



Voice Design to Support Young Children's Agency in Child-Agent Interaction

Layne Jackson Hubbard
layne.hubbard@colorado.edu
Department of Computer Science
University of Colorado Boulder
Boulder, Colorado, USA

Shanli Ding
Department of Computer Science
University of Colorado Boulder
Boulder, Colorado, USA

Vananh Le
Department of Computer Science
University of Colorado Boulder
Boulder, USA

Pilyoung Kim
Department of Psychology
University of Denver
Denver, USA

Tom Yeh
Department of Computer Science
University of Colorado Boulder
Boulder, Colorado, USA



Figure 1: A four-year-old child creates a collage, then converses with their stuffed animal to tell a story about their creation.

ABSTRACT

Agency is essential to play. As we design conversational agents for early childhood, how might we increase the child-centeredness of our approaches? Giving children agency and control in choosing their agent representations might contribute to the overall playfulness of our designs. In this study with 33 children ages 4–5 years old, we engaged children in a creative storytelling interaction with conversational agents in stuffed animal embodiments. Young children conversed with the stuffed animal agents to tell stories about their creative play, engaging in question and answer conversation from 2 minutes to 24 minutes. We then interviewed the children about their perceptions of the agent's voice, and their ideas for agent voices, dialogues, and interactions. From babies to robot daddies, we discover three themes from children's suggestions: Family Voices,

Robot Voices, and Character Voices. Additionally, children desire agents who (1) scaffold creative play in addition to storytelling, (2) foster personal, social, and emotional connections, and (3) support children's agency and control. Across these themes, we recommend design strategies to support the overall playful child-centeredness of conversational agent design.

CCS CONCEPTS

- Human-centered computing → User centered design; User studies;
- Social and professional topics → Children.

KEYWORDS

child-agent interaction, conversational agents, creativity support, storytelling, agency, play

ACM Reference Format:

Layne Jackson Hubbard, Shanli Ding, Vananh Le, Pilyoung Kim, and Tom Yeh. 2021. Voice Design to Support Young Children's Agency in Child-Agent Interaction. In *3rd Conference on Conversational User Interfaces (CUI '21)*, July 27–29, 2021, Bilbao (online), Spain. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3469595.3469604>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CUI '21, July 27–29, 2021, Bilbao (online), Spain

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8998-3/21/07.

<https://doi.org/10.1145/3469595.3469604>

1 INTRODUCTION AND RELATED WORK

Agency is essential to play [61], and supporting children's agency in child-agent interaction can foster playful child-centeredness in our designs [15, 18, 46]. And while researchers are increasingly studying the importance of providing parents and caregivers with agency and control over the use of conversational agents in home and therapeutic settings [5, 21, 39], designers of computational and expressive technologies must continue to work toward supporting the agency of children through playful self-discovery [20]. The need to cultivate learning through play [62] must challenge designers to examine the elements of play in our designs. In this work, we explore multiple ways to support children's agency — from open-ended creative storytelling [6, 22, 26, 31, 52, 54], to selecting their agent's voice [2, 18, 60] and their agent's embodiment [2]. We then ask children to reflect on their experience in the interaction. In this way, young children contribute child-generated ideas to the space of child-agent interaction design.

Meaningfully involving children in technology design can support their agency and empowerment [19, 24, 29, 50], and asking young children questions about their drawings can help them communicate their ideas [23, 26–28]. Open-ended questioning in particular can be empowering. Not only can open-ended questioning support children's expressive ability [7], narrative skills [7, 42], and verbal participation in both child-parent interactions [30, 33, 38] as well as child-agent interactions [56–59], but open-ended questioning also provides children with autonomy in how they choose to respond and how they choose to express their narrative. In this work, we use open-ended questioning via child-agent interaction to support children's agency in interactive storytelling.

In interactive system design, autonomy and self-expression are two critical elements of agency [35] and agency itself is important for creating enjoyable user experiences [45]. Here, agency is the power to take meaningful action [40], and involves exercising control over "processes, motivation, action, and environment" [3]. Agency arises when one's own self is the source for communication [53], further connecting agency to self-expression. Robots can support children's agency by inviting children to participate in self-expression activities [4, 14] or by giving children opportunities to shape the development of a story [14]. Within interactive storytelling, agency is fostered by having control over the characters, interactions, events, and outcomes of a story [41]. The more that children can engage in open-ended self-expression, the more their agency is facilitated [4]. Finally, exercising control within agent personalization can also support young children's agency. By providing methods for customizing the agent's voice [34, 43], the agent's sound effects [1], or the agent's embodiment [49], interaction designers can support children in communicating their own visions for child-agent interaction.

To explore young children's ideas for playful conversational agent design, we recruit 33 children ages 4–5 years old across 10 U.S. states for a remote child-agent interaction session (Figure 1). Using a range of materials found in their homes, children freely craft or build, then choose a stuffed animal to guide them in telling a story about their creation. The stuffed animal agent asks scaffolded, open-ended questions about the child's creative play, and they collaborate in a multi-turn sequence to construct a reflective story. Finally, we

ask children and parents semi-structured interview questions to gather their feedback. Our key research question examines young children's perspectives and ideas for child-agent interaction design — specifically exploring the design of the agent's voice and dialogue scaffolds.

From 33 child interview transcripts, we use affinity diagramming and thematic analysis techniques to understand young children's perspectives on the agent's voices and their ideas for future agent voices, dialogues, and interactions. The data suggests three distinct themes of agent voices recommended by young children: Family Voices, Robot Voices, and Character Voices. Further, children desire agents who (1) scaffold creative play in addition to storytelling, (2) foster personal, social, and emotional connections, and (3) support children's agency and control. We examine strategies to incorporate children's themes for agent voices, dialogues, and interactions into future designs. And, we explore strategies to support child-centeredness in child-agent interaction. For example, although most major providers of text-to-speech processing services offer many options for synthesized voices, these options are almost exclusively adult-sounding voices — and do not commonly include the child-centered voices that children suggest. We conclude that finding ways to support young children's agency and control in choosing their agents' representations and interactions — as well as developing child-centered voices — might contribute to the overall playfulness of conversational agent design.

