Children Asking Questions: Speech Interface Reformulations and Personification Preferences

Svetlana Yarosh[1], Stryker Thompson[1], Kathleen Watson[1], Alice Chase[1], Ashwin Senthilkumar[2], Ye Yuan[1], A.J. Bernheim Brush[3]

[2]Eden Prairie High School
Eden Prairie, MN, USA
ashwin.s320@gmail.com

[1]University of Minnesota
Minneapolis, MN
{lana, thom5043, watso559, chase263, yuan0191}@umn.edu

[3]Microsoft Research
Redmond, WA, USA
ajbrush@microsoft.com

ABSTRACT

The pervasive availability of voice assistants may support children in finding answers to informational queries by removing the literacy requirements of text search (e.g., typing, spelling). However, most such systems are not designed for the specific needs and preferences of children and may struggle with understanding the intent of their questions. In our investigation, we observed 87 children and 27 adults interacting with three Wizard-of-Oz speech interfaces to arrive at answers to questions that required reformulation. We found that many children and some adults required help to reach an effective question reformulation. We report the common types of reformulations (both effective and ineffective ones). We also compared three versions of speech interfaces with different approaches to referring to itself (personification) and to the participant (naming personalization). We found that children preferred personified interfaces, but naming personalization did not affect preference. We connect our findings to implications for design of speech systems for families.

Author Keywords

Children; family; speech interfaces; voice assistants; search; information seeking; personification; personalization

ACM Computing Classification System

CCS → Human-centered computing → Human computer interaction (HCI) → Interaction devices → **Sound-based input / output**

INTRODUCTION

Speech interfaces like Apple Siri, Amazon Alexa, Microsoft Cortana, and Google Assistant are becoming increasingly available in homes, bringing to life many of the possible benefits of speech-based interaction identified by HCI researchers as early as the 1990s (e.g., [7,42,45]). In-home speech interfaces are used for a wide variety of tasks including playing music, timers, and informational queries. Beyond commercial systems, many libraries and toolkits make it easier than ever to add speech input to arbitrary systems [49]. Speech interfaces appear to be a particularly compelling interaction modality for children, because they reduce or limit the need for literacy and typing [41]. However, there are a number of challenges that make it difficult for commercial systems to support use by children: recognizing children's speech is notoriously hard (e.g., [21]), recognizing intent is even more difficult (e.g., [33,53]), and COPPA laws may make it hard for companies to study and accommodate children's needs (e.g., [1]). Speech interfaces and assistants are also traditionally designed for and tested with adults, and children may have different information needs and preferences.

In this investigation, we focus on speech interfaces for answering children's informational queries. We chose to focus on informational queries because previous work shows that the plurality of family speech interactions with a home voice kiosk [4] and of children's interactions with Apple's Siri [36] fell into the "information seeking" and "web search" categories. Additionally, speech interfaces provide an opportunity to address significant literacy-related issues that children currently face when searching online [10,11]. To develop a better understanding of children's practices and preferences, we conducted an investigation with 87 children ages 5–12 interacting with three variations of a Wizard-of-Oz speech interface.

Our primary research question is ***(RQ1) How do children structure and restructure informational queries towards a speech interface?*** We answer this question through a qualitative content analysis of children's errors and reformulations when using our system. Second, we were interested in the role of specific speech interfaces' design choices on children's preferences. In particular, we investigate open questions on the role of personification (i.e., named agent speaking in first person) and personalization (i.e., knowing and referring to personal information about the user) in a speech interface. Thus, our second research question is ***(RQ2) How does a speech interface's personification and naming personalization affect children's experience and preference?*** We answer this question through both a quantitative analysis of children's preference between three

IDC '18, June 19–22, 2018, Trondheim, Norway
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5152-2/18/06…$15.00
https://doi.org/10.1145/3202185.3202207

speech interface versions and a qualitative analysis of their interactions towards those interfaces. Finally, speech interfaces implemented as information appliances (e.g., Amazon Echo, Google Home) may need to account for the preferences and needs of both children and adults [3]. As such, our third research question is ***(RQ3) How do the experiences and preferences of adults compare to those of children?*** We address this question by comparing both qualitative and quantitative aspects of children's interaction with our speech interfaces from those of 27 adults answering the same informational queries.

In the remainder of this paper, we begin by situating our work in the body of previous investigations on spoken dialog systems, personification and personalization in speech interfaces, and children's informational queries. We describe our methods, including the demographics of our participants, the speech interfaces used, and the procedure of the investigation. We discuss our analysis and findings in response to each of the above research questions. Finally, we conclude with a discussion that provides specific implications for the design of future speech interfaces for homes.

