Voice Agents Supporting High-Quality Social Play

Luiza Superti Pantoja
Informatics
University of Iowa
Iowa City, IA, USA
luiza-supertipantoja@uiowa.edu

Kyle Diederich
Computer Science
University of Iowa
Iowa City, IA, USA
kyle-diederich@uiowa.edu

Liam Crawford
Computer Science
University of Iowa
Iowa City, IA, USA
liam-crawford@uiowa.edu

Juan Pablo Hourcade
Computer Science
University of Iowa
Iowa City, IA, USA
juanpablo-hourcade@uiowa.edu

ABSTRACT

While the design of Voice User Interfaces (VUIs) has mostly focused on applications for adults, VUIs also provide potential advantages for young children in enabling concurrent interactions with the physical and social world. Current applications for young children focus mostly on media playing, answering questions, and highly-structured activities. There is an opportunity to go beyond these applications by using VUIs to support less structured, developmentally appropriate activities. In this paper, we describe our first step in pursuing this opportunity through an exploration of voice agents to facilitate high-quality social play guided by a partnership with eight 3-4 year old children. During 24 design sessions, we explored making voice agents tangible and enabling children to control what voice agents say. After analyzing the sessions, we learned voice agents could help keep children socially engaged in play and children liked incorporating the agents with the physical aspects of their play. On the other hand, enabling children to control the voice agents caused distractions from play.

CCS CONCEPTS

CCS → Human-centered computing → Human computer interaction (HCI) → Interaction paradigms → Natural language interfaces

KEYWORDS

Children; preschool; voice user interfaces; voice agents; tangible user interfaces; communication; play;

ACM Reference format:

Luiza Superti Pantoja, Kyle Diederich, Liam Crawford and Juan Pablo Hourcade. 2019. Voice Agents Supporting High-Quality Social Play. In *Proceedings of ACM Interaction Design and Children conference (IDC'19). ACM, Boise, ID, USA, 10 pages.* https://doi.org/10.1145/3311927.3323151

IDC'19, June 12-15, 2019, Boise, ID, USA

ACM 978-1-4503-6690-8/19/06…$15.00
https://doi.org/10.1145/3311927.3323151

1 INTRODUCTION

Voice user interfaces (VUIs) have been increasing their prominence in the past few years. *Alexa* [70], *Siri* [71], *Cortana* [72], and *Google Assistant* [73] are prominent in advertising and appear critical to all major software companies' strategies for engaging with users. Television and online commercials show these systems enabling users to interact with computers in situations where it would be inconvenient or difficult to do so using a mobile device or traditional computer [1]. Cooking, exercising, or taking a bath are no longer a barrier to getting a recipe, making music selections, or getting a weather forecast.

VUIs also offer something new to children under the age of five: a way to interact with computers that may be compatible with developmentally beneficial engagement with the social and physical world [8,14,56,66]. This is a significant departure from interaction with mobile devices, which typically require full visual and manual engagement. VUIs could therefore alleviate some of the concerns about interactive devices socially isolating children [65] and keeping them from interacting with the physical world around them. In addition, since VUIs are ideal for avoiding typing and reading, they can provide advantages to children who cannot yet read or write. Amazon, for example, already offers the Echo Dot with Alexa for kids, which promises to let children play media (e.g., music, stories) and ask factual questions [49,74].

Researchers have mostly studied VUIs with children under the age of five in the context of understanding communication breakdowns, controlling media, asking factual questions, pursuing highly-structured activities, or understanding the perception of VUIs' personal qualities, such as intelligence. There is an opportunity to begin exploring less structured contexts, such as supporting high-quality social play, which has been associated with multiple positive outcomes [8,10,53].

To begin filling this gap, we conducted 24 sessions with eight children to explore the wider opportunities of VUIs for 3-4 year old children focusing on facilitating 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 this paper, we contribute a qualitative description of our explorations, which provides findings related to voice agents supporting social play and design recommendations for future VUI applications for this age group. Our analysis suggests that context-aware,

tangible, portable voice agents may help keep children socially engaged in play and that children like to integrate voice agents in the physical aspects of their play. We also identify challenges associated with children's interest in controlling voice agents and provide other useful findings (e.g., need to slow down speech synthesis) for future VUI design for young children.

2 RELATED WORK

Below, we first discuss the characteristics of high-quality social play for children under the age of five. We end this section with a discussion of research on VUIs for children and other related technologies such as interactive dolls and robots.

2.1 High-Quality Social Play

Ages 3 and 4 are a crucial time for development, in particular for building the foundations of executive function skills (e.g., selective attention, planning, cognitive flexibility) through self-regulation [43]. They are also a time for the development of curiosity, creativity, imagination, social play, cooperation, language and communication, and storytelling [23]. An important type of activity for this age group that, to our knowledge, VUIs do not yet support, is high-quality social play. High-quality social play 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 [14,60,64,66]. Several studies have identified the positive short and long-term impact of this type of play, including enhanced self-regulation and executive functions [10,11,27,28,52], which in turn lead to improvements in mathematical ability [13,18,26,51], reading, emergent literacy and vocabulary [13,51], theory of mind [21], and creativity [8,53].

A well-defined approach to high-quality social play, which we use in the research described in this paper comes from the *Tools of the Mind* (*ToM*) curriculum [14]. Multiple large studies provide evidence of this approach's positive impact on children's executive function skills and academic achievement [5,10,27]. *ToM* draws its inspiration from Vygotsky's ideas, as well as those of his students, Elkonin and Leont'ev, on the role of social interaction, including social play, and external tools in child development [14]. More specifically, Vygotsky's views on development emphasize that children's development of skills and concepts occurs first socially (with help from others) and then individually [66]. In line with Vygotsky's views, *ToM* makes a strong emphasis on teachers scaffolding play activities and children purposefully collaborating with peers (e.g., planning play activities) [14].

With respect to make-believe play, Vygotsky's observation was that it leads to children regulating their behavior, in particular inhibiting behavior that does not fit the make-believe context [14,66]. This self-regulation starts with physical behaviors, is followed by social behaviors, and then by cognitive processes such as attention [14]. Vygotsky also inspired another aspect of *ToM* play, which is the use of generic props [14]. The use of generic props such as basic shapes made of soft materials, enables children to use and reuse the props to represent different objects based on the necessities of play [14]. Vygotsky hypothesized that playing in this manner can help children develop abstract thought [66]. The main challenge with adopting the *ToM* approach is that it requires a significant amount of teacher training, with a recent study including five days of instruction in addition to in-class coaching sessions every other week [10].

There are opportunities for interactive technologies to lower barriers to young children's high-quality social 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 this paper, we begin to address this gap with an exploration of VUIs to facilitate high-quality social play in the style of *ToM* through an extensive partnership with 3 to 4 year old children.