2 INTERACTION DESIGN

Our child-agent interaction design is informed by a series of formative studies [26–28]. Here, we visited preschool classrooms and explored the ways in which caregivers ask questions to guide children in telling stories about their creative artifacts. This method of asking questions to develop stories about creative play is based on the Storybook Journey curriculum [37], which examines storytelling as a developmentally appropriate form of reflection. We analyzed transcript data to categorize patterns of inquiry and abstracted an open-ended 'serve and return' model to describe the caregivers' story scaffolding methods. In the 'serve and return' early childhood model of contingent reciprocity, cooperation occurs as partners appropriately respond to each other's input [48].



Figure 2: A graphic of our initial child-robot interaction. A mobile application running the voice interaction is inserted into a stuffed animal (stuffie) embodiment. We instruct the children, "We put the phone in the stuffie, to make the stuffie talk."

2.1 Interactive Robotic Object (IRO)

We diagrammed the resulting conversational model into an abstract state machine and iteratively developed its implementation into an interactive robotic object (IRO). The result is a conversational agent that asks scaffolded questions; in return, the child constructs responses to tell a story. After asking a question, the conversational agent waits for the child's response. After the child responds, the agent asks a new question with the aim of scaffolding the development of a story. A mobile application runs the voice interaction, and the smartphone or speaker is inserted into a stuffed animal embodiment (Figure 2). Through this screenless design, we aim to support children's attentiveness in real world contexts. And by using everyday materials to construct our interactive robotic object — such as smartphones, speakers, and stuffed animals — we aim to develop a tool that can be used in a variety of settings.

To adapt the child-agent interaction for a remote study, we employ a hybrid Wizard of Oz method. We ask children to first choose a stuffed animal from their home, and we explain that we will use a computer, a smartphone, and a robot voice to make their stuffed animal talk. When the child is ready to tell their story, we ask them to place their stuffed animal atop their parent's smartphone. We then operate the agent's text-to-speech voice prompting through the smartphone. In this way, children have control in co-locating the agent's stuffed animal body with the agent's voice.

2.2 Text-to-Speech Voice Design

To create the agent's voice, we selected from commonly available voices across major text-to-speech systems, including Google Cloud, IBM Watson, Amazon Polly, and Microsoft Azure. Although these major text-to-speech systems offer many voice options for designers to use within voice applications, these systems mainly offer adult-sounding voices and do not commonly include children's voices or character voices in their available options. Among the adult-sounding voices, we selected voices that were as naturalistic as possible. Here, we focused on voices that used punctuation triggers in the text (such as "?" and "!") to modify the intonation of the agent's speech — such as raised intonation at the end of a sentence to indicate a question. Of the adult-sounding voices, we then selected a woman's voice and a man's voice for the child-agent interaction. By providing these two options for the agent's voice, we aimed to (1) examine children's feedback on the two commonly available voice categories in text-to-speech systems, and (2) support children's agency in controlling their agent's voice.

2.3 Scaffolded, Open-Ended Questioning

To start the child-robot interaction, the stuffed animal introduces itself to the child, "*Hi, I'm the story helper. I will help you tell a story about what you made. I will ask you lots of questions to help you tell a story.*" We designed the robot's questions with the goal of being sufficiently open-ended to integrate within a wide variety of children's creative play activities. From the children's perspective, their conversation experiences with the robot consists of four stages: (1) The robot initiating conversation, (2) the robot asking story beginning questions, (3) the robot asking story follow-up questions, and (4) the robot asking story ending questions. Throughout, the robot uses open-ended 'wh-' questions to elicit children's verbal

expression. For example, the robot initiates storytelling by asking "*What did you make?*" and "*Tell me a story about what you made.*" The robot proceeds by asking scaffolded questions to guide the child in telling a story about their creation.

By saying "*The end*" or "*I'm all done,*" the child signals to the agent to end their storytelling session. The robot then asks the child to name their story to guide the child in synthesizing their story's theme. Finally, the agent transitions the interaction by asking a series of reflective questions to foster iteration, including "*Next time you make something, what are you going to make?*" and "*Next time you tell a story, what is it going to be about?*" Through the agent's scaffolded, open-ended questioning and stuffed animal embodiment we aim to support our design goal of playful child-centeredness while supporting children's agency within the storytelling interaction.

3 METHODS

Due to the ongoing COVID-19 pandemic in the United States, we conducted a remote study in participants' homes to prioritize the safety of all involved. We limited our study session to 30-minutes with the goal accommodating the differing needs and background of parents across the socioeconomic spectrum — who may have competing caregiving, work, and household demands. Our goal in recruitment was to represent families from diverse regions, backgrounds, family structures, income levels, and educational experiences.



Figure 3: The creative table setup for a child and parent participant in our study. The parent offered magnet tiles, crayons, scissors, glue, cotton balls, and paper for their child's creative play. The five-year-old child gathered four stuffed animals — a bear, a dinosaur, a seahorse, and a lamb — to embody the agent during their child-agent interaction session. The child then chose their dinosaur for the agent, who asked open-ended questions to help the child tell a ~10 minute story about "The Big Bad Wolf."

3.1 Participants

We recruited 33 children ages 4–5 ($M = 4.95$ years, $SD = 0.58$ years, 17 females and 16 males) and their parents. Of the 33 families in our study, one family had two children ages 4 and 5 years old. To support their family dynamic, we included both siblings in our study. We recruited participants through child care centers, family

services email lists, community advocacy groups, and social media announcements shared broadly in neighborhood parent-social groups.

Participants resided in 22 cities from 10 states across the U.S. (Arizona, California, Colorado, Idaho, Illinois, Michigan, Montana, New York, Oregon, and Texas). In Michigan, for example, the participants' cities were as rural as Mikado (population 899 in 2018), as suburban as Highland (pop. 20,179) or Waterford (pop. 72,948), and as urban as Detroit (pop. 672,662) [11].

