SLaTE 2019: 8th ISCA Workshop on Speech and Language Technology in Education
20-21 September 2019, Graz, Austria

A robotic interface for the administration of language, literacy, and speech pathology assessments for children

Gary Yeung[1], Alison L. Bailey[2], Amber Afshan[1], Morgan Tinkler[1], Marlen Q. Pérez[2], Alejandra Martin[2], Anahit A. Pogossian[2], Samuel Spaulding[3], Hae Won Park[3], Manushaqe Muco[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

garyyeung@g.ucla.edu

Abstract

A preliminary implementation of a robotic interface for the administration of language, literacy, and speech pathology assessments for children is presented. This robot assessment protocol will be used for several ongoing studies to improve the performance of educational robots for children. The robot used is JIBO, a personal assistant-style robot capable of expressing itself with its poseable body. JIBO's implementation is intended for children as young as 4 years old. JIBO is designed to have friendly interactions with young children while administering assessments such as the evaluation of pronunciation, alphabetic knowledge, and explanatory discourse. Additionally, this implementation is currently being used to collect a speech database of such assessments being administered to children.

Index Terms: human-robot interaction, child speech, educational assessment

1. Robotic Setup

The robot to be used for this series of studies is JIBO, a robot originally developed to be a personal assistant with a simplistic design. Figure 1 shows a picture of a black JIBO during an administration session. JIBO is approximately 30 cm in height, 2 kg in weight, and has either a white or black body. In general, JIBO is meant to be placed on a table, desk, or counter and contains no wheels or other mechanisms to travel to a different location. JIBO's main feature is its ability to express itself with body poses and gestures using just a torso and head.

The center of JIBO's face has a 6.3×11 cm touch screen. The screen usually displays JIBO's eye (a grey circle) that moves along with JIBO's body (e.g., looking to the right as JIBO turns its body right). However, the screen can also be used for other purposes such as displaying text, images, or videos. The touch aspects can be used to identify objects or make selections on the screen. Two cameras are located above the screen. A microphone array is located across JIBO's head for identifying the direction of a sound. Two speakers are located on each side of JIBO's head for speech or music.

JIBO is currently being evaluated in several pilot studies. At this time, JIBO is controlled using a Wizard-of-Oz setup with an "operator" controlling JIBO through a graphical user interface that runs on a Linux-based operating system. JIBO's software is based on the Robot Operating System (ROS).

Figure 1: *An example recording session.*

2. Interactions and Assessments

The goal of our current study is to provide JIBO with more advanced capabilities to interact naturally with children while running any number of educational assessments. To this extent, a child speech database is currently being collected using several assessments implemented in JIBO. JIBO's current protocol is based on our initial implementation [1]. Furthermore, a number of additional assessments and interactions have been added to JIBO to improve the quality of data collection and enjoyment of the child subjects. These are documented in this section.

2.1. 3rd Goldman-Fristoe Test of Articulation

JIBO is currently equipped to administer the 3rd Goldman-Fristoe Test of Articulation (GFTA3) in its base form. The GFTA3 is an assessment of pronunciation and articulation of consonant and vowel sounds commonly performed by clinical speech pathologists. The GFTA3 is separated into two main parts. GFTA3 Sounds-in-Words (SIW) is a picture identification task consisting of 60 pictures. In this task, a picture is shown to the child often accompanied by the prompt "What is this?" GFTA3 Sounds-in-Sentences (SIS) is a "repeat-after-me" task in which the child repeats the sentences JIBO says. These sentences are usually part of a story, and the story changes in difficulty depending on the age of the child. See [1] for additional details.

2.2. Letters and Numbers

JIBO has two new assessment games to assess a child's knowledge of letters and spelling. For younger children ($< 1^{st}$ grade), the Alphabet Train Game evaluates a child's ability to identify letters. JIBO shows a scrolling train on its screen with a letter on each car, and the child attempts to identify each letter. For older children, the Spelling Game can assess a child's ability to spell basic words. A word is shown on JIBO's screen along with a corresponding picture (e.g., a picture of a hat and the word "hat"). After the child reads the word, the word disappears, and JIBO waits for the child to attempt to spell the word.

In addition to assessments of letters and spelling, JIBO has two new ways to evaluate a child's knowledge of numbers and math. For younger children, the Finger Game can assess a child's counting skills. JIBO shows the child a picture of hands with some fingers raised and asks the child to count the number of fingers raised aloud. For older children, the Real World Math Game can assess a child's ability to apply counting and math skills to more interesting real-life situations. This could vary from addition using candy (e.g., "How many pieces of candy do you have now?") to a picture of a child's birthday (e.g., "How old is the child?" with five candles on the birthday cake).