2.2 VUIs for Children

Wide use of VUIs as personal assistants has only become a reality in the 2010s [7,16], with research on VUIs for young children also concentrated in this decade with a significant increase in publications since 2017. Recurring research topics include the use of commercial systems [48,59], communication challenges [24,69], as well as systems specifically designed for children [2,42,44,46]. The findings suggest that children typically use commercial systems to explore interactions, seek information, or make requests (e.g., for media to be played) [29,48]. However, these interactions were usually marred by poor speech recognition [29,48,59]. Children's difficulty communicating with VUIs was the topic of recent investigations with 3-5 year old children [24], and 5-12 year old children [69].

Findings in prior research relevant to our work include: 1) the use of fantasy, curiosity, and self-disclosure by voice agents to keep children engaged [19,30,42], 2) the advantages of physical representations of agents in order to evoke basic social expectations [22,31,61], 3) using a combination of concrete directives intermingled with compliments to manage activities [61], 4) avoiding the use of unexpected knowledge [45], 5) remembering prior interactions [46], and 6) agents being aware of context and able to converse [68]. Researchers have also raised concerns about voice agent technology both in terms of privacy [37,39,40] and the possibility of these technologies negatively influencing children's behavior and values [67].

In our investigation, we contribute an exploration of voice agents to support a lightly-structured activity incorporating the use of fantasy, while exploring a variety of design choices and configurations (e.g., tangible, screen-based, various levels of contextual awareness) for voice agents to support *ToM* style play.

3 RESEARCH GOALS

Our research goal was to explore the design of voice agents to support high-quality social play in the style of *ToM* [10,14]. 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 augment what adult facilitators can do to scaffold *ToM* play activities, learn about each design's impact on key *ToM* play components that arise from Vygotsky's theories, such as social engagement in role-play and the use of physical props, and understand the characteristics of children's interactions with the agents.

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. All the participants used a mobile device at home and their favorite apps reported by parents were: Osmo, PBS Kids, Amazon prime, Netflix, YouTube Kids, and varied games. The average number of children in their households was 1.77. We obtained permission to conduct research from our institution's Human Subjects Office and obtained consents from all parents. Children only participated in a given session if they wanted to.

4.2 Research Activities

Our research activities adapted participatory design methods developed with elementary school children [34] 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 3-4 year old children. In addition to the children, two to four research team members and one teacher were always present in the room where we conducted the sessions.

After making use of the app to set up play, which took 2 to 5 minutes, the children proceeded to engage in play using generic physical props, as recommended by the *ToM* curriculum [14]. This portion of the design sessions typically took about 15 minutes. Our exploration of voice agents occurred in this portion of the activities, as well as all the observations we discuss in this paper. As described in Table 1, we explored a variety of configurations for voice agents, led by children's suggestions, including researcher-controlled-speech agents where researchers typed text to control what voice agents said (a static and a portable version, see Figure 2), portable and screen-based agents with speech controlled through an app (see Figure 3) that could be used by children or researchers, and using a "turned-off" portable, tangible agent with no speech. Only one voice agent was active in any given design session.

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 [38] 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 [38], an activity encouraged by the *ToM* curriculum [14]. We always presented the app on a tablet. Below we provide a detailed description of each configuration:

Table 1. Outline of Design Sessions.

Session Number	Activities
1-3, 6	Warm-up sessions intended for children to get used to *ToM* style play and working with our team of researchers (no voice agents).
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 2).
18	"Turned-off" portable, tangible agent.
23, 24	Tablet app for child/researcher to control screen-based, animated agent (see Figure 3).

Figure 1: Screen capture from Space Explorers story.

4.2.1 Researcher controlled static, voice agent. Our initial exploration of voice agents involved a setup where researchers typed text to control what a voice agent said with the purpose of encouraging children to stay engaged in *ToM* style play. Our first static iteration consisted of a small Bluetooth speaker inside a paper box that looked like a character (the voice agent characters looked like small versions of the characters in the stories, see Figure 2). Because the paper box representing the voice agent was too delicate to pick up, we told children they could talk to it and touch it, but not move it, similar to the way they would interact with a device like the *Amazon Echo Dot*.

Figure 2: On the left, static versions of voice agents we called miniBear, miniCat and miniBot. On the right, the portable tangible voice agent with closed top and Bluetooth speaker on the front.

4.2.2 Researcher controlled portable, tangible voice agent. During the use of the static version of the voice agent, it became clear through children's actions and requests that they wanted an agent they could pick up. Hence, we replaced the paper boxes with foam blocks. We printed the face and top of the character on thick paper (see Figure 2) and attached it to the foam block with Velcro. The top portion secured the Bluetooth speaker, which fit inside a carved space in the foam block. We continued using the same physical setup for the voice agents in our remaining sessions, except for the last two where we explored a screen-based representation.

4.2.3 Tablet app for child and researcher to control portable, tangible voice agent. When children realized that we were controlling the voice agents, some insisted on wanting to control the agents' speech. We then iteratively developed a tablet-based speech-control app. The initial app prototype allowed participants to select random comments from a set of four categories: 1) new events to expand play, 2) feelings about the play environment, 3) facts about the play environment, 4) reactions to play. The final version of the app generated speech based on three sets of choices: to whom the speech was directed (from the set of characters being played by children), a subject related to the story on which the play was based (e.g., food, drink, nature), and a theme (e.g. events, facts, feelings) (see Figure 3). It was set up with the goal of having the voice agent say something relevant to the play activity that could encourage continued engagement. Making the same selections multiple times yielded different speech, as the voice agent would, for example, speak about a different type of food and express a different feeling about it. Both children (after brief instruction and demonstration from researchers) and researchers used the app to control the voice agents.

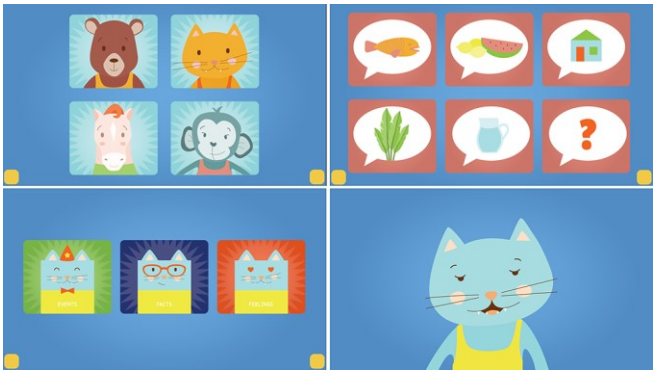


Figure 3: Tablet app to control voice agent enabling selection of to whom the speech is directed, the topic, and feeling, fact, or event. The last screen shows the image of the animated agent used in the 4.2.5 setup.

