Explorations of Voice User Interfaces for 3 to 4 Year Old Children

Luiza Superti Pantoja
University of Iowa
Department of Informatics
luiza-supertipantoja@uiowa.edu

Kyle Diederich
University of Iowa
Department of Computer Science
kyle-diederich@uiowa.edu

Liam Crawford
University of Iowa
Department of Computer Science
liam-crawford@uiowa.edu

Juan Pablo Hourcade
University of Iowa
Department of Computer Science
juanpablo-hourcade@uiowa.edu

ABSTRACT

The design of Voice User Interfaces (VUIs) has mostly focused on applications for adults, but VUIs provide potential advantages to young children in enabling concurrent interactions with the physical and social world. Current applications for young children focus on media playing, answering questions, and highly-structured activities. There is an opportunity to go beyond these applications by using VUIs to support high-quality, creative social play. In this paper, we describe our first step in pursuing this opportunity with 24 design sessions guided by a partnership with eight 3 to 4 year old children. In a social play setting, we learned that children wanted to physically interact with the voice agents and VUIs could redirect behaviors and promote social interactions.

KEYWORDS

Children; Preschool; Play; Voice User Interfaces; Tangible User Interfaces; Communication

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK.

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3312802

1 INTRODUCTION

Voice user interfaces (VUIs) are becoming more relevant and present in families' homes [12], providing opportunities for children to interact with computers at early ages. Children under the age of five could potentially use VUIs in a developmentally beneficial way by interacting with technology while still engaging with the social and physical world [2]. In addition, VUIs can provide an alternative way to interact with computers for children who cannot yet read or write.

So far, young children's interactions with VUIs were typically studied in the context of understanding communication breakdowns, controlling media, asking factual questions, pursuing highly-structured activities, or understanding the perception of VUIs' personal qualities. There is an opportunity to explore other contexts such as supporting high-quality social play, which has been associated with multiple positive outcomes [1]. To begin filling this gap, we conducted 24 design sessions with eight 3 to 4 year old children using VUIs to facilitate high-quality social play activities. Our partnership with children guided the exploration and led us to investigate making voice agents tangible and enabling children to control what voice agents say. In our analysis of these design sessions we found that children wanted to physically interact with voice agents and that the agents could redirect children's behaviors and promote interactions with their peers.

2 RELATED WORK

2.1 High-quality Social Play and Voice Agents for Young Children

No research to date has explored how VUIs may support high-quality social play. This type of activity typically involves groups of children engaged in pretend play that includes common goals, planning, role-play, interactive social dialogue and negotiation, improvisation, and the use of generic physical props as opposed to realistic toys [2]. Several studies have identified the positive short and long-term impact of this type of play, including enhanced self-regulation and executive functions (e.g., [1]). Therefore, supporting high-quality social play can help to develop these skills, which lead to improvements in mathematical ability, reading, emergent literacy and vocabulary, theory of mind, and creativity [8]. A significant portion of American children could benefit from this type of activity since there is indications that many children arrive in Kindergarten without the self-regulation and executive function skills needed to succeed in school [9].

There are opportunities for interactive technologies to lower barriers to young children's high-quality play by scaffolding such activities. VUIs could play a role by integrating with physical, social play, without requiring the visual and motor engagement necessary to use screen-based apps. In spite of VUIs' increasing popularity [12], current applications for children include a wide variety of structured activities (e.g., [4]) that often struggle to recognize young children's speech and intentions [4]. Findings in prior research relevant to our work include the use of fantasy, curiosity, and agents' self-disclosure to keep children engaged (e.g., [3]), the advantages of physical representations of agents [10], using a combination of concrete directives intermingled with compliments to manage activities [10], avoiding the use of unexpected knowledge [7], and voice agents being aware of context and able to converse [11].

Table 1: Outline of Design Sessions

Session number	*Condition*
1 - 3, 6	Warm-up sessions intended for children to get used to high-quality social play and working with our team of researchers (no VUIs).
4, 5, 7, 8	Researcher controlled static voice agent (see Figure 2).
9 - 14	Researcher controlled portable, tangible voice agent (see Figure 3).
15 – 17, 19 - 22	Tablet app for child/researcher to control portable, tangible voice agent (see Figure 4).
18	"Turned-off" portable, tangible agent.
23, 24	Tablet app for child/researcher to control screen-based, animated agent (see Figure 5).

