Media Characters in Inclusive CUIs for Preschoolers

Jennifer Rodriguez
Director, Research and
Inclusive Design, PBS KIDS
Public Broadcasting Service
Arlington, VA 22202, USA
jmrodriguez@pbs.org

Silvia Lovato
Sr Director, Research and
Evaluation, PBS KIDS
Public Broadcasting Service
Arlington, VA 22202, USA
slovato@pbs.org

CSCW 2021 CUI@CSCW Workshop, October 23-27, 2021, held virtually.

Abstract
PBS KIDS is developing socially contingent conversational user experiences (CUX) with media characters using an approach that supports voice, tap, and gesture inputs. In June 2020, we built the prototype "Elinor Talks" wherein children responded to terminal and open-ended questions posed by a familiar character through a video chat experience that aimed to provide contextual, reliable, and immediate feedback. Initial research conducted in December 2020 revealed challenges and opportunities inherent in the technology and in young children's exposure to and prior knowledge of CUX and gestural experiences. We share lessons learned and explore the potential for voice and gestural communication inputs to support learning in special populations (e.g. deafness, autism, cerebral palsy, Down Syndrome). Related to this work, we hope to develop design guidelines for voice and gesture-based user experiences that can support the varied needs of children.

Author Keywords
Authors' choice; of terms; separated; by semicolons; include commas, within terms only; required.

CSS Concepts
• **Human-centered computing~Human computer interaction (HCI)**; *Conversational User Interfaces*;

Children; User studies
https://dl.acm.org/ccs/ccs_flat.cfm

Introduction

In June 2020, PBS KIDS developed an iOS app prototype to power a socially contingent conversational agent capable of contextual, reliable, and immediate feedback in the form of PBS KIDS character, Elinor Rabbit, from the preschool series *Elinor Wonders Why.* The "Elinor Talks" prototype is a video chat experience that gives young learners alternative inputs -- voice and gesture -- to have a conversation with Elinor about observable phenomena (e.g., "I wonder why spiders have tiny hairs on their legs."). The intent of the prototype was to use principles of Universal Design for Learning to provide all children with multiple ways to engage and "show what they know" – acknowledging that young children are often better at demonstrating knowledge than answering questions – and multiple ways to interact beyond default touch-screen inputs [2] it included a variety of closed and open-ended prompts and flexible input options: touch, voice, and gesture. Terminal conversational prompts could be answered with a microphone-enabled "yes" or "no" reply or a camera-enabled thumbs up/thumbs down gesture. More open-ended questions (e.g., "What do you think of spiders?") were designed to elicit verbal responses with on-screen visual cues to signal turn-taking in the conversation, and with the conversation agent programmed to wait for (and prompt) responses.

The "Elinor Talks" prototype was tested in December 2020 with a sample of ten (n=10) preschoolers (age 4). The group included one English Language Learner (ELL), a child with a mild speech impediment, and families of mixed income. PBS KIDS contracted with a third-party research firm, Rockman Et Al, to conduct the sessions, some of which happened in-person during the COVID19 pandemic. These in-person testing scenarios occurred in a lab setting with children sometimes wearing masks as they conversed with Elinor.

Key research questions were:

- [APPEAL] To what extent is this type of play experience appealing to young children?
- [USABILITY] To what extent are young children able to successfully engage voice-based and gesture-based play?
- [LEARNING] What do young children learn from the play experience?

Social Contingency in "Elinor Talks"

Socially contingent interactions involve a two-way exchange that is appropriate in content and length with immediate and reliable responses [5]. Sometimes young children experience one-sided conversations through the direct address of a fictional TV character talking "with" them, as when a character asks a question to the audience and waits a moment for a response. The "Elinor Talks" prototype aimed to deliver a more realistic and immersive version of that experience through an informal video chat format with Elinor Rabbit. The main learning goals of the prototype were to engage children in using inquiry practices and related vocabulary, such as observation with augmentative tools like magnifiers and binoculars. The vocabulary-related goal seemed especially promising because empirical studies of socially contingent video chat experiences with toddlers showed they could learn

Figure 1: Selection menu where children can choose which tool is appropriate for the observation-related mission.

novel words from video chats and suggested that "socially contingent interactions are powerful catalysts for word learning" [5].

Building on this social contingency research, the production and research teams also considered the theoretical work of Alfred Bandura's Social Learning Theory [1] and his ideas that children are surrounded by many "influential models" including parents, peers, and characters on TV [1]. Children can learn behaviors from these models because of their capacity to (1) hold the child's attention; (2) support retention; and (3) motivate with rewards. Research has shown that children can develop parasocial relationships with characters and learn when the interaction is not reciprocal [4].

Therefore, the impact a sense of reciprocity and social contingency might have in the media experience of preschoolers. Could a character support the development of behaviors and skills through guided and, even, game-like conversational interactions? These questions guided the design of the "Elinor Talks" prototype. Another critical influence in the prototype's design were the principles of Universal Design for Learning.