4.2.4 "Turned-off" portable, tangible agent. Given the limitations of a tablet app to control speech through a small set of options, we decided to explore "turning off" the voice agent and encouraging children to use their own voices instead to make the voice agents speak.

4.2.5 Tablet app for child/researcher to control screen-based, animated agent. To understand the impact of the tangible aspects of voice agents, we also decided to conduct sessions where the representation of the agent was on the screen. We used sprite animations for the screen-based agent and incorporated them into the existing tablet app to control agent speech described in section 4.2.3. The animated agent appeared on the screen speaking after children made speech choices. In both sessions, the facilitator held the device for children and supported children taking turns making the agent speak.

4.3 Technical Setup

The technical components to our system included: 1) an *Ancord Micro* Bluetooth speaker (shown in Figure 2), 2) two laptop computers (to run the researcher-controlled system), 3) two tablets to run the apps (*Microsoft Surface Pro 4*, or an *iPad 4th generation*). We used the *Amazon Polly* Text-to-Speech service to generate all voice agent/character speech [75], which allowed us to create different voices and personalities for all characters. Additionally, *Amazon Polly* supports *Speech Synthesis Markup Language* (SSML) [76] so we could fine-tune the generated speech using breaks, emphasis, prosody, and so forth.

4.4 Analysis

We conducted a qualitative analysis of our design sessions by coding 430 minutes of video data using *BORIS* [33]. Two researchers coded the videos identifying children's interactions with the voice agents and the periods of time when individual children were not engaged in play. Interactions with voice agents included children conversing, reacting to, or physically manipulating the voice agents. Non-engagement in play included children interacting with mobile devices instead of playing with their peers or with the physical props or getting distracted from

the play activity (e.g. using props out of the play context). 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 the coding identified as including children's interactions with voice agents. Three researchers grouped the resulting 127 excerpts into themes using affinity diagramming and group discussions, focusing on children's social engagement in play and the use of physical props, both key aspects of *ToM* play [14], as well as characterizing children's interactions with the agents in an activity that is less structured than those studied in the past.

5 FINDINGS

Below, we present the themes identified through our analysis organized in three areas: social engagement in play, use of physical props, and interactions with agents. We present each theme with respective subthemes illustrated through corresponding excerpts (all names in each excerpt are pseudonyms), followed by related design recommendations including discussions of links to prior research.

5.1 Social Engagement in *ToM* Style Play

The introduction of voice agents in the social play environment impacted social interaction dynamics, with varying outcomes depending on how we structured activities.

5.1.1 Voice agents promoted peer interactions. We observed situations where the voice agents stimulated children to communicate and engage in social activities with their peers. Simple compliments or suggestions involving something a child was making were usually good avenues for promoting peer interactions, such as the one portrayed below from session 6:

Agent miniCat: *Cat, can you give a piece of cake to Horse and Monkey?*
Dora gives a prop to Eduardo, who is playing Monkey.
Eduardo pretends to eat the prop and then puts it aside.
Dora holds a prop.
Dora: *Here is my slice.*
Agent miniCat: *Monkey, did you like the cake?*
Dora goes toward Eduardo.
Dora: *Did you like the cake, Monkey?*
Eduardo: *Yes.*
Then, Dora turns back to agent miniCat.
Dora: *He said yes.*

5.1.2 Voice agents could redirect behavior to keep children socially engaged in ToM style 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. In the following example from session 7, agent miniCat redirected children to re-engage with a shelter that was part of the play storyline:

Ahmed places a prop on agent miniCat's head.
Ahmed: *Do you like your hat?*
Agent miniCat: *That's too big.*
Jessica: *That's too big. He doesn't like it.*
Agent miniCat: *What about the shelter?*
Eduardo goes to the floor, arranges some props, and sits down.
Eduardo: *I am in the shelter!*
Ahmed looks at his peer and follows him by sitting down too.

5.1.3 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 *ToM* style 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"). Here is a positive example that happened on session 10:

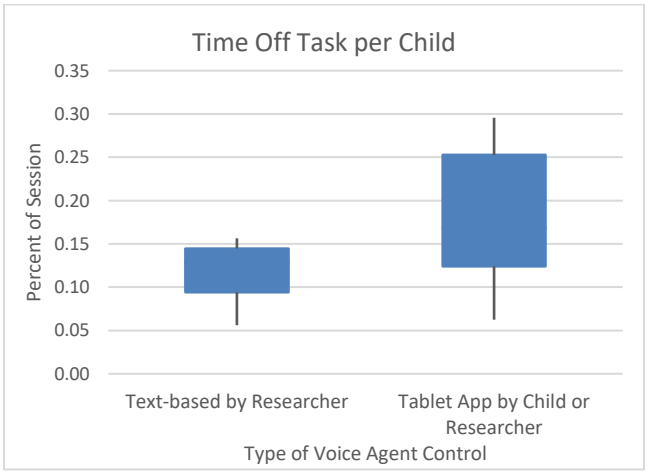
Agent miniBot: *Bear, could you get me a drink?*
Michael: *He said Bear give me a drink.*
Michael gets a prop.
Michael: *Here is your apple juice.*
Agent miniBot: *Yummy.*
Michael approaches agent miniBot.
Michael: *What did you say?*
Alice: *I think he said yummy.*
Agent miniBot: *Yummy.*
Michael pretends to give juice to agent miniBot.
Agent miniBot: *Yummy. Yummy. Yummy.*
Michael: *Here is some milk!*
Agent miniBot: *Sorry, I broke!*
Agent miniBot: *Help me monkey!*
Michael: *He said, 'Help me Monkey.' Who is the monkey?*
Alice: *Okay!*

5.1.4 Children as mediators of voice agents. We observed children acting as mediators of voice agents, by repeating what agents said to their peers. It is another way in which the voice agents promoted social interactions. 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. Here are excerpts from sessions 6 and 9 illustrating this behavior:

Agent miniCat: *You broke the shelter.*
Maggie approaches agent miniCat.
Maggie: *What?*
Agent miniCat: *You broke the shelter.*
Maggie goes toward the boys.
Maggie: *She says that you broke the shelter.*

Agent miniCat: *Looks great! Good job, Horse.*
Michael ignores agent miniCat.
Dora: *He said, good job Horse.*
Michael smiles and brings a prop to put on agent miniCat's head. The kids then place hats on agent miniCat's head.
Michael: *Do you like your new hat?*

5.1.5 The tablet app to control speech could distract children from ToM style play activities. While children took ownership over their control of the speech-control tablet app, the mobile device competed for children's attention and distracted them from playing with their peers. Children issued speech control commands through the tablet app an average of 1.96 times per minute during sessions 15-17 and 19-24, and our coding confirms that they typically spent more time off task in the sessions where they had access to the app to control speech, than in the ones they did not (see Figure 4). There was also more variability in time off task for sessions where children had access to the app to control speech.