RELATED WORK

We outline previous work in spoken dialog systems and speech interfaces for children and adults to situate the three research questions we are exploring in this investigation.

Spoken Dialog Systems & Speech Recognition

Spoken dialog systems and speech recognition systems are becoming more and more sophisticated, with systems like Apple Siri, Amazon Alexa, Microsoft Cortana, and Google Assistant available as off-the-shelf products. Recent advances in the field focus on personalizing to the needs of specific users [6], maintaining an understanding of context over multiple turns of interaction [8], and asking appropriate clarifying questions to guide the user [30]. However, most of these advances target adult users.

A number of papers provide evidence that speech is a promising mode of computer interaction for children (e.g., [41]). Proposed applications of such systems are as diverse as play/entertainment (e.g., [5,22,28,32,33]), social skill practice (e.g., [15]), literacy education (e.g., [21,39,45]), and in-car entertainment (e.g., [18]). However, spoken dialog systems for children is an area with significant potential for improvement. First, speech recognition with children is notoriously difficult (e.g., [17,26,50,55]), especially for spontaneous utterances which have been described as "disfluent and ungrammatical" [21]. Second, recognizing *intent* is significantly more difficult than just recognizing speech, even with adults [53]. While several investigations have sought to generate datasets of children's speech (e.g., [29]), datasets of children's spontaneous utterances are lacking [41]. Our work begins addressing this gap, however we are not seeking to improve the state of the art in speech recognition or spoken dialog systems, but rather answer several open questions in speech interface design.

Personification & Personalization in Speech Interfaces

One open question in speech interfaces for the home is the role of personalization and personification. HCI researchers have pursued personalized and personified interaction with speech interfaces since the early 1990s [38], though whether these kinds of agents lead to the best user experience remains an open question even after decades of investigation (e.g., [52]). For example, one previous study found that personified agent-like output from a speech device led adults to interact with it in ways that were not well supported by their system (e.g., asking the device direct questions) [24]. Adults interacting with Siri reported negative responses to a similar mismatch between the level of personification and actual capabilities of the speech interface [37]. An investigation of adults' interaction with a virtual receptionist found that people attributed different amounts of personification to the receptionist and interacted differently with it based on that attribution [31]. Another investigation focused on voice output found no effects of output "embodiment" (speech coming directly from smart objects vs. a disembodied home control agent vs. a home controlled agent that was also embodied as an on-screen avatar) on users' experience, though the disembodied voice was slightly preferred overall [47]. Nonetheless, most current commercial speech interfaces (e.g., Siri, Amazon Alexa, Google Assistant, Cortana) are both personified and personalized.

The role of the user's age in personification and personalization preferences is even less clear. One paper suggested that "factual" versus "social" interaction with a dialog system may be an inherent user preference rather than an age-dependent characteristic [56]. Other preliminary work with speech interfaces in the home shows that children interact with voice systems differently than adults. For example, children are more likely to include social exchanges with the system (e.g., "bye!") [4]. Older participants are more likely than younger children to have a negative affective response to speech interfaces that violate privacy expectations (e.g., know information that the child didn't explicitly tell them) [34]. Outside of speech interfaces, at least a few investigations have suggested that personified search agents may help children interpret query results and found that this approach worked best with 8- and 9-year-olds (older children found it to be too "childish") [20]. Systems can be fairly accurate at distinguishing between adult and children's speech [43], however little is known about how a system could then adjust its personification and personalization in the most appropriate way to ages or preferences of users. Our work addresses this gap, with the goal of leading to more tailored speech appliances for families (e.g., as suggested in [3]).

Children's Informational Query Practices

While there has been substantial previous work on speech interfaces for children's entertainment (e.g., [18,22,32,33]), we focus on informational queries as the specific context of interaction with speech interfaces. There are two reasons for this decision. First, these interactions represent an important use case for speech interfaces. Previous work revealed that the plurality (30%) of family speech interactions with a home voice kiosk [4] and the plurality (45%) of children's interactions with Apple's Siri [36] fell into the information seeking and web search categories. Second, there is substantial potential for leveraging speech interfaces to address a number of challenges that children face with current text-based search practices. Log analysis has found that children's web searches are frequently "unsuccessful" and "confused" [13,14]. The major challenges identified in previous investigations of children's web search include spelling, typing, and query formulation [10,11]. Spelling and typing difficulties may be amplified by children's tendency to use longer, natural language queries [12,27] (though this may be culturally dependent, as a study of German children's search found that they were more likely to have shorter queries [19]). Speech interfaces may remove the barriers of spelling and typing, allowing children to focus on the task of query formulation.