3 RESEARCH GOALS

Our research goal was to explore voice agents to support high-quality social play in the style of the well-defined, Vygotsky-inspired, *Tools of the Mind (ToM)* curriculum [2], for which there is substantial evidence of positive impacts on children's self-regulation and academic performance [1]. In our previous experience facilitating *ToM* play, we identified a challenge in keeping children socially engaged in play. Therefore, we wanted to explore using voice agents to scaffold the *ToM* play activities and increase social engagement in play.

4 METHOD

4.1 Participants

We recruited eight children (4M, 4F) from a preschool located in a city with a population of about 100,000 people in the United States. The childcare center serves mostly middleclass families and all the participants used a mobile device at home. We obtained permission to conduct research from our institution's Human Subjects Office and obtained consents from all parents. Children only participated in activities if they wanted to.

4.2 Research Activities

Our research activities adapted participatory design methods developed with elementary school children [5] to work with children under the age of five by enabling them to contribute ideas both verbally and by acting out their experiences. We conducted 24 design sessions at the children's small preschool (one classroom per age level), as described in **Table 1**, video recording every session. All research team members had prior experience facilitating 11 sessions of play in the style of *ToM* with another group of 3 to 4 year old children. In addition to the children, two to four research team members and one teacher were always present in the room. In the first three sessions, we worked with all the children together. In the remaining sessions, we worked with no more than four children at a time. After completing a session, the adult members of the design team met to debrief, to note any lessons learned, and to decide on the next directions for the research activities.

We leveraged an existing app designed to support *ToM* style play [6] that introduces children to stories and characters to provide a common context for play (see Figure 1). The same app included a play planner that enabled children to plan play by selecting the character they wanted to role-play [6], an activity encouraged by the *ToM* curriculum [2]. We always presented the app on a tablet (either a *Microsoft Surface Pro 4*, or an *iPad 4th generation*). After making use of the app to set up play, which took two to five minutes, the children proceeded to engage in play using generic physical props, as recommended by the *ToM* curriculum [2]. This portion of the design sessions typically took about 15 minutes. Our exploration of VUIs occurred in this portion of the activities, as well as all the observations we discuss in this paper.

Figure 1: Screen capture from Space Explorers story.

Figure 2: Static versions of miniBot, miniCat and miniBear.

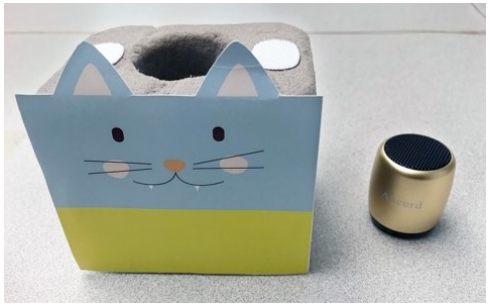

Figure 3: Portable, tangible voice agent with open top and Bluetooth speaker on the side.

As described in **Table 1**, we explored a variety of configurations for voice agents including researcher-controlled-speech agents where researchers typed text to control what voice agents said (a static and a portable version, see Figures 2 and 3), portable and screen-based agents with speech controlled through an app (see Figures 4 and 5) that could be used by children or researchers, and using a "turned-off" portable, tangible agent with no speech. We used the *Amazon Polly* Text-to-Speech service to generate all voice agent/character speech. Only one voice agent was active in any given design session.

4.3 Analysis

We conducted a qualitative analysis of our design sessions by coding 430 minutes of video data. Two researchers coded the videos identifying children's interactions with the voice agents. The Cohen's Kappa value of agreement for a randomly selected session was .849 (for 131 codes for different events). Four researchers transcribed all portions of the videos that included children's interactions with voice agents. Three researchers grouped the resulting 127 excerpts into themes using affinity diagramming and group discussion.

5 RESULTS AND DISCUSSION

5.1 Interactions with VUIs in the context of High-Quality Social Play

The introduction of VUIs in the social play environment impacted social interaction dynamics in positive and negative ways. Next, we highlight the major findings we observed in our explorations.

Children as mediators of voice agents. We observed children acting as mediators of the voice agent by repeating what the agent said to their peers. Potential reasons for this behavior include their peers not listening to or understanding the agent's comments, or that they wanted peers to collaborate or take some action regarding what the agent said.