Children varied in the type of educational programs they attended, with 42.42% (14) attending preschool or pre-kindergarten, 45.45% (15) attending kindergarten, 3% (1) child attending a home-based preschool program, 3% (1) child attending a home-based kindergarten program, and 6.1% (2) without an educational program.

Parents identified their children's racial and ethnic backgrounds: 18.75% (6) identified as being of Hispanic or Latino descent, 68.3% (28) as White/Caucasian, 14.6% (6) as Black or African American, 7.3% (3) as Asian; 7.3% (3) as American Indian/Alaska Native, and 7.3% as Other (mixed race). Six children (18.75%) spoke a language in addition to English (Shona, Mandarin Chinese, and Spanish).

Parents were predominantly female (93.9% (31), 6.1% (2) male). Ten (30.3%) parents were the sole caregivers in their household, whereas 69.7% (23) of parents had an additional caregiver in the household. The total number of people in participants' households ranged from 2 to 7 people ($M = 4.18$, $SD = 0.98$).

Families reported household incomes as low as \$4,000 USD per year, and as high as \$400,000 USD per year. The national median household income in the United States was \$63,703 in 2019 [12] — of our participants, 61.29% (19) families had a household income above the national median, while 38.71%, (12) families had a household income below the national median.

Parents' educational levels varied as well. Eleven parents (33.3%) reported having a Bachelor's degree. Eleven parents reported having less than a Bachelor's degree (6.1% (2) with a GED, 6.1% (2) with a high school diploma, and 21.1% (7) with an Associate's degree). And eleven parents reported having greater than a Bachelor's degree (27.3% (9) with a Master's degree, 3% (1) with an M.D./Ph.D., and 3% (1) with a J.D.)

3.2 Protocol

3.2.1 Setup. Parents were given instructions prior to their session: (1) Gather a variety of materials to support their children's creative play. Include, for example: paper, crayons, markers, paint, stickers, glue, paper scraps, paper shapes, dried leaves, interlocking bricks, magnet tiles, and bristled blocks. (2) Setup a table or desk with the creative supplies (Figure 3). (3) Have their child choose a stuffed animal to use in the reflective storytelling activity (Figure 4). We then used the mobile application for Zoom conferencing to host the study session. We asked families to connect via audio without using video in order to (1) respect participants' privacy in their home environments, and (2) maintain focus on the child, parent, and child-robot interactions — rather than on the researcher's visual persona.

3.2.2 Creative Activity. First, the researcher asked the child to describe their creative materials ("What things do you have to create



Figure 4: This four-year-old child gathered five of their stuffed animals for the agent embodiment. The child then chose a tiger stuffed animal for their agent and conversed for ~24 minutes to tell a story about interplanetary travel entitled "The Space Story From Fly Guy."

with today?"), and then invited the child to create something using the materials provided. Children were given options for their activity: "You can make your family, your school, a forest, a garden, or your own idea." Children were given options as a way to support their initial ideation if needed, but were also reminded that, "You can make it however you want." Children were given 5 minutes to make their creation, and could ask their parents for help or work together.

3.2.3 Selection of the Agent's Voice. Before children began their child-robot interaction, we explained that "We will use a computer, your phone, and a robot voice to make your stuffed animal talk." We then invited children to choose their stuffed animal robot's voice, in order to support their agency and control in the interaction. Children could choose between a naturalistic-sounding woman's voice ("Do you want your robot to have a girl voice?"), or a naturalistic-sounding man's voice ("Or do you want your robot to have a boy voice?") Children chose their voice based only on this verbal description.

3.2.4 Reflective Storytelling Interaction. Next, the researcher instructed the child: "Now, your stuffed animal is going to ask you questions to help you tell a story about what you made. We're going to use a computer, a phone, and a robot voice to make your stuffed animal talk. When you're ready, pick up your stuffie and sit it down on top of the phone to make your stuffie talk." The researcher then used our story scaffolding software and text-to-speech (TTS) synthesized voices to guide the child in a spoken question-and-answer (Q&A) dialogue, in which the stuffed animal agent asks questions and the child improvises responses to the agent's prompts.

3.2.5 Child Interview. After the child-agent reflective storytelling interaction, the researcher rejoined the audio call and asked the child semi-structured questions about the agent voice they heard, about their suggestions for agent voices, and about their suggestions for what other things the agent could do or say: "I'm building stuffed animal robots to help kids tell stories about the things they make, and I'm wondering: What should the robot sound like? What kind of voice

do you want it to have? What did you think of the voice you heard? What else should the robot say?"

3.2.6 Parent Interview. The following day, we called the parent participant for a 20 minute interview. We asked parents semi-structured questions about their observations of their child during and after the child-agent interaction. Parents shared their insights while observing their children in the interaction, as well as their shared discussions about the interaction after-the-fact.



Figure 5: This four-year-old child built a creation with paper, scissors, glue, and fabric scraps, then talked with their stuffed animal agent for ~11 minutes to tell story about horses, fairies, clowns, and a tooth that got lost in the forest.

3.3 Data Analysis

We transcribed the audio recordings of the child-agent interactions, child interviews, and parent interviews. We then used qualitative techniques including affinity diagramming and thematic analysis in order to explore insights related to our research questions. Affinity diagramming [13, 47] is used in human computer interaction (HCI) and design research to analyze qualitative data from observational and user interviews [32, 36]. In this study, we used affinity diagramming techniques to spatially cluster insights from the semi-structured interviews with children and parents. Next, we used thematic analysis [9, 16] – a qualitative analysis technique used in both psychology and interaction design [10, 55] – to generate themes from the clustered data and use those themes to code the interview data. To examine children's engagement and expression in the child-robot interaction, we used quantitative measures including word count, conversational turn count, and length of interaction.

4 FINDINGS

When conversing with the stuffed animal agent to tell a story about their creative play (Figure 5), children ages 4–5 years old engaged with the agent between 2.27 and 24.53 minutes (median = 7.25 minutes). Children spoke between 26 and 960 words (median = 298

words). And children engaged in 10 to 75 conversational turns with the agent (median = 24 turns). The conversational turn count (CTC) refers to the number of turn-taking pairs within the interaction – each time the agent speaks and the child responds constitutes one turn-taking pair [17]. These findings demonstrate the wide range of our participants' verbal expression and conversational engagement within the child-agent interaction (Table 1). In this range, we observe how open-ended questioning can support a *low floor* and *high ceiling* for the highly variable verbal abilities of early childhood [25, 44, 51] within child-agent interaction.