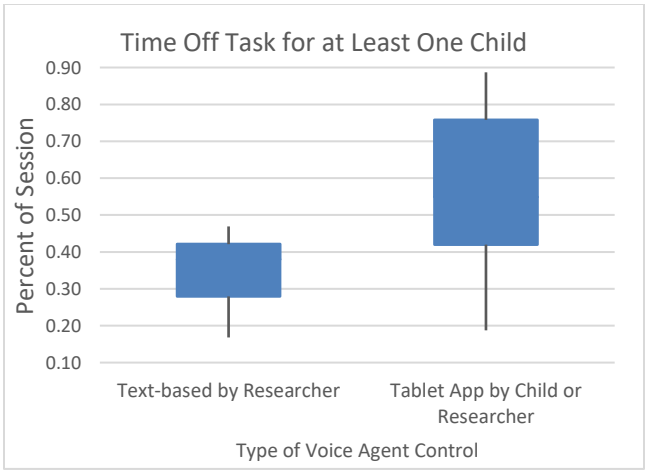


Figure 4: Box plots of children's time spent off task. Text-based voice agent control by researcher includes sessions 4-5 and 7-14, while tablet app by child or researcher includes sessions 15-17 and 19-22. Time off task for at least one child shows the percent of a session in which at least one child was off task. Time off task per child divides the sum of time off task for every child in a session over the number of children in a session (divided by session length to get percent of session).

Among the reasons we observed for time off task was children's strong interest in using the app to control voice agents and the resulting inability of researchers to use voice agents to reach out to children who were off task. Below is an example from session 24, where one child was only interested in interacting with the tablet app. That behavior influenced the quality of play:

> The researcher holds the device and Michael keeps making agent miniCat speak.
> Agent miniCat: *I heard a baby animal crying.*
> Researcher: *Did you guys hear that too?*
> Maggie: *Where is the baby animal?*
> Researcher: *Are you going to find the baby animal?*
> Maggie: *Jessica, do you want to go into the jungle and find the baby animal crying?*
> Jessica: *No. I will stay here and make a house for miniCat.*
> Maggie: *Michael, do you want to go into the jungle and find the baby animal?*
> Michael ignores Maggie and keeps interacting with the tablet app.
> Jessica keeps building with the physical props and Maggie joins her.

On the positive side, the excerpt below from session 21 shows a child taking ownership of his control of the voice agent while exploring the speech options of the interface:

> Michael interacts with the tablet.
> Agent miniCat: *Listen cat, let's build a boat using the trees.*
> Michael: *Dora, that's you, he just talked to you, because I made him to.*

5.1.6 Design recommendations. Designers can use voice agents to enhance young children's high-quality social play by promoting social interactions and redirecting activities toward social role-play. Voice agent interventions are likely to work best by making suggestions and providing positive reinforcement for behavior that fits high-quality social play, such as collaborating with a peer or re-engaging with a storyline. Previous research found that it was important to offer children a good mix between task-oriented speech and positive reinforcements [61], and this lesson still applied to the young children who participated in our research activities. In addition, considering children are more likely to use polite social exchanges with speech systems [16] voice agents should attempt to be polite in their interactions with young children (e.g., saying "please" and "thank you") and avoid authoritative comments. The children who participated in our explorations tended to respond well to compliments, task-oriented suggestions, and humor.

Likewise, designers should carefully consider the appropriateness of the use of tablet apps during high-quality social play. Previous research suggests that visual design, sound effects, and even the touchscreen interface can either engage or distract young children [57]. During our design sessions touchscreen interfaces distracted some children from interacting with their peers even when an adult was facilitating their interactions by using prompts and holding the device. However,

it is important to note that in other contexts, such as healthcare, distractions may be welcome [15].

5.2 Use of Physical Props

When we introduced voice agents to children, they were interested in physically interacting with them right away, which included the use of physical props. With the static version of the voice agents, children could physically bring props to them, but they could not move the agents. Once we made the voice agents portable and tangible, the type of interactions with physical props changed. Below, we outline some of the ways in which children interacted with the voice agents through physical props.

5.2.1 Children augmented interactions with the voice agents by using physical props. One of the characteristics of our design sessions was the physical props that children used to represent a wide variety of objects (e.g., fish, cake, a glass of water) and build whatever they could imagine (e.g., houses, caves, trees, spaceships). Throughout both the static and portable, tangible voice agent design sessions, we observed many instances when children incorporated physical props with the voice agents. For example, they used hats to cover them; or they used blocks as hats, as food, or as beverages, to pretend they were feeding the voice agents, as these excerpts from sessions 5 and 9 show:

> Agent miniCat: *I like that blue hat on you.*
> Dora: *Thank you.*
> Dora: *That's funny.*
> Eduardo covers agent miniCat with a black hat.
> Dora smiles.
> Dora: *Let's cover him up!*
> Agent miniCat: *It's dark in here.*
> The kids keep putting one hat on top of the other, and Eduardo approaches agent miniCat to check if he can still hear it.

> Maggie approaches agent miniCat.
> Maggie: *Hey, let's see if this hat fits you!*
> Maggie puts the hat on top of agent miniCat.
> Maggie: *Oh, it fits you!*
> Ahmed and Michael approach agent miniCat and both pick up props and pretend to feed agent miniCat.

Once we made the voice agents portable and tangible, the children started to place them inside of their constructions, or built shelters and cities around them as shown in the excerpt below extracted from session 20:

> Agent miniBear: *Horse, will you make something new for me?*
> Richard: *Something light for me?*
> Researcher: *Something new!*
> Richard: *Okay, miniBear, making something new. Are you sure miniBear?*
> Richard: *Okay you go right there…*
> Richard places agent miniBear down.
> Richard: *…and I'll build all around you.*

5.2.2 Comparison with screen-based animated agent. In the last two design sessions, we introduced children to the screen-based, animated version of our voice agent miniCat (see Figure 3). They responded with a mixture of surprise and disappointment for not having the tangible agent. Here is a sample reaction from session 23:

> Researcher: *You said you wanted to make miniCat say something. What do you want to make miniCat say?*
> Eduardo: *Where is miniCat?*
> Researcher: *miniCat is right here.*
> Researcher points at the tablet.
> Eduardo selects speech options on the tablet and the animated version of agent miniCat appears and speaks.
> Eduardo and Ahmed: *Whaaaaat?!*

5.2.3 Design recommendations. Young 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. Therefore, it is important to consider the tangible affordances of a physical representation of a voice agent for young children such that it can be part of physical play, including the use of physical props. This recommendation resonates with previous research findings suggesting that tangible interfaces were better at supporting children's active collaboration and more appropriate for younger children to refine their fine motor skills [3,6,36,62,63].