Voice agents promoted peer interactions. The voice agents stimulated children to communicate and engage in social activities with their peers during play. Simple compliments or suggestions involving something a child was making were usually good avenues for promoting peer interactions (e.g. "miniCat: Cat, can you give a piece of cake to Horse and Monkey?").

Voice agents could redirect behavior. Voice agents became facilitators of positive, high-quality social play. When children listened to a voice agent, they tended to reply to prompts by either conversing with the agent or acting on its suggestions. Voice agents were thus a good avenue to redirect children's focus to participate in story-oriented play activities.

Children reacted best to a combination of task-oriented suggestions and positive reinforcements. We observed that when voice agents made suggestions or expressed compliments for behaviors that fit high-quality social play, these tended to promote positive interactions with the voice agent as well as positive play outcomes. On the other hand, authoritarian comments (e.g., "I want food now") caused surprise in children and made them complain to the researchers about the agent (e.g., "He is not being nice").

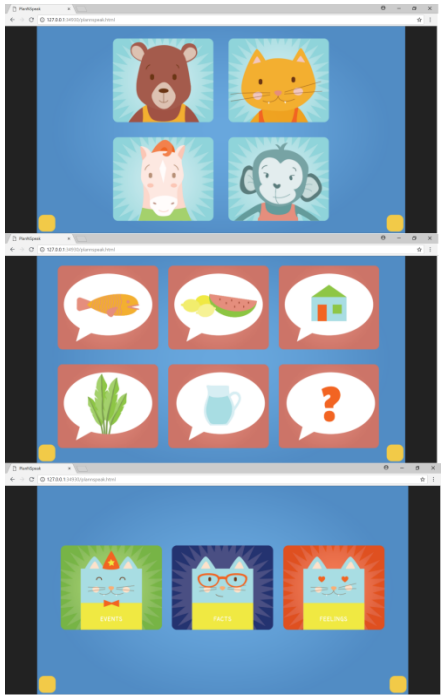

Figure 4: Tablet app to control voice agent enabling selection of who the speech is directed to topic, and feeling, fact, or event.

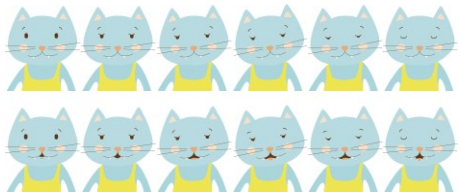

Figure 5: Sprite image of the animated agent.

The voice agents' lack of context frustrated children when using the tablet app to control speech. When we switched the control of the voice agents from researchers to the first version of the tablet app (see Figure 4), children noticed the difference calling the voice agents "weird", getting frustrated, and even shouting at the voice agents. The challenges had to do mainly with the tablet app generating random speech related to a story, which lacked the contextual information that researchers had in the previous sessions.

The tablet app to control speech could distract children from high-quality social play. The tablet app competed for children's attention and distracted them from play with their peers. For instance, in one of the sessions a child was only interested in interacting with the tablet app and that influenced the quality of the social play.

Children augmented interactions with the voice agents by using physical props. Throughout both the static and portable, tangible voice agent design sessions, children frequently incorporated physical props into the voice agents. For example, they used hats to cover the voice agents, they used blocks as hats, and placed the voice agents inside their constructions. Children also used blocks as food and beverages to pretend they were feeding the voice agents.

Children were curious about how the voice agents worked. During sessions with researcher-controlled-speech agents, children expressed curiosity about how the voice agents were speaking. Once they discovered that researchers were controlling the speech, they continued interacting with the voice agents with the same level of interest and engagement.

Children's stereotypes affected their interactions with the characters depicted by the voice agents. The characters in the stories children experienced depicted gender-neutral animals or robots (see Figures 1 and 2), all had similar levels of importance and differed only in their unique ability. However, we observed great affection from both girls and boys toward miniCat, with behaviors such as petting the tangible agent, verbally expressing their love (e.g., "I love you, miniCat"), and holding it gently.

5.2 Opportunities and Ethical Concerns

An interesting finding from our explorations was that young children continued to engage with the voice agents as if they were autonomous in spite of knowing adults were controlling the speech. This highlights an opportunity for exploring VUIs as an alternative way for adults to communicate with children using speech synthesis.