For example, children with lower word counts typically responded to the agent's initial prompts with just a few words per response. Yet even with shorter responses these children could still engage in prolonged turn-taking with the agent. And after repeated prompting by the agent, some children increased their word count per response – suggesting that the agent's scaffolding was effective in supporting the children in increased self-expression and agency. Conversely, some children with higher word counts had lower conversational turn counts due to their tendency to speak in multiple sentences without pausing frequently to receive a new prompt from the agent. Whereas as other children with higher word counts also had higher conversational turns as they often spoke a sentence then paused for their agent's next prompt. This flexibility in turn-taking also suggests ways that open-ended questioning might support children's agency in self-expression – in, for instance, choosing how much input they desire from the agent in scaffolding their narrative development. By providing children with flexibility, designers can support children's differing needs and desires within child-agent interaction.

4.1 Children's Selection of an Agent's Voice

In their child-robot storytelling sessions, 32 of the 33 children selected a voice for their robot. Of the 32 children who selected a voice, 18 children (56.25%) chose the "boy" robot voice, whereas 14 children (43.75%) chose the "girl" voice. Of the 16 male children, 16 (100%) selected the "boy" voice. Of the 17 female children, 14 (82.35%) chose the "girl" voice. 3 (17.64%) female children chose the "boy" voice in their first session.

4.2 Children's Feedback on Machine-Synthesized Voices

During the child interview, we asked children semi-structured interview questions about their perceptions of the agent's machine-synthesized voice. Eight children responded to the prompt, *"What did you think of the robot's voice?"* Several children responded with a simple value judgement, such as *"I like it"* or *"Good."* Another child elaborated, *"I likes his voice because he just was on the phone and I just talked to him for a long time before until he needed to leave."*

4.2.1 Expectations of Character-Voice Alignment. Other children commented on the characteristics of the voice, as well as the implementation of the synthesized voice: *"[It] sounds like a boy and girl at the same time. Like it was a little shaky a lot."* As another child shared, *"I didn't like it because it sounds like a human more."* Here, they were comparing their expectation of a "robot voice" with the machine-synthesized human voice. Another child was similarly

Table 1: Children’s Speech, Language, and Engagement Measures Within the Child-Agent Interaction

	Length of Interaction	Word Count	Conversational Turn Count
Median	7.25 minutes	298 words	24 turns
Min	2.27 minutes	26 words	10 turns
Max	24.53 minutes	960 words	75 turns
Range	17.28 minutes	934 words	65 turns

disappointed in the characteristics of the voice, *"Well, you could just turn off that voice. Because, um, this is not a voice that matches my stuffed animal. It's only a Pikachu, so you only go like 'pika.' Like Pikachu."* This child expected the robot's voice to match the voice of their stuffed animal.

4.2.2 Open-Ended Voice Mappings. In their elaborated responses, children highlight the difference between their expectations or desires for the conversational agent's voice, and the machine-synthesized human they heard. The verbal labels of "*girl voice*", "*boy voice*", and "*robot's voice*" may have restricted the children's interpretation of the resulting voice. Instead of providing verbal or textual labels, voice designers might consider ways to widen children's interpretations — such as allowing children to select a voice by playing a voice sample, instead of using a verbal or textual description. This may allow for more open-endedness in their ability to map the agent's voice onto their existing stuffed animals (Figure 6). Similarly, expanding the set of voices that children can choose from may allow for increased child-centeredness as they create their own mappings between the voices and their stuffed animal embodiments.



Figure 6: This five-year-old child chose two stuffed animals for their agent embodiment — an owl and an elephant. After drawing a picture and telling a ~10 minute story about "Flying Birds" with one stuffed animal agent, the child then chose to draw another picture for storytelling with their other stuffed animal.

4.3 Children’s Ideas for Agent Voice Design

Next, we asked children semi-structured interview questions about their ideas for agent voice design. We then used affinity diagramming and thematic analysis techniques to iteratively code, cluster, categorize, and extract themes from the children's insights. Twenty-seven children responded to the prompt: *"We're building stuffed animal robots to help kids tell stories. I'm wondering — what should the robot sound like? What kind of voice should it have?"* Three overall themes emerged: (1) Family, (2) Robots, and (3) Characters.

4.3.1 Family Voices. Family was the most explored theme by children, and their responses span several categories: (1) Baby Voices, (2) Kid Voices [girls, boys], (3) Teenager Voices, and (4) Adult Voices [moms, dads].

One child suggested *"a little baby voice"* and another child suggested a range of voices: *"A baby voice. Like a kid voice. Even also a teenager voice."* Of all voices across all themes and categories, *"a girl voice"* was the most frequent response, though some children mixed this request with others, such as *"some girl voice and some boy voice."* The many requests for baby, girl, boy, and teenager voices suggest that prioritizing children's voices might support the overall child-centeredness of our conversational agent designs.

One child elaborated on the quality of the voice, *"Um, maybe like a girl voice. Not like shaky a lot like how it was. And like real. Like a real girl voice talking."* This suggests a potential preference for human-recorded voices rather than machine-synthesized voices. Designers might think about ways — such as puppeteering — to support human-human in addition to human-agent interactions.

Children who suggested adult human voices tended to describe them as wanting to hear a *"mom"* or *"dad"* voice, emphasizing the theme of family. Designers might extend this notion through an intergenerational approach that includes the voices of elders and grandparents as well. By considering the range of voices across families, designers of conversational agents can foster child-centeredness by connecting to children's home lives.

4.3.2 Robot Voices. Many children expressed a desire for increasingly robotic voices (Figure 7). Here, children's responses spanned two categories: (1) Typical Robots, and (2) Fun Robots.