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. This outcome is similar to a previous study that examined 4-10 year-old children's interactions with conversational characters [41], finding there was no need for perfectly realistic-looking human characters to elicit natural behaviors from children. Another advantage of physical representations of agents is to evoke basic social expectations of face-to-face communication [22]. In our experience, even though our voice agents only had a static facial expression, young children were able to relate to them affectionately.

5.3 Interactions with Voice Agents

During our extensive explorations, there were other relevant aspects that arose from children's interactions with voice agents that may inform future design of voice agents for this age group beyond applications to high-quality social play. Below we outline the general findings related to young children's interactions with the voice agents.

5.3.1 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 Figure 1). All had similar levels of importance and differed only in their unique ability. The voice agent physical representations looked like mini versions of these characters (see Figure 2). Despite the similarities in the way in which we introduced children to story characters and voice agents, children treated voice agents differently depending on the character they were depicting. For example, 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 carefully.

Here are two examples of physical and verbal exchanges between miniCat, a girl, and a boy, from session 7:

> Maggie picks up a prop and places it on agent miniCat's head.
> Maggie: *Here's your hat.*
> Agent miniCat: *Nice hat!*
> Maggie pets agent miniCat once more and then removes the hat.
>
> Eduardo (who is playing the horse character) approaches agent miniCat.
> Agent miniCat: *Horse, you are so skilled.*
> Eduardo pets agent miniCat while smiling. Then, he gives a prop to agent miniCat.
> Eduardo: *There you go little guy! Here is a nice cake.*

Children were just as likely to interact physically with miniCat (mean=35% of the time over all sessions) and miniBot (mean=31% of the time over all sessions). They were also just as likely to interact verbally with the two voice agents (miniCat mean=9.1% of the time over all sessions; miniBot mean=10.0% of the time over all sessions). However, the quality of the interactions was different. Children demonstrated less warmth in their interactions with miniBot, which translated into a tendency to be rougher on the physical handling of the voice agent. Here is one example of physical and verbal interactions that happened between children and miniBot during session 8:

> Agent miniBot: *Hello friends, miniBot here to say hi.*
> All the kids turn to miniBot and approach it.
> Dora and Jessica: *Hi!*
> Ahmed waves at agent miniBot and Richard smacks agent miniBot with a prop.
> Dora places a prop on top of miniBot.
> miniBot: *Nice to meet you!*
> Dora: *Nice to meet you too.*
> Richard places a hat on top of agent miniBot and agent miniBot falls off the table.
> Researcher picks agent miniBot up.
> Researcher: *Be careful with miniBot.*
> Ahmed: *Richard, you are being really mean.*

5.3.2 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, which generated random speech related to a story, 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 lacking the contextual information that researchers had in the previous sessions. These challenges led us to provide more control in the tablet app (see Figure 3) over what the voice agents would say, solving some, but not all problems. Below are two examples of interactions lacking context from sessions 15 and 21, respectively:

> Alice puts a hat on agent miniCat.
> Alice interacts with the tablet app to control speech.
> Agent miniCat: *Hi, I am miniCat.*
> Alice lifts the hat to shout at agent miniCat.
> Alice: *I already know you are miniCat!*
> Maggie: *Wait, wait, wait.*
> Alice: *Let's cover him all up.*
> Dora interacts with the tablet app to control speech.
> Agent miniCat: *Did we get enough trees for building?*
> Dora: *Noooo!* [Exasperated]
>
> Agent miniCat: *Guess what bear, I'd like to eat some fresh fish.*
> Michael: *Okay…you already ate a fish!*
> Researcher: *Apparently, you've got to get more.*
> Michael: *Ugggh*

5.3.3 Children were curious about how the voice agents worked and who controlled them. 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 voice agents, they continued interacting with the voice agents with the same level of interest and engagement, with the only added difference being occasional requests for the agent to say something. As we moved forward with our sessions, children's desire to control the voice agent evolved into requests of specific phrases they wanted the voice agents to say. Here are excerpts from sessions 6 and 24 illustrating the kinds of speech that interested children:

> Ahmed looks at the researcher.
> Ahmed: *Hey, make it say, "I like cake!"*
> Agent miniCat: *I like cake.*
> Ahmed laughs aloud.
> Ahmed: *I love cake with a yolk.*
> All three children are close to agent miniCat smiling and laughing.
> Michael opens agent miniCat's top lid and shows Eduardo the speaker inside it. Then, he turns to researcher.
> Michael: *Can you make him say, "I hate cake?"*
> Agent miniCat: *I hate cake.*
> The kids laugh.
>
> Maggie: *I wish you would say, "I have a big fear of water."*

5.3.4 Children expected voice agents to say something at specific times. Children's initial requests mostly related to making the voice agents say something in a particular moment (e.g., when physically approaching the agent). On average, children made 0.33 requests per minute to control the voice agents' speech during sessions when researchers were controlling it. Here are two examples of children's requests for the voice agent to say something, extracted from session 4:

> The kids start building with props, and Michael covers agent miniCat with a hat.
> Michael: *Say something!*
>
> Maggie approaches agent miniCat, putting her ear close to it.
> Maggie: *Say something, cat.*

5.3.5 Children disliked playing with "turnea-off" portable, tangible voice agents. Since the tablet app gave children limited control over voice agents' speech, we explored telling children to use their own voices instead, as if controlling a puppet. The result was a hectic session with a lot of interventions from researchers and the teacher who was present. Children did not accept the idea very well, the boys in particular. While there were many physical interactions with the agents (58.2% of session time), their verbal interactions were very low (3.4% of session time as compared to other similar sessions where children verbally interacted with the agent for an average 12.31% of session time). That said, toward the end of the session, two girls played as we suggested, as demonstrated by the following excerpt from session 18:

> Jessica gets agent miniCat from the pair of boys.
> Together with Maggie, they show off the house they had been building for agent miniCat.
> Maggie: [to agent miniCat] *This is your house that we built for you.*
> Maggie as miniCat: *It's wonderful.*
> Jessica as miniCat: *This is amazing.*

5.3.6 Difficulty understanding speech synthesis. In our initial sessions, we used *Amazon Polly's* default speech settings [75]. We noticed children sometimes had difficulty understanding what the voice agents said. We later slowed the speed of the voices by adjusting prosody settings, which solved the problem. Below is an example of a misunderstanding extracted from session 14:

> Agent miniBear: *You are making great use of materials.*
> Michael (shouting): *She wants some Cheerios!!!!*
> The whole room laughs.