Moreover, children clearly favored interacting with a portable, tangible voice agent over a screen-based one, or a physical representation they could not pick up and incorporate in their play. We also discovered that a minimal physical representation of a character was good enough for young children to have an interest in engaging with it. It appears that there is no need of complex fidelity or complex materials. At the same time, more advanced setups could include sensors or props to understand context, and more actuators on tangible voice agents to engage curiosity.

On the other hand, one ethical concern related to VUIs and young children is the degree to which children will follow the voice agent's instructions. During our design sessions, we observed children following suggestions or performing tasks directed by the agent. This finding might raise

ACKNOWLEDGMENTS
We would like to thank the teachers and staff at the preschool that hosted our research activities, as well as the children who participated. Luiza Superti Pantoja's work was funded by a fellowship from the CAPES Foundation (99999.013579/2013-04).

concerns about who controls the agent and how children may be influenced by voice agents. Therefore, it is important to consider the potential risks of unsupervised young children interacting with VUIs.

6 CONCLUSIONS

This paper described an exploration of the design of VUIs in the context of high-quality social play. Our partnership with 3 to 4 year old children guided the research directions throughout 24 design sessions. The findings from our explorations demonstrate the potential of voice agents for facilitating creative, social, lightly-structured activities.

REFERENCES

[1] Clancy Blair and C. Cybele Raver. 2014. Closing the Achievement Gap through Modification of Neurocognitive and Neuroendocrine Function: Results from a Cluster Randomized Controlled Trial of an Innovative Approach to the Education of Children in Kindergarten. *PLoS ONE* 9, 11: e112393. https://doi.org/10.1371/journal.pone.0112393

[2] Elena Bodrova and Deborah J.. Leong. 2007. *Tools of the mind*. Pearson.

[3] Franziska Burger, Joost Broekens, and Mark A Neerincx. 2016. Fostering relatedness between children and virtual agents through reciprocal self-disclosure. In *Benelux Conference on Artificial Intelligence*, 137–154.

[4] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. Hey Google is it OK if I eat you?: Initial Explorations in Child-Agent Interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children*, 595–600.

[5] Mona Leigh Guha, Allison Druin, and Jerry Alan Fails. 2013. Cooperative Inquiry revisited: Reflections of the past and guidelines for the future of intergenerational co-design. *International Journal of Child-Computer Interaction* 1, 1: 14–23. https://doi.org/10.1016/j.ijcci.2012.08.003

[6] Juan Pablo Hourcade, Luiza Superti Pantoja, Kyle Diederich, Liam Crawford, and Glenda Revelle. 2017. The 3Cs for preschool children's technology: create, connect, communicate. *interactions* 24, 4: 70–73. https://doi.org/10.1145/3096461

[7] Iolanda Leite and Jill Fain Lehman. 2016. The Robot Who Knew Too Much: Toward Understanding the Privacy/Personalization Trade-Off in Child-Robot Conversation. In *Proceedings of the The 15th International Conference on Interaction Design and Children* (IDC '16), 379–387. https://doi.org/10.1145/2930674.2930687

[8] Megan M. McClelland, Claire E. Cameron, Carol McDonald Connor, Carrie L. Farris, Abigail M. Jewkes, and Frederick J. Morrison. 2007. Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology* 43, 4: 947–959. https://doi.org/10.1037/0012-1649.43.4.947

[9] Sara E. Rimm-Kaufman, Tim W. Curby, Kevin J. Grimm, Lori Nathanson, and Laura L. Brock. 2009. The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology* 45, 4: 958–972. https://doi.org/10.1037/a0015861

[10] Erik Strommen. 1998. When the interface is a talking dinosaur: learning across media with ActiMates Barney. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 288–295.

[11] Julia Woodward, Zari McFadden, Nicole Shiver, Amir Ben-hayon, Jason C Yip, and Lisa Anthony. 2018. Using Co-Design to Examine How Children Conceptualize Intelligent Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 575.

[12] Svetlana Yarosh, Stryker Thompson, Kathleen Watson, Alice Chase, Ashwin Senthilkumar, Ye Yuan, and A. J. Bernheim Brush. 2018. Children asking questions: speech interface reformulations and personification preferences. In *Proceedings of the 17th ACM Conference on Interaction Design and Children - IDC '18*, 300–312. https://doi.org/10.1145/3202185.3202207