Several children suggested prototypical robot sounds and voices: *"A normal robot. Beep bop bop. Beep beep bop. Like a real robot."* Here, children often used sounds such as *"do-dee do-dee-do"* to communicate their ideas. One child responded that they wanted a *"robot sound"* and another responded that they wanted it to sound *"like a robot robot"*. And some children combined machine representations with human characteristics: *"Like Daddy sometime pretend he's a*

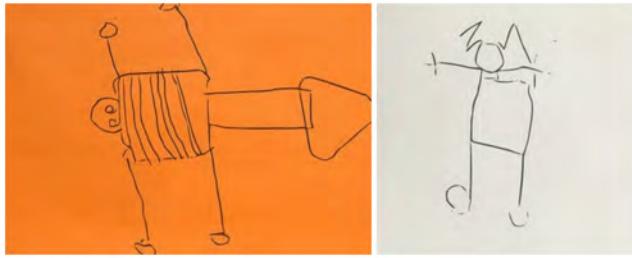


Figure 7: (Left) This five-year-old child first drew a picture of a scorpion and talked with their stuffed animal agent for ~10 minutes to tell a story. **(Right)** After sharing their ideas for the robot's voice in the interview, this child drew a picture of how they wanted their robot to look and sound — *"like a real robot"* with *"lightning bolts."*

robot and make funny voice. Daddy robot." This suggests an opportunity for playful robot voices in supporting joyful, child-centered interactions.

While conversational agents for adult populations are increasingly human-sounding, our data suggests that young children may benefit from exploring distinctly machine-like sounds and voices. For some young children, realistic human voices may not align with their experience of the agent's behaviors. Designers of conversational agents might consider using a combination of robot sounds, voices, and even prototypical robot embodiments in their designs. Robotic representations may allow young children to playfully navigate their interactions with the agent, and may also help them develop appropriate expectations of the agent's capabilities.

4.3.3 Character Voices. Finally, a few children suggested voices modeled after their stuffed animals' characters, including a *"Pikachu voice"* and *"Buzz Lightyear."* Sometimes children excitedly mimicked the character voices, or giggled or laughed while emulating their characters' sounds. Two siblings who took turns telling stories also took turns practicing their character voices. This suggests two categories of character voices: (1) Fun Characters, and (2) Silly Characters. By supporting a variety of expressive character voices in their conversational agents (such as animals, monsters, as super-heros), designers can support children's playfulness and agency in choosing voices that align with their characters or emotions.

4.4 Children's Ideas for Conversational Interaction Design

We also asked children about their ideas on how the robot should engage with them. Twenty children responded to the prompt, *"I'm wondering — what other questions do you want the robot to ask you? What else should the robot say?"* Through affinity diagramming, we extracted three main themes from the children's responses: (1) Scaffold Creative Play in Addition to Storytelling, (2) Foster Personal, Social, and Emotional Connections, and (3) Support Children's Agency and Control.

4.4.1 Scaffold Creative Play in Addition to Storytelling. Children expressed ways for conversational agents to further engage with them in creative and story exploration. Agents could invite children to

"color some more" or ask, *"What colors are you using?"* Agents could ask *"about the LEGOs"* or ask, *"Did you make some shapes?"* Agents could also help children expand on the contents of their stories through contextual questions such as, *"Where is space? Where is the planet?"* Children also thought of open-ended ways that agents could use to help children tell stories. One child suggested that agents ask, *"What is happening to your story? What did they do in the story?"* Here, children highlight their interests in both creative and story scaffolding. By combining open-ended and contextual methods, designers of conversational agents might discover new ways to enrich their conversational scaffolds.

4.4.2 Foster Personal, Social, and Emotional Connections. Children also suggested ways that agents could connect with them personally, by asking *"Hi kids! What are you up to?"* or *"What are you doing today?"* One child explored ways that agents could support the child's own role in their story, such as by asking *"What did you do in your book? Do you have a dog in your book? Are you a police officer in your book?"* Children seemed interested in having interactions that help them reflect on and explore their own selves — in addition to their creative play.

One child explored the bi-directionality of child-agent roles in an emotional context, *"I want to give it a hug"* and the agent should *"give me a hug."* Here, designers might consider methods for supporting young children in expressing their social-emotional needs too.

4.4.3 Support Children's Agency and Control. Children indicated a desire to have increased control of the agent, as well as to switch roles with the agent. While one child asked for *"more questions,"* another child remarked: *"It was kind of annoying me because it was like a lot of questions. Ask lower amount of questions."* These opposing perspectives might align with differences in young children's individual needs. By providing children with ways to control the interaction, designers can support children in meeting their own needs.

Children suggested potential for controlling what the agents say and the questions that the agents ask. *"They can ask anything you want,"* said one child; and *"he can say whatever you want,"* suggested another. One child elaborated on this dynamic: *"Can kids say everything your stuffie says? Kids should ask the questions too."* To increase children's agency and control, designers might consider ways for children to input questions into the system, or support children in puppeteering the agent. Puppeteering the agent may also foster peer-to-peer or sibling learning, wherein one child puppeteers the questioning for another.

Children explored ways they could switch roles with the agent in both creating and storytelling contexts. *"They will play crafts if the kids say crafts,"* suggested one child — indicating a role reversal from the agent prompting the children in creative play. Multiple children also requested that agents share stories of their own. *"They should make a story,"* said one child; and agents could *"tell a stuffed animal story,"* said another. Here, switching roles may expand the opportunities for children to learn and build their skills within child-agent interaction — as they think about how to control and scaffold their own ideas for creative exploration. Further, by providing children with increased agency and control, designers might increase the child-centeredness of their child-agent interactions.