Query reformulation in response to a misrecognition or misunderstanding is an important aspect of interacting with speech interfaces. This has been found to be a challenging task even for adults, who employed strategies as diverse as word substitution, phrase re-ordering, and phonetic emphasis [25,37]. Other previous work on speech-based home automation controls investigated adults' responses to different types of errors (i.e., ones leading to stagnation, regression, or partial progress towards a goal) [47], again reiterating that reformulation may be challenging. It may be even more difficult for children, but little is known about children's practices with speech interface query reformulation. Previous exploratory work that has examined children's interaction with Apple's Siri through a content review of YouTube videos found that children had substantial trouble dealing with speech interface errors, relying mostly on phonetic emphasis in reformulation [36]. Additionally, in a study with a similar WoZ design to ours, Oviatt et al. found that children change the prosodic (e.g., speed, pitch) qualities of their speech to match an animated agent's [48], but did not investigate semantic reformulations or adaptations. One of our goals in this work is to understand how children deal with restating or restructuring queries when they cannot get to an answer.

METHODS

In this section, we describe our setting, participants, and detail the system setup and procedure to support replication.

Setting

The study took place in the research outreach building at the Minnesota (MN) State Fair. The Driven2Discover building is a permanent facility on the State Fairgrounds, visited daily by thousands of fair attendees who are representative of the population in the Minnesota.[1] The permanent facility provided a relatively private and quiet space for the study to take place (e.g., other studies were separated by curtains and the building was protected from the general hustle and bustle of the Fair). There were two study stations set up in the building booth (see Figure 1), as well as another station for gathering initial information and consent from the children and parents.

Participants

We recruited child participants (ages 5–12), along with a parent or guardian who could provide consent, give information about the child, and potentially participate in this study. One benefit of our chosen study setting was the ability to recruit and include a greater diversity of families than most lab-based investigations. Families passing through the research facility and choosing to participate in the study were representative of Fair attendees as a whole. For example, 25% of the visitors were from rural counties (consistent with state population) and 28% of the parents did not have a college degree (consistent with published demographics statistics from the 2016 State Fair).

During the course of the study, 87 children completed the procedure (57% female; *M = 9 years old, SD = 1.99*). Parents or guardians were given one of three options:

1. If there was an empty station, they could attempt the tasks themselves (otherwise, we prioritized children).
2. They could sit next to their child to help read the questions and write answers (any assistance given by parents beyond reading and writing help, such as hints or prompts, was logged and included in our analysis).
3. They could do neither and wait for their child to complete the study.

Given these options, 27 adults completed option one, providing us with a base of comparison for adult participants (48% female; *M = 47 years old*, *SD = 10.61*). All

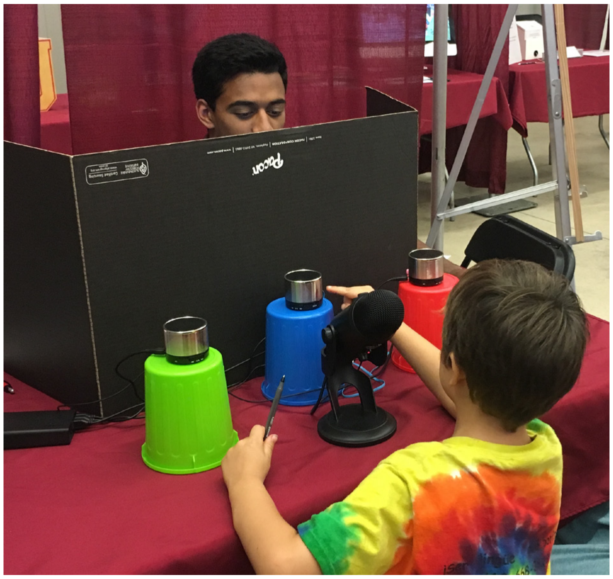


Figure 1. A child points to one of the interfaces in the study setup. Each interface is represented as a plastic bin with a speaker. The Wizard-of-Oz (visible behind the cardboard divider) controls the text-to-speech output of each device.

[1] http://d2d.umn.edu/

participants (adults and children) were either native English speakers or agreed with the statement "I am comfortable speaking English." We also requested parents to tell us which (if any) voice assistants they and their children have used in the past (e.g., Apple Siri, Amazon Alexa). 96% of the adults and 94% of the children had used one or more different voice assistants in the past. Parents were more likely to have used these interfaces infrequently (mode response was "less than once a week"), while children were more likely to use these interfaces frequently (mode response was "multiple times per day").