5.3.7 Design recommendations. Children are likely to bring with them stereotypes about voice agents based on their perceived characteristics or personality. For instance, a recent study found that 5-12 year-old children were more willing to ask a variety of questions to personified interfaces [69]. Another study with 5-8 year old children focusing on robot voices [58], indicated that children could not reliably assign gender to a robot based on its voice, but preferred robots that matched their gender. In our experience, children's stereotypes about the type of being the character represented impacted their behavior toward the voice agent, similar to what has been observed with adults extending gender and ethnic stereotypes to computers and reacting to personality traits [55]. Designers should carefully select appearances and other outward characteristics that are likely to elicit constructive behavior from children [4].

Additionally, voice agents that lack contextual information can be counterproductive and may not be a good fit for supporting lightly-structured activities. This finding indicates that it is important to young children that agents be aware of context and able to converse, which is consistent with previous recommendations for embodied conversational agents [22] and older children's expectations of intelligent user interfaces [30,68]. Systems with context awareness could take into account prior events in their speech [47] and initiate speech at socially appropriate times.

Children's interest in controlling the speech of voice agents could be leveraged for a variety of empowering activities (e.g., programming, learning about grammar), but our experiences suggest such control may not be appropriate for social activities. Regardless of the setting, designers should be aware that default speech synthesizer settings for adults may not work well for young children. These were likely tested with adults and intended for tasks such as providing directions while driving, which require quick speech. Designers should experiment with the wide range of options available from modern speech synthesizers [75,76] to find ideal settings for their target audience.

6 DISCUSSION

In this section, we reflect on our findings beyond the discussions included in the design recommendations and discuss potential implications on future designs of VUIs for young children.

6.1 Voice Agents to Facilitate High-Quality Social Play

Our experiences over these 24 design sessions suggest that voice agents can help facilitate high-quality social play, primarily by keeping children socially engaged in play and prompting novel uses of physical props during play. Out of the setups we explored, the tangible, portable agents with speech controlled by researchers worked best. Making the voice agents tangible and portable enabled children to put the agents at the center of their play, augment the agents with physical props, embed the agents in their constructions, and stay fully engaged with the physical and social space around them.

Researchers controlling speech ensured contextual awareness of the voice agents, which was not possible with the voice control app. Contextual awareness was critical for making the voice agents credible play partners in a fluid and lightly-structured activity. Children were sensitive toward what the voice agent said, sometimes finding the version of the voice agent controlled through the app to be disappointing, which impacted their engagement in play.

The addition of voice agents augmented adults' abilities to scaffold the play activities, redirect children's attention, and create opportunities for children to use their creativity in the social play context. In that sense, the voice agent became a tool that supported the target behaviors of *ToM* style play [14].

While the children were very interested in controlling the speech of the voice agents, giving them the ability to do so had a detrimental effect on children's engagement in play. We do not believe this is a dead end though. The strong interest by children points at opportunities for incorporating speech control into other applications.

6.2 Considerations for Voice Agents' Interactions with Young Children

An interesting finding from our explorations was that young children continued to engage with voice agents as if they were autonomous despite knowing adults were controlling what the voice agents said. This highlights an opportunity for exploration of VUIs as an alternative avenue for adults to communicate with children. For instance, it could be a playful way to redirect children's attention toward their chores or provide another way for caregivers to play with children by controlling what the voice agent says. In our design sessions, the children were also curious and interested in learning how things work, so designers should consider disclosing how a voice agent operates in terms that children can understand.

We believe that personification was critical to motivate children to engage in play as shown in previous research where children asked questions about agents' experiences and preferences [16,69]. In our explorations, we found that a simple representation of a character (e.g., square shape featuring a simple representation of an animal or robot face) sufficed for young children to be interested in engaging with the voice agent. We expect that some voice agents may represent characters already known by children [61], in which case they will bring a specific set of expectations. However, in other situations, there may be an opportunity to develop new characters specifically for a voice agent. Based on our experiences, it is important to design the characters represented by voice agents such that children interact with them in ways that will benefit the target activity.

Another aspect to consider when designing a voice agent for children under the age of five is its tangible affordances. Young children are still developing their senses and communicating through touch with the world around them [35]. Based on our observations, combining tangible affordances with voice agents can promote collaboration, social skills, and integration with physical play. Such interactive characteristics will likely be beneficial for other lightly-structured activities.

6.3 Ethical Concerns

There are many ethical concerns regarding young children's interactions with technology [39] and each kind of technology has its own peculiarities. VUIs, if they make use of speech recognition, are usually connected to cloud-based services, which poses privacy concerns. In other words, children's devices could be constantly sending speech heard around them to a company's online servers [40]. We avoided making use of speech recognition in our research activities since it was not necessary to fulfill our research goal, but also to avoid privacy issues and sending human subjects data to third parties.

Besides privacy, another worrying aspect of voice technologies is that they could influence children's values. For instance, a study of a commercial smart doll's impact on 4-10 year old children's judgments found the doll could influence children's moral transgressions (e.g., taking out a toy during snack time, hitting another child), but was unsuccessful in persuading children to disobey an instruction [67]. In our research sessions, we observed children following suggestions or performing tasks directed by the agent, which may raise concerns about who controls the agent. Therefore, it is important to consider the risks of unsupervised young children interacting with VUIs.

6.4 Limitations

We conducted our research in the context of supporting high-quality social play in the style of the well-defined, Vygotsky-inspired, *Tools of the Mind (ToM)* curriculum [14] at a local preschool. Our findings may not apply to different settings and other types of high-quality social play. In addition, only eight children participated in our research, which biases our findings toward their needs, abilities, and preferences. However, we believe the small number of children was appropriate for an exploratory research phase with many sessions reviewed in depth.

Since our goal was to explore voice agent design to support *ToM* style play, we did not compare *ToM* style play with and without voice agent support. Such a comparison is a long-term goal, and we conducted the research presented in this paper to better understand the tradeoffs in the implementation of voice agents to support *ToM* style play. Our findings may inform the design of future controlled experiments. In addition, we did not compare our voice agent sessions to sessions without a voice agent because the warmup sessions without agents involved time for children to get used to the researchers and three out of four had all eight children together, while the remaining sessions included at most four children. Because we did not conduct a controlled experiment, we provide only descriptive statistics in our findings, as we believe it is not appropriate to use inferential statistics given our approach.

7 CONCLUSION

This paper described an exploration of the design of voice agents in the context of high-quality social play with 3-4 year old children. Our partnership with children guided the research directions in our 24 design sessions. The exploration of a diverse set of voice agent setups enabled us to learn about tradeoffs in voice agent design. Our findings suggest that researcher-controlled tangible, portable voice agents can effectively support high-quality social play in the style of the *ToM* curriculum by keeping children socially engaged in play and enabling the integration of the voice agent with physical prop play. We also learned that although children wanted to control the speech of voice agents, giving them the ability to do so was detrimental to high-quality social play. In addition, we discussed important ethical considerations and opportunities for the design of future VUIs.