Table 2: Minimal Availability of Child-Centered Synthesized Voices in Major TTS Systems

	Amazon Polly	Google Cloud	IBM Watson	Microsoft Azure
Baby Voices	n/a	n/a	n/a	n/a
Kid Voices	Yes	n/a	n/a	n/a
Robotic Voices	n/a	n/a	n/a	n/a
Character Voices	n/a	n/a	n/a	n/a

4.5 Parent Insights on Agent Voice Design

Parents largely echoed the children's desire for more child-like and playful voices. Because many children perceived the agent's voice as an adult, it became an "*authority figure*" for some children instead of a playmate. "*When she talks to her teachers she uses a smaller voice. She regarded the robot as an adult too.*" Parents suggested "*kid voices*" that are "*higher pitched*" and "*kid-sounding*" as a way to signal to children that they can be playful in the interaction. Otherwise — because of many children's familiarity with remote learning during the COVID-19 pandemic — parents warned that adult voices create a perception that the child-agent interaction is a form of "*homework*" or "*schoolwork*." Conversational agent designers can instead support playfulness in our interactions by incorporating children's voices into our designs.

Finally, parents frequently reported in parent interviews that their children "*really enjoyed choosing different tones of voice*" by "*giving them options and letting them decide*." Conversational agent designers can support the child-centeredness of our designs by providing children with options for choosing their agent's voice.

4.6 Availability of Child-Centered Synthesized Voice Options

After categorizing young children's ideas and preferences for agent voice design, we then examined major text-to-speech (TTS) processing systems to understand how children's ideas are — or are not — supported (Table 2). Because these systems provide voice synthesis services used widely by application developers and designers, the availability of voice options within these systems directly impact the voices used by researchers and developers alike. We examined major TTS processing systems, specifically Amazon Polly, Google Cloud, IBM Watson, and Microsoft Azure. While they each provide many options for adult-sounding voices, only Amazon Polly provided options for child voices. Amazon Polly provided one female child voice (*Ivy*) and two male child voices (*Justin* and *Kevin*). These child voices are only available in English. Of the other, non-adult voices suggested by young children for agent voice design — including baby voices, robotic voices, and fun or silly character voices — none are offered by these major TTS systems.

5 DISCUSSION AND FUTURE WORK

In this work, we focus on young children's experience with a machine-synthesized agent voice, their ideas for agent voice design, and their ideas for conversational interaction design. We engaged children ages 4–5 years old in a creative storytelling activity —

facilitated by open-ended prompting from a stuffed animal conversational agent — then asked children and parents semi-structured questions to gather their insights for child-agent interaction design.

Young children's expectations about the agent's human-sounding voice didn't always align with their stuffed animal characters, or with their expectations of robot characteristics. Children generated many ideas and expressed a desire for three types of agent voices: Family Voices, Robot Voices, and Character Voices. Yet we found that most major text-to-speech (TTS) processing systems do not provide these child-centered voices in their array of synthesized voice options. Providing child-centered voice options is a ripe area for innovation with potential for impact — the offerings of major TTS systems are widely integrated into applications used by researchers and developers. By providing child-centered voice options, TTS systems will influence the design of child-agent interactions across a range of applications and contexts.

Within family voices, children shared a desire for child-like agents with baby voices, kid voices, and teenager voices. Within robot voices, children requested prototypical machine sounds and monotone robotic voices, as well as expressive and fun robots. And within character voices, children desired both silly and fun characters who they can map to their own favorite stuffed animals at home. Parents echoed the importance of using of child-like voices and providing children with choices for their agent's voice. Finally, children generated three categories of ideas for conversational interaction design. They desire agents who: (1) Scaffold Creative Play in Addition to Storytelling, (2) Foster Personal, Social, and Emotional Connections, and (3) Support Children's Agency and Control.

When designing for child-agent interaction in the early years, we encourage interaction designers to provide children with choices for their agent's voice, and to include options for child-like and playful voices — specifically family voices (including kid voices), robot voices, and character voices. Similarly, we recommend that the developers of text-to-speech processing systems expand their speech offerings beyond adult-like voices to include these child-centered voices. In this way, voice designers can incorporate these expanded offerings into their designs for children. We also encourage designers of conversational agents to incorporate open-ended questioning into their designs to support children's agency and self-expression, and to design for children's three categories of ideas for child-agent interaction. Here, designers can consider how to scaffold both creative and reflective play, how to support social-emotional connections, and how to provide children with increased agency and control.

A limitation of this study is our use of stuffed animal embodiments alone. Because we focused only on stuffed animal agents, this

may have limited children's reflection and imagination when generating new agent voices and interactions. To better explore a range of conversational agent design, future work could benefit from engaging children with other agent forms – from digital avatars to disembodied voices to humanoid platforms. In this way, children might imagine voices, dialogue, and scaffolds that align with a diversity of agent implementations. In future work, we aim to explore ways to provide children with increased agency and control when interacting with the agent. In addition to voice interaction, we wonder how multimodal input such as tactile interaction might support young children with differing verbal skills in flexibly controlling the agent's voice, prompting, and scaffolds. Through open-ended questioning and child-led agent customization, designers of child-agent interaction can foster children's agency and self-expression in their creative play experiences. As the late maker-movement pioneer Mike Eisenberg [8] challenges us, "The time is now right to return our attention to children's intellectual narratives about minds, thought, learning, and the nature of the self" [20]. To do so, we must design opportunities for children to explore their ideas through agency and play.

ACKNOWLEDGMENTS

We thank all of the phenomenal children and families who participated in this research across Arizona, California, Colorado, Idaho, Illinois, Michigan, Montana, New York, Oregon, and Texas. Funding for this research was provided by the Piton Foundation and the OpenIDEO Early Childhood Innovation Prize. This research was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1650115, as well as the NSF National AI Institute for Student-AI Teaming (iSAT) under grant DRL 2019805. The opinions expressed are those of the authors and do not represent views of the NSF.