Universal Design for Learning

Universal Design for Learning originates from the field of Universal Design developed by architects and engineers to increase access to buildings and products. The PBS KIDS production team developing the "Elinor Talks" prototype was well-versed in applying principles of Universal Design for Learning and focused on designing multiple ways to help learners engage or "show what they know" and multiple ways to interact [2]. This supports the inclusion of a variety of closed and open-ended prompts and flexible input options (touch, voice, and gesture) in the prototype design.

Prototype Development

This *Elinor Talks* game prototype was designed to test conversational interaction between a character and a player. It was developed in partnership with the development firm FS Studio and uses speech recognition (supported by the Apple Speech framework) and gesture recognition (supported by Google's Mediapipe). In it, children can choose from one of two possible investigations or mini-games - the bird-watching mission (which supports gestural interactions) OR the spider observation mission (which supports conversational user experience or CUX interactions).

In the spider observing mini-game, players experience interacting with Elinor primarily through the use of their voice, with touch-enabled buttons as fallbacks. Prompts are largely terminal (i.e., anticipating a "yes" or "no" reply) and the conversational agent "listens" for short replies before responding. In a couple of instances Elinor also prompts for more open-ended responses. Here are some example prompts:

- *What do you think of spiders? Do you like them, or do you think they're a little scary?* (open-ended)
- *Do you want to help me?* (terminal, listening for yes/no)
- *Can you help me choose which tool I should use - the magnifying glass or the measuring tool?* (terminal, listening for options presented)
- *Count the spider's legs with me!*

Figure 1: Demonstration of Google Mediapipe gesture tool hand tracking capabilities

In the birdwatching mini-game, players interact with Elinor using gesture and touch-enabled inputs. Elinor prompts children to help her investigate a distant baby bird and to select the best tool (i.e., binoculars vs. tape measure) for the job. Children are prompted to respond with the following gestures: finger counts; thumbs up/down; making their hands into the shape of tools like binoculars (see Figure 2); and flapping their wings to mimic the baby bird when it is hungry.

As an alternative or fallback for players who are unable to make these gestures or whose gestures are not read correctly, on-screen buttons with icons appear as alternative inputs (i.e., image of thumbs up or thumbs down). Prompts included:

- *How old are you?* (prompting player to show age with finger count)
- *Do you want to try and help me?* (prompting player by modeling and saying thumbs up/down)
- *Which tool do you want to use - flashlight or binoculars?* (prompting player by modeling and saying tool options)

Research Methods

Ten families with four-year-old children participated in testing the 'Elinor Talks" app (six girls), representing a mix of children enrolled in full-time and part-time preschool programs as well as those who were not attending preschool. The group included one English-language learner and a mix of lower to middle-income families.

Participating families agreed to an in-office testing session where they could be observed by a member of the Rockman Et Al research team from an adjoining room, separated by a one-way mirror. Participants were asked a series of questions before and after playing two mini-games within the Elinor Talks prototype. The prototype was also equipped to save conversation transcripts locally. Session data was reset between sessions.

Findings

Appeal of the "Elinor Talks" prototype

Both children and parents liked the games and appreciated the conversational nature of the game. Children were often eager to play the mini-games again, and when time permitted, they were allowed to do so and found that the second or third times playing were just as appealing as the first. A few parents also commented on the novelty of the games. All of the players realized they were having a conversation with a character from the game, rather than with a real person, though the interactions with Elinor still seemed genuine and served as further evidence of their engagement.

Usability of the "Elinor Talks" prototype

Voice input seemed to work more consistently than gesture-based inputs. Voice input also seemed more intuitive than the gestures in response to Elinor, however, use of the "thumbs-up" and "thumbs-down" signs for "yes "and "no" was easily understood. Voice inputs also required less prompting on the part of a researcher or parent to help facilitate successful in-game interactions. No participants knew anyone who communicated primarily through signs, and most had a hard time thinking of ways that they could say things

Figure 3: A 4-year-old research participant trying the gestural mode.

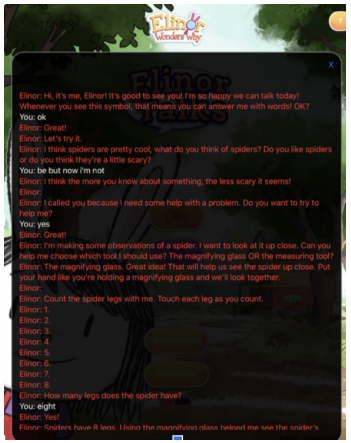

Figure 4: A sample conversation transcript from testing the CUX mode of the prototype

through signs prior to gameplay. Some of the children knew signs such as "thank-you," "please," and "more," because of having used "baby signs." One mother also noted that this type of game might be a fun way to help teach sign language. Another key research finding was that gesture-based inputs required precise conditions to work (i.e., body position 18 inches from the camera, precise lighting to avoid shadows). Working out the appropriate position so that the game could recognize gestures correctly was a persistent challenge. This stemmed from players' different standing and sitting heights, as well as distances from the device, even though the researcher provided tips on how to position themselves. It is also worth noting that children typically hold a tablet in their hands or on their lap when they play, so it may warrant further investigation to see if children can figure out how to position the iPad successfully for gesture-based inputs on their own.