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 in the design sessions. Luiza Superti Pantoja's work was funded by a fellowship from the CAPES Foundation (99999.013579/2013-04).

SELECTION AND PARTICIPATION OF CHILDREN

Our university IRB reviewed and approved our research activities. Parents of participating children received a written invitation and mailed a signed informed consent form to us prior to their children beginning participation in our research activities. Children only participated in a design session if they desired. If not, we gave them the option to observe. There was always a teacher present in the room and we minimized privacy issues by not using a cloud-based service for the VUI. All research data is stored in secure cloud-based storage approved by our university's IRB.

REFERENCES

[1] amazon. *Alexa Loses Her Voice – Amazon Super Bowl LII Commercial*. Retrieved September 20, 2018 from https://www.youtube.com/watch?v=J6-8DQALGt4&frags=pl%2Cwn

[2] Sean Andrist, Iolanda Leite, and Jill Lehman. 2013. Fun and Fair: Influencing Turn-taking in a Multi-party Game with a Virtual Agent. In *Proceedings of the 12th International Conference on Interaction Design and Children* (IDC '13), 352–355. https://doi.org/10.1145/2485760.2485800

[3] Alissa N. Antle and Alyssa F. Wise. 2013. Getting Down to Details: Using Theories of Cognition and Learning to Inform Tangible User Interface Design. *Interacting with Computers* 25, 1: 1–20. https://doi.org/10.1093/iwc/iws007

[4] J. P. Baggaley. 1985. Design of a TV Character with Visual Appeal for Preschool Children. *Journal of Educational Television* 11, 1: 41–48. https://doi.org/10.1080/0260741850110106

[5] W. Steven Barnett, Kwanghee Jung, Donald J. Yarosz, Jessica Thomas, Amy Hornbeck, Robert Stechuk, and Susan Burns. 2008. Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early childhood research quarterly* 23, 3: 299–313.

[6] G.E. Baykal, I. Veryeri Alaca, A.E. Yantaç, and T. Göksun. 2018. A review on complementary natures of tangible user interfaces (TUIs) and early spatial learning. *International Journal of Child-Computer Interaction* 16: 104–113. https://doi.org/10.1016/j.ijcci.2018.01.003

[7] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottrjdge. 2018. Understanding the Long-Term Use of Smart Speaker Assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3: 91.

[8] Doris Bergen. 2002. The role of pretend play in children's cognitive development. *Early Childhood Research & Practice* 4, 1: n1.

[10] 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

[11] Clancy Blair and C. Cybele Raver. 2015. School Readiness and Self-Regulation: A Developmental Psychobiological Approach. *Annual Review of Psychology* 66, 1: 711–731. https://doi.org/10.1146/annurev-psych-010814-015221

[12] Clancy Blair and Rachel Peters Razza. 2007. Relating Effortful Control, Executive Function, and False Belief Understanding to Emerging Math and Literacy Ability in Kindergarten. *Child Development* 78, 2: 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

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

[15] Alyssa A Breitbach, Cassandra A Jenn, and Gary Milavetz. 2017. Distraction techniques for immunizations in children and adolescents. *Journal of the American Pharmacists Association* 57, 3: 414–415.

[16] A. J. Brush, Paul Johns, Kori Inkpen, and Brian Meyers. 2011. Speech@Home: An Exploratory Study. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '11), 617–632. https://doi.org/10.1145/1979742.1979657

[17] Rebecca Bull and Gaia Scerif. 2001. Executive Functioning as a Predictor of Children's Mathematics Ability: Inhibition, Switching, and Working Memory. *Developmental Neuropsychology* 19, 3: 273–293. https://doi.org/10.1207/S15326942DN1903_3

[19] 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.

[20] Stephanie M. Carlson and Louis J. Moses. 2001. Individual Differences in Inhibitory Control and Children's Theory of Mind. *Child Development* 72, 4: 1032–1053. https://doi.org/10.1111/1467-8624.00333

[22] Justine Cassell. 2001. Embodied conversational agents: representation and intelligence in user interfaces. *AI magazine* 22, 4: 67.

[23] CDC. *Milestones Moments*. Retrieved September 3, 2014 from https://www.cdc.gov/ncbddd/actearly/pdf/parents_pdfs/milestonemomentseng508.pdf

[24] Yi Cheng, Kate Yen, Yeqi Chen, Sijin Chen, and Alexis Hiniker. 2018. Why doesn't it work?: voice-driven interfaces and young children's communication repair strategies. In *Proceedings of the 17th ACM Conference on Interaction Design and Children - IDC '18*, 337–348. https://doi.org/10.1145/3202185.3202749

[25] Caron A. C. Clark, Verena E. Pritchard, and Lianne J. Woodward. 2010. Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology* 46, 5: 1176–1191. https://doi.org/10.1037/a0019672

[27] Adele Diamond, W Steven Barnett, Jessica Thomas, and Sarah Munro. 2007. Preschool Program Improves Cognitive Control. *Science (New York, N.Y.)* 318, 5855: 1387–1388. https://doi.org/10.1126/science.1151148

[28] Adele Diamond and Kathleen Lee. 2011. Interventions shown to aid executive function development in children 4 to 12 years old. *Science* 333, 6045: 959–964.

[29] 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.

[30] Stefania Druga, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2018. How smart are the smart toys?: children and parents' agent interaction and intelligence attribution. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 231–240.

[31] Allison Druin. 1988. Noobie: The animal design playstation. *ACM SIGCHI Bulletin* 20, 1: 45–53.

[32] Kimberly Andrews Espy, Melanie M. McDiarmid, Mary F. Cwik, Melissa Meade Stalets, Arlena Hamby, and Theresa E. Senn. 2004. The Contribution of Executive Functions to Emergent Mathematic Skills in Preschool Children. *Developmental Neuropsychology* 26, 1: 465–486. https://doi.org/10.1207/s15326942dn2601_6

[33] Olivier Friard and Marco Gamba. 2016. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution* 7, 11: 1325–1330. https://doi.org/10.1111/2041-210X.12584

[34] 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

[35] Matthew J. Hertenstein. 2002. Touch: Its Communicative Functions in Infancy. *Human Development* 45, 2: 70–94. https://doi.org/10.1159/000048154

[36] Michael S. Horn, R. Jordan Crouser, and Marina U. Bers. 2012. Tangible interaction and learning: the case for a hybrid approach. *Personal and Ubiquitous Computing* 16, 4: 379–389. https://doi.org/10.1007/s00779-011-0404-2

[37] Juan Pablo Hourcade, Alissa N Antle, Lisa Anthony, Jerry Alan Fails, Ole Sejer Iversen, Elisa Rubegni, Mikael Skov, Petr Slovak, Greg Walsh, and Anja Zeising. 2018. Child-computer interaction, ubiquitous technologies, and big data. *interactions* 25, 6: 78–81.