REFERENCES

- [1] Bryan Alexander. 2017. *The New Digital Storytelling: Creating Narratives with New Media—Revised and Updated Edition*. Abc-clio.
- [2] Matthew P Aylett, Selina Jeanne Sutton, and Yolanda Vazquez-Alvarez. 2019. The right kind of unnatural: designing a robot voice. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. 1–2.
- [3] Albert Bandura. 2001. Social cognitive theory: An agentic perspective. *Annual review of psychology* 52, 1 (2001), 1–26.
- [4] Claudio Baraldi. 2008. Promoting self-expression in classroom interactions. *Childhood* 15, 2 (2008), 239–257.
- [5] Erin Beneteau, Ashley Boone, Yuxing Wu, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2020. Parenting with Alexa: Exploring the Introduction of Smart Speakers on Family Dynamics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [6] Marina Umaschi Bers and Justine Cassell. 1998. Interactive storytelling systems for children: Using technology to explore language and identity. *Journal of Interactive Learning Research* 9 (1998), 183–215.
- [7] Maria Birbili and Ioanna Karagiorgou. 2009. Helping children and their parents ask better questions: An intervention study. *Journal of Research in Childhood Education* 24, 1 (2009), 18–31.
- [8] Paulo Blikstein, Yasmin Kafai, and Roy Pea. 2019. Mike Eisenberg: A One of a Kind Pioneer in the Learning Sciences.
- [9] Richard E Boyatzis. 1998. *Transforming qualitative information: Thematic analysis and code development*. sage.
- [10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [11] US Census Bureau. [n.d.]. Census.gov. <https://www.census.gov/en.html>
- [12] US Census Bureau. 2020. Income and Poverty in the United States: 2019. [https://www.census.gov/library/publications/2020/demo/p60-270.html#:~:text=Medianhouseholdincomewas\\\$protect\(T1\\\$textdollar68,703, andTableA-1](https://www.census.gov/library/publications/2020/demo/p60-270.html#:~:text=Medianhouseholdincomewas\$protect(T1\$textdollar68,703, andTableA-1)
- [13] Hugh Byer and Karen Holtzblatt. 1998. Contextual Design—Defining Customer-Centred Systems.
- [14] Justine Cassell and Kimiko Ryokai. 2001. Making space for voice: Technologies to support children's fantasy and storytelling. *Personal and ubiquitous computing* 5, 3 (2001), 169–190.
- [15] Fabio Catania, Micol Spiale, Giulia Cosentino, and Franca Garzotto. 2020. What is the Best Action for Children to "Wake Up" and "Put to Sleep" a Conversational Agent? A Multi-Criteria Decision Analysis Approach. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–10.
- [16] Victoria Clarke and Virginia Braun. 2014. Thematic analysis. In *Encyclopedia of critical psychology*. Springer, 1947–1952.
- [17] Seamus Donnelly and Evan Kidd. 2021. The Longitudinal Relationship Between Conversational Turn-Taking and Vocabulary Growth in Early Language Development. *Child Development* 92, 2 (2021), 609–625.
- [18] 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.
- [19] Allison Druin. 2002. The role of children in the design of new technology. *Behaviour and information technology* 21, 1 (2002), 1–25.
- [20] Michael Eisenberg, Sherry Hsi, and Hyunjoo Oh. 2017. Machines and minds: The new cognitive science, and the potential evolution of children's intuitions about thinking. *International Journal of Child-Computer Interaction* 14 (2017), 1–4. <https://doi.org/10.1016/j.ijcci.2017.06.001>
- [21] Saad Elbeleidy, Aubrey Shick, and Tom Williams. 2021. Teleoperation Interface Usage in Robot-Assisted Childhood ASD Therapy. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 162–166.
- [22] Jerry Alan Fails, Allison Druin, and Mona Leigh Guha. 2014. Interactive storytelling: interacting with people, environment, and technology. *International Journal of Arts and Technology* 7, 1 (2014), 112–124.
- [23] Mona Leigh Guha, Allison Druin, Gene Chipman, Jerry Alan Fails, Sante Simms, and Allison Farber. 2004. Mixing ideas: a new technique for working with young children as design partners. In *Proceedings of the 2004 conference on Interaction design and children: building a community*. 35–42.
- [24] 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 (2013), 14–23.
- [25] Erika Hoff. 2006. How social contexts support and shape language development. *Developmental review* 26, 1 (2006), 55–88.
- [26] Layne Jackson Hubbard, Yifan Chen, Eliana Colunga, Pilyoung Kim, and Tom Yeh. 2021. Child-Robot Interaction to Integrate Reflective Storytelling Into Creative Play. In *Creativity and Cognition*. 1–8. <https://doi.org/10.1145/3450741.3465254>
- [27] Layne Jackson Hubbard, Chen Hao Cheng, Boskin Erkocevic, Dylan Cassady, and Andrea Chamorro. 2018. MindScribe: Reflective Inquiry through Scaffolded Storytelling for Low-Income and Multilingual Early Childhood Communities. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 345–346. <https://doi.org/10.1145/3173386.3177823>
- [28] Layne Jackson Hubbard, Boskin Erkocevic, Dylan Cassady, Chen Hao Cheng, Andrea Chamorro, and Tom Yeh. 2018. MindScribe: Toward Intelligently Augmented Interactions in Highly Variable Early Childhood Environments. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*. ACM, 1–2. <https://doi.org/10.1145/3180308.3180319>
- [29] Ole Sejer Iversen, Rachel Charlotte Smith, and Christian Dindler. 2018. From computational thinking to computational empowerment: a 21st century PD agenda. In *Proceedings of the 15th Participatory Design Conference: Full Papers—Volume 1*. 1–11.
- [30] Jennifer Yusun Kang, Young-Suk Kim, and Barbara Alexander Pan. 2009. Five-year-olds' book talk and story retelling: Contributions of mother–child joint bookreading. *First Language* 29, 3 (2009), 243–265.
- [31] Jacqueline Kory and Cynthia Breazeal. 2014. Storytelling with robots: Learning companions for preschool children's language development. In *The 23rd IEEE international symposium on robot and human interactive communication*. IEEE, 643–648.
- [32] Ilpo Koskinen, John Zimmerman, Thomas Binder, Johan Redstrom, and Stephan Wensveen. 2013. Design research through practice: From the lab, field, and showroom. *IEEE Transactions on Professional Communication* 56, 3 (2013), 262–263.
- [33] Yana Kuchirkko, Catherine S Tamis-LeMonda, Rufan Luo, and Eva Liang. 2016. 'What happened next?': Developmental changes in mothers' questions to children. *Journal of Early Childhood Literacy* 16, 4 (2016), 498–521.
- [34] Kwan-Min Lee and Clifford Nass. 2005. Social-psychological origins of feelings of presence: Creating social presence with machine-generated voices. *Media Psychology* 7, 1 (2005), 31–45.
- [35] Mike EU Lighthart, Mark A Neerincx, and Koen V Hindriks. 2020. Design patterns for an interactive storytelling robot to support children's engagement and agency. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 409–418.
- [36] Andrés Lucero. 2015. Using affinity diagrams to evaluate interactive prototypes. In *IFIP Conference on Human-Computer Interaction*. Springer, 231–248.
- [37] Sue McCord. 1995. *The storybook journey: pathways to literacy through story and play*. Merrill.