Participants demonstrated varying levels of success depending on the prompt. For example: the prototype often struggled to correctly read players' gesture-based response to the question "how old are you?" Some children struggled to show the correct number of fingers to represent their age—one player had to ask his mom how to show his age with his fingers. Players were far more successful, however, in their use of the "thumbs up" and "thumbs down" gestures. There was one instance where a player held two "thumbs up" to indicate her agreement that the baby bird was flying. Players were also successful in using the binocular gesture to select that observation tool. Nobody tried the gesture for flashlight because they'd already realized it wasn't the correct tool to help them take a closer look at the baby bird. "Flap" and "clap" sound similar and, perhaps because of this, a few children clapped, rather than flapping their arms when prompted to help show the baby bird how to fly.

Timing and turn-taking in voice mode

While the concept of responding verbally came naturally to most players, there were a few instances where players began to answer before Elinor had finished asking a question. Likewise, there were also instances where players were not done answering when Elinor started to speak again. For example, it sometimes took children a while to think of someone they wanted to share their observations with. It was also common for participants to nod (or shake) their heads or say "yeah" or "um hum" instead of "yes." Based on conversation transcripts, it seems that the game was recognizing some instances where a player had said "yeah" rather than "yes." Parents sometimes had to remind their child to talk out loud, to talk a little louder, and/or repeat themselves. This was especially true in cases where participants opted to leave their masks on for all or a part of the playtesting sessions—and we noted that it was easier for voice inputs to be recognized when players weren't wearing a mask. However, even in instances where players had to try a few times to get Elinor to hear them, they were generally patient and ultimately seemed to be successful overall.

Switching from one input type to another

Players would often continue using inputs from the first mini-game they had played as they were playing the second mini-game. For example, several players continued to use the "thumbs up" gesture during the spider game if they'd played the bird game first, whereas some continued to answer questions verbally

in the bird game if they'd played the spider game first. There were a few instances in both mini games where players thought they could respond with their answer, but the narration continued for a bit longer so their responses had to be repeated (Rockman, 2020).

Learning in the "Elinor Talks" prototype
Many players seemed pick up facts about birds and spiders as they were playing. After playing the spider-themed conversation, a few players stated that they had learned that the hair on spiders' legs kept them from sticking to a web. A few were able to recall, correctly, that a spider had 8 legs, but some recalled incorrect numbers, e.g., 9 or 10. This could have been a result of the counting lag, and the fact that some children continued to count on their own as they were tapping, rather than counting along with Elinor.

Players understood that birds had to be older before they could fly—a fact which they likely knew before playing this game. A few noted that birds fly by flapping their wings and a couple recalled that flying would ultimately help the baby bird be able to eat, e.g., "flying to get the worms".

Future Work

Conversational agents have advanced in their ability to understand and respond to users [3] as the PBS KIDS team has observed in the "Elinor Talks" prototype research and development work described in this paper. While CUI is still in its infancy, with many usability challenges for the general population of young children, we believe in its potential to support the needs of learners with varied abilities. To an even greater extent, gestural recognition technologies and how young children interact with them in the context of social communication is an even more open area for exploration.

Future research and development goals for this prototype include: (1) testing with children who have diagnosed communication disorders (in particular expressive language disorders) and deaf children and children who are hard of hearing (2) refining, based on improvements in the technology tools since 2020, and redesigning the prototype based on findings from testing that includes children from identified special populations; (3) testing the improved prototype with a mix of children from the general population and children with more specific criteria (communication disorders and deaf); and (4) developing design guidelines for voice and gesture social communication user experiences so that the varied needs of children are planned for and supported.

References

[1] Bandura, A., & McClelland, D. C. (1977). Social learning theory (Vol. 1). Prentice Hall: Englewood cliffs.

[2] Center for Applied Special Technology (CAST), "The UDL Guidelines." *UDL*, 6 Oct. 2020, udlguidelines.cast.org/.

[3] Kennedy, James, et al. "Child Speech Recognition in Human-Robot Interaction." *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 2017, doi:10.1145/2909824.3020229

[4] Richards, M. N., & Calvert, S. L. (2016). Media characters, Parasocial relationships, and the social aspects OF Children's learning across media platforms. *Media Exposure During Infancy and*

Early Childhood, 141–163. https://doi.org/10.1007/978-3-319-45102-2_9

[5] Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child development*, *85*(3), 956-970.