2.3. Explanatory Discourse

JIBO has several explanation tasks it can administer to evaluate a child's discourse skills. For these tasks, JIBO converses with the child about open-ended reasoning tasks or their daily routines. A few examples of questions are given in the remainder of this section.

For an example of a discourse on a personal routine, JIBO can show the child a picture of a boy and girl brushing their teeth and say, "Here is a picture of a boy and girl brushing their teeth. Tell me how you clean your teeth." JIBO follows the question with "Why do you clean your teeth?" and "Can you explain to a friend how to clean their teeth?"

For a discourse on reasoning, JIBO can show the child a picture of four animals (bird, cat, elephant, fish) and ask the child which animal is the "odd-one-out." JIBO follows the question by asking why the child answered the way they did.

For a discourse on problem solving, JIBO can ask the child to look at several blocks in a container and figure out how many blocks are in the container. The blocks can be attached or detached to each other to assist with solving the problem. Afterwards, JIBO asks the child what they did to figure out how many blocks were in the container. Finally, JIBO asks how the child's strategy helped with solving the problem.

3. Data Collection

3.1. Participants

Our current data collection includes approximately 60 pre-kindergarten students (4-5 years old) and 70 kindergarten students (5-6 years old) from a university demonstration school in the southwestern United States. Social robots were introduced in classrooms as part of an early technology inquiry-based curriculum. Approximately 40% of the student population is enrolled in English-Spanish dual immersion language programs.

The continuing data collection will be longitudinal. Speech data from the same students, now in kindergarten and 1^{st} grade, using similar assessments is currently being collected. Additionally, speech from a new cohort of pre-kindergarten children will be collected.

3.2. Recording Setup

Audio was captured using a Logitech C390e webcam placed at a 30°-45° angle to the child, approximately 50 cm away from where the child was seated. JIBO was placed directly in front of the child on a table or desk, approximately 50 cm away from the child. The "operator" was located behind JIBO during the session. As JIBO's development is still at a preliminary stage, to handle unexpected interactions between JIBO and the child, an "instructor," an education researcher, also sat nearby. An example recording session with JIBO and a child is shown in Figure 1.

4. Ongoing Research

The ongoing research motivated by this project includes improvement of child automatic speech recognition (ASR) [2], design of robotic systems and evaluation systems [3, 4], and increasing the effectiveness of educational evaluations. The data collected using JIBO will be critical to accomplish these goals.

5. Conclusion

In this paper, we have described the basic functionality of JIBO and its potential for administering educational assessments for young children. Several of the implemented assessments were described. Additionally, we have provided a brief description of the speech database currently being collected and the participants being recorded.

A number of improvements are still necessary. First, for the system to be fully autonomous, JIBO needs the ability to both recognize a child's speech and evaluate their performance. Additionally, JIBO would likely need a mechanism to identify when the child is getting tired or frustrated during the session, to which JIBO can respond by changing topics, taking a break, or ending the session. This would likely require some form of emotion detection using speech and/or video. Our ultimate goal is to have interactions between JIBO and the child be less reliant on a library of audio and movement files. However, this would require dramatic improvements to child ASR and natural language understanding (NLU) systems.

6. Acknowledgements

This work was supported in part by NSF Grant #1734380.

7. References

[1] G. Yeung, A. L. Bailey, A. Afshan, M. Q. Pérez, A. Martin, S. Spaulding, H. W. Park, A. Alwan, and C. Breazeal, "Towards the development of personalized learning companion robots for early speech and language assessment," in *Proc. of the Annual Meeting of the American Educational Research Association (AERA)*, 2019.

[2] G. Yeung and A. Alwan, "On the difficulties of automatic speech recognition for kindergarten-aged children," in *Proc. of INTERSPEECH*, 2018, pp. 1661–1665.

[3] S. Spaulding, H. Chen, S. Ali, M. Kulinski, and C. Breazeal, "A social robot system for modeling children's word pronunciation," in *Proc. of the International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, 2018, pp. 1658–1666.

[4] G. Yeung, A. Afshan, K. E. Ozgun, K. Kaewtip, S. M. Lulich, and A. Alwan, "Predicting clinical evaluations of children's speech with limited data using exemplar word template references," in *Proc. of the Workshop on Speech and Language Technology in Education (SLaTE)*, 2017, pp. 161–166.