[38] Joyce H McNeill and Susan A Fowler. 1999. Let's talk: Encouraging mother-child conversations during story reading. *Journal of Early Intervention* 22, 1 (1999), 51–69.

[39] Emily McReynolds, Sarah Hubbard, Timothy Lau, Aditya Saraf, Maya Cakmak, and Franziska Roesner. 2017. Toys that listen: A study of parents, children, and internet-connected toys. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5197–5207.

[40] Janet Horowitz Murray and Janet H Murray. 2017. *Hamlet on the holodeck: The future of narrative in cyberspace*. MIT press.

[41] Marija Nakevska, Anika van der Sanden, Mathias Funk, Jun Hu, and Matthias Rautenberg. 2017. Interactive storytelling in a mixed reality environment: the effects of interactivity on user experiences. *Entertainment computing* 21 (2017), 97–104.

[42] Carole Peterson, Beulah Jesso, and Allyssa McCabe. 1999. Encouraging narratives in preschoolers: An intervention study. *Journal of child language* 26, 1 (1999), 49–67.

[43] Amanda Purington, Jessie G Taft, Shruti Sannon, Natalya N Bazarova, and Samuel Hardman Taylor. 2017. "Alexa is my new BFF" Social Roles, User Satisfaction, and Personification of the Amazon Echo. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*. 2853–2859.

[44] Mitchel Resnick and Ken Robinson. 2017. *Lifelong kindergarten: Cultivating creativity through projects, passion, peers, and play*. MIT press.

[45] Christian Roth and Hartmut Koenitz. 2016. Evaluating the user experience of interactive digital narrative. In *Proceedings of the 1st International Workshop on Multimedia Alternate Realities*. 31–36.

[46] Supraja Sankaran, Chao Zhang, Mathias Funk, Henk Aarts, and Panos Markopoulos. 2020. Do I have a say? Using conversational agents to re-imagine human-machine autonomy. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–3.

[47] Raymond Scupin. 1997. The KJ method: A technique for analyzing data derived from Japanese ethnology. *Human organization* (1997), 233–237.

[48] Jack P. Shonkoff and Susan Nall Bales. 2011. Science Does Not Speak for Itself: Translating Child Development Research for the Public and Its Policymakers. *Child Development* 82, 1 (2011), 17–32.

[49] Claudia Sino, Sylvia van Der Pal, Olivier A Blanson Henkemans, Anouk Keizer, Bert PB Bierman, Rosemarijn Looije, and Mark A Neerincx. 2018. Friendship with a robot: Children's perception of similarity between a robot's physical and virtual embodiment that supports diabetes self-management. *Patient education and counseling* 101, 7 (2018), 1248–1255.

[50] Katta Spiel, Laura Malinvern, Judith Good, and Christopher Frauenberger. 2017. Participatory evaluation with autistic children. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5755–5766.

[51] Marie A Stadler and Gay Cuming Ward. 2005. Supporting the narrative development of young children. *Early Childhood Education Journal* 33, 2 (2005), 73–80.

[52] Ming Sun, Iolanda Leite, Jill Fain Lehman, and Boyang Li. 2017. Collaborative storytelling between robot and child: A feasibility study. In *Proceedings of the 2017 Conference on Interaction Design and Children*. 205–214.

[53] SS Sundar. 2006. Self as Source: Agency and Customization in Interactive Media Top Paper. In *Proc. the Annual Meeting of the International Communication Association*.

[54] Cristina Sylla, Íris Susana Pires Pereira, and Gabriela Sá. 2019. Designing manipulative tools for creative multi and cross-cultural storytelling. In *Proceedings of the 2019 on Creativity and Cognition*. 396–406.

[55] Miguel A Teruel, Elena Navarro, Pascual González, Víctor López-Jaquero, and Francisco Montero. 2016. Applying thematic analysis to define an awareness interpretation for collaborative computer games. *Information and Software Technology* 74 (2016), 17–44.

[56] Ying Xu, Dakuo Wang, Penelope Collins, Hyelim Lee, and Mark Warschauer. 2021. Same benefits, different communication patterns: Comparing Children's reading with a conversational agent vs. a human partner. *Computers & Education* 161 (2021), 104059.

[57] Ying Xu and Mark Warschauer. 2020. "Elinor Is Talking to Me on the Screen!" Integrating Conversational Agents into Children's Television Programming. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[58] Ying Xu and Mark Warschauer. 2020. Exploring young children's engagement in joint reading with a conversational agent. In *Proceedings of the Interaction Design and Children Conference*. 216–228.

[59] Ying Xu and Mark Warschauer. 2020. Wonder with elinor: designing a socially contingent video viewing experience. In *Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts*. 251–255.

[60] Ye Yuan, Stryker Thompson, Kathleen Watson, Alice Chase, Ashwin Senthilkumar, AJ Bernheim Brush, and Svetlana Yarosh. 2019. Speech interface reformulations and voice assistant personification preferences of children and parents. *International Journal of Child-Computer Interaction* 21 (2019), 77–88.

[61] Jennifer M Zosh, Kathy Hirsh-Pasek, Emily J Hopkins, Hanne Jensen, Claire Liu, Dave Neale, S Lynne Solis, and David Whitebread. 2018. Accessing the inaccessible: Redefining play as a spectrum. *Frontiers in psychology* 9 (2018), 1124.

[62] Jennifer N. Zosh. 2017. *Learning through play: a review of the evidence*. LEGO Fundation.