102, 2007.

4. C. E. Snow, M. V. Porche, P. O. Tabors, and S. R. Harris, *Is literacy enough? Pathways to academic success for adolescents*. Baltimore, MD: Paul H. Brookes Publishing, 2007.

5. A. Biemiller and N. Slonim, "Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition," *Journal of Educational Psychology*, 93(3), pp. 498-520, 2001.

6. National Center for Education Statistics, *The Condition of Education 2015 (NCES 2015-144)*. Washington D.C.: National Center for Education Statistics, 2017. [Online]. Available: https://nces.ed.gov/pubs2015/2015144.pdf

7. C. E. Snow, "The theoretical basis for relationships between language and literacy in development," *Journal of Research in Childhood Education*, 6(1), pp. 5-10, 1991.

8. S. S. Asaridou, Ö. E. Demir-Lira, S. Goldin-Meadow, and S. L. Small, "The pace of vocabulary growth during preschool predicts cortical structure at school age," *Neuropsychologia*, 98, pp. 13-23, 2017.

9. G. Wells, "Dialogic inquiry in education." In C. D. Lee and P. Smagorinsky (eds.). *Vygotskian Perspectives on Literacy Research*, Cambridge: Cambridge University Press, pp. 51-85, 2000.

10. A. L. Bailey, "Theoretical and developmental issues to consider in the assessment of young students' English language proficiency." In M. K. Wolf and Y. G. Butler (eds.). *English Language Proficiency Assessments for Young Learners: (Innovations in Language Learning and Assessment at ETS)*, New York, NY: Routledge, pp. 25-40, 2017.

11. A. L. Bailey, "Assessing the language of young learners." In E. Shohamy and N. H. Hornberger (eds.). *Encyclopedia of language and education, Vol. 7: Language testing and assessment*, Berlin: Springer, pp. 379-398, 2008.

12. H. W. Park, M. Gelsomini, J. J. Lee, and C. Breazeal, "Telling stories to robots: The effect of backchanneling on a child's storytelling," in *Proc. of the AMC/IEEE International Conference on Human-Robot Interaction (HRI)*, 2017, pp. 100-108.

13. A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User Modeling and User-Adapted Interaction*, 4(4), pp. 253-278, 1994.

14. M. C. Desmarais and R. S. Baker, "A review of recent advances in learner and skill modeling in intelligent learning environments," *User Modeling and User-Adapted Interaction*, 22(1-2), pp. 9-38, 2012.

15. M. V. Yudelson, K. R. Koedinger, and G. J. Gordon, "Individualized Bayesian knowledge tracing models," in *Proc. of the International Conference on Artificial Intelligence in Education*, 2013, pp. 171-180.

16. B. P. Woolf, *Building intelligent interactive tutors: Student-centered strategies for revolutionizing E-learning*. Burlington, MA: Morgan Kaufmann, 2010.

17. J. Kennedy, S. Lemaignan, C. Montassier, P. Lavalade, B. Irfan, F. Papadopoulos, E. Senft, and T. Belpaeme, "Child speech recognition in human-robot interaction: Evaluations and recommendations," in *Proc. of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2017, pp. 82-90.

18. G. Yeung and A. Alwan, "On the difficulties of automatic speech recognition for kindergarten-aged children," in *Proc. of INTERSPEECH*, 2018.

19. J. M. Kory, "Storytelling with robots: Effects of robot language level on children's language learning," Ph.D. dissertation, Massachusetts Institute of Technology, 2014.

20. J. M. Kory and C. Breazeal, "Storytelling with robots: Learning companions for preschool children's language development," in *Proc. of the IEEE International Symposium on Robot and Human Interactive Communication*, 2014, pp. 643-648.

Table 1. Statistics of the parental survey regarding participants' languages known, languages spoken, reading habits, and familiarity with basic technology.

Questions	Responses	% of pre-K Participants	% of K Participants
What languages are spoken in the child's household?	Just English	77%	78%
	English and another language	14%	13%
	Another language besides English	9%	9%
What languages does the child speak most often?	Just English	86%	87%
	English and another language	9%	9%
	Another language besides English	5%	4%
How often does the parent/guardian read to the child?	Every day	100%	91%
	Once a week	0%	9%
	Almost never	0%	0%
How often does the child read by himself/herself?	Every day	59%	70%
	Once a week	14%	17%
	Almost never	27%	13%
How often does the child use a computer or tablet/phone for reading?	Every day	14%	30%
	Once a week	32%	18%
	Once a month	18%	22%
	Almost never	36%	30%
How often does the child use a computer or tablet/phone for games?	Every day	18%	39%
	Once a week	46%	22%
	Once a month	9%	22%
	Almost never	27%	17%

Figure 1. An example setup of a child and JIBO interacting along with an “instructor” to the side of the child and an “operator” behind JIBO.