[38] 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

[39] Juan Pablo Hourcade, Anja Zeising, Ole Sejer Iversen, Narcis Pares, Michael Eisenberg, Chris Quintana, and Mikael B Skov. 2017. Child-Computer Interaction SIG: Ethics and Values. 1334–1337.

[40] Juan Pablo Hourcade, Anja Zeising, Ole Sejer Iversen, Mikael B Skov, Alissa N Antle, Lisa Anthony, Jerry Alan Fails, and Greg Walsh. 2018. Child-Computer Interaction SIG: Ubiquity and Big Data--A Changing Technology Landscape for Children. SIG07.

[41] Jennifer Hyde, Sara Kiesler, Jessica K. Hodgins, and Elizabeth J. Carter. 2014. Conversing with children: cartoon and video people elicit similar conversational behaviors. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 1787–1796. https://doi.org/10.1145/2556288.2557280

[42] Theofanis Kannetis, Alexandros Potamianos, and Georgios N. Yannakakis. 2009. Fantasy, Curiosity and Challenge As Adaptation Indicators in Multimodal Dialogue Systems for Preschoolers. In *Proceedings of the 2Nd Workshop on Child, Computer and Interaction* (WOCCI '09), 1:1–1:6. https://doi.org/10.1145/1640377.1640378

[43] Liisa Klenberg, Marit Korkman, and Pekka Lahti-Nuuttila. 2001. Differential Development of Attention and Executive Functions in 3- to 12-Year-Old Finnish Children. *Developmental Neuropsychology* 20, 1: 407–428. https://doi.org/10.1207/S15326942DN2001_6

[44] 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

[46] Iolanda Leite, André Pereira, and Jill Fain Lehman. 2017. Persistent memory in repeated child-robot conversations. In *Proceedings of the 2017 Conference on Interaction Design and Children*, 238–247.

[48] Silvia Lovato and Anne Marie Piper. 2015. Siri, is this you?: Understanding young children's interactions with voice input systems. In *Proceedings of the 14th International Conference on Interaction Design and Children*, 335–338.

[49] Ewa Luger and Abigail Sellen. 2016. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297.

[50] 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

[52] Candice M. Mottweiler and Marjorie Taylor. 2014. Elaborated role play and creativity in preschool age children. *Psychology of Aesthetics, Creativity, and the Arts* 8, 3: 277–286. https://doi.org/10.1037/a0036083

[53] PAULA Y. Mullineaux and LISABETH F. Dilalla. 2009. Preschool Pretend Play Behaviors and Early Adolescent Creativity. *The Journal of Creative Behavior* 43, 1: 41–57. https://doi.org/10.1002/j.2162-6057.2009.tb01305.x

[55] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues* 56, 1: 81–103.

[56] Jean Piaget and Barbel Inhelder. 1969. *The psychology of the child*. Basic books.

[57] Jenny S. Radesky, Jayna Schumacher, and Barry Zuckerman. 2015. Mobile and Interactive Media Use by Young Children: The Good, the Bad, and the Unknown. *Pediatrics* 135, 1: 1. https://doi.org/10.1542/peds.2014-2251

[58] Anara Sandygulova and Gregory MP O'Hare. 2018. Age-and Gender-Based Differences in Children's Interactions with a Gender-Matching Robot. *International Journal of Social Robotics*: 1–14.

[59] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18*, 857–868. https://doi.org/10.1145/3196709.3196772

[60] Katherine J Short-Meyerson and Leonard J Abbeduto. 1997. Preschoolers' communication during scripted interactions. *Journal of child language* 24, 02: 469–493.

[61] Erik Strommen. 1998. When the Interface is a Talking Dinosaur: Learning Across Media with ActiMates Barney. 8.

[62] C. Sylla, I. S. P. Pereira, C. P. Coutinho, and P. Branco. 2016. Digital Manipulatives as Scaffolds for Preschoolers' Language Development. *IEEE Transactions on Emerging Topics in Computing* 4, 3: 439–449. https://doi.org/10.1109/TETC.2015.2500099

[63] Cristina Sylla, Clara Coutinho, Pedro Branco, and Wolfgang Müller. 2015. Investigating the use of digital manipulatives for storytelling in pre-school. *International Journal of Child-Computer Interaction* 6: 39–48.

[64] Jeffrey Trawick-Smith, Jennifer Wolff, Marley Koschel, and Jamie Vallarelli. 2015. Effects of Toys on the Play Quality of Preschool Children: Influence of Gender, Ethnicity, and Socioeconomic Status. *Early Childhood Education Journal* 43, 4: 249–256. https://doi.org/10.1007/s10643-014-0644-7

[65] Sherry Turkle. 2012. *Alone together: Why we expect more from technology and less from each other*. Basic books, New York, NY, USA.

[66] Lev S Vygotsky. 1967. Play and its role in the mental development of the child. *Soviet psychology* 5, 3: 6–18.

[67] Randi Williams, Christian Vázquez Machado, Stefania Druga, Cynthia Breazeal, and Pattie Maes. 2018. My doll says it's ok: a study of children's conformity to a talking doll. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 625–631.

[68] 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.

[69] 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

[70] Echo & Alexa - Amazon Devices - Amazon Official Site. Retrieved September 20, 2018 from https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b?ie=UTF8&node=9818047011

[71] Siri. *Apple*. Retrieved September 20, 2018 from https://www.apple.com/siri/

[72] Personal Digital Assistant - Cortana Home Assistant - Microsoft. *Microsoft Cortana, your intelligent assistant*. Retrieved September 20, 2018 from https://www.microsoft.com/en-us/cortana

[73] Google Assistant - Just Say "Hey Google" and Make Google Do It. *Google Assistant - Just Say "Hey Google" and Make Google Do It*. Retrieved September 20, 2018 from https://assistant.google.com/

[74] Echo Dot Kids Edition | a smart speaker with Alexa for kids. Retrieved September 20, 2018 from https://www.amazon.com/amazon-echo-dot-kids-edition/dp/B077JFK5YH

[75] Amazon Polly. Retrieved January 14, 2019 from https://aws.amazon.com/polly/

[76] SSML Tags Supported by Amazon Polly - Amazon Polly. Retrieved January 14, 2019 from https://docs.aws.amazon.com/polly/latest/dg/supported-ssml.html