Systems

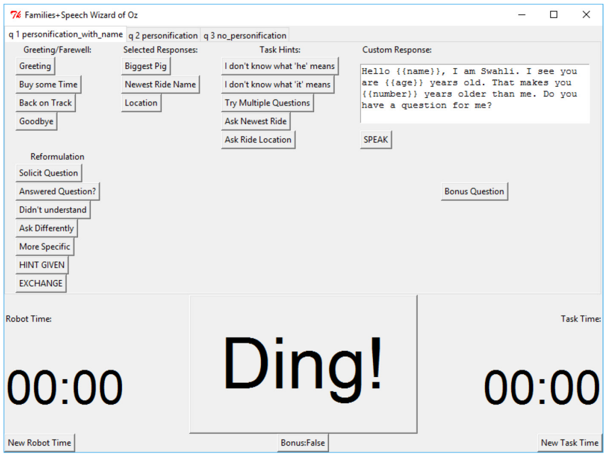
To address our research questions, we developed multiple versions of a voice assistant. We discuss the physical hardware employed, the software and Wizard-of-Oz training, and the specific modifications made in each condition.

Equipment

Each study station housed three variations of speech interfaces, each represented as a different plastic housing and an AISBR 3W wired speaker (see Figure 1) to allow participants to more easily distinguish between and refer to each interface. All participants were audio recorded using a Blue Yeti USB microphone, which contains a tri-capsule array to support field recording of human voices at a 48kHz sample rate and 16bit bit rate. All of these peripherals were connected to a laptop running the Wizard-of-Oz's control software and Audacity recording software.

Wizard-of-Oz Controls and Training

To focus our investigation on the role of question reformulations (rather than recognition factors), we chose to use a Wizard-of-Oz (WoZ) technique to simulate a voice assistant. This allowed us to have human-quality speech recognition, removing this confound from the investigation. As in other WoZ studies (e.g., [9]), to allow the Wizards to provide near real-time response to the participants, we had to develop custom WoZ control software. The team created a Python GUI (see Figure 2) to allow the Wizard to quickly and consistently select common responses and statements, as well as directly edit response text when modifications were necessary. The Wizards followed a specific script to ensure that participants received consistent responses from the system. Response text was converted to speech using Microsoft's *Zira* voice (female voice with an American accent) and the *pyttsx* Python text-to-speech library.



```
Hello {{name}}, I am Swahli. I see you
are {{age}} years old. That makes you
{{number}} years older than me. Do you
have a question for me?
```

Figure 2. A custom software interface guided the Wizard. Many common interactions (e.g., greetings, hints) were pre-programmed. Ad-hoc interactions could be typed directly into the response box or edited from an existing scripted response. After a participant stated their question, a "ding!" sound provided feedback that the "system" heard it. The interface guided the Wizard through condition order and logged interactions.

Five Wizards supported this investigation, allowing work to occur in shorter shifts to avoid inconsistencies due to fatigue. All Wizards used a common protocol and guide for contingency responses and trained together through multiple piloting sessions to increase consistency in Wizard response. Wizards were visible to the participants, though a privacy screen hid their immediate actions (see Figure 1). Participants were told that the Wizards were there to help the voice assistants. Due to our significant GUI shortcuts and piloting, most responses were quick and required no typing, revealing little about the Wizard's role. Due to the increasingly common use of human computation as a technique in computing systems, the IRB did not view WoZ to be an example of deception so no debriefing was required with participants who did not inquire about the specific functioning of the system. Only two adults expressed suspicion that the researcher had a greater role and they were debriefed after the study. None of the children expressed suspicion or inquired about how the system worked.

Three Conditions

To answer our questions regarding the role of personification and naming personalization, we built three variations of the speech system (supplementary materials include examples of full scripts of interaction with each system):

- **Voice Search System** – the non-personified and non-personalized system never referred to itself in first person, gave only task-related responses, and did not mention the participant's name, e.g.: "*Welcome to the voice search system. Please, say your question.*"
- **Fraga**[2] – the personified and non-personalized system referred to itself in first person, gave some responses that were not task-related, but did not mention the participant's name, e.g.: "*Hello, I am Fraga. Do you have a question for me?*"
- **Swali**[2] – the personified and personalized system referred to itself in first person, gave some responses that were not task-related, and periodically referred to the participant's name and age, e.g.: "*Hello Jake, I am Swali. I see you are 6 years old. That makes you 5.5 years older than me. Do you have a question for me?*"

[2] "Fraga" and "Swali" mean "question" in Swedish and Swahili, respectively.

Given the limited nature of children's attention spans and that reliable preference measures for children require within-subjects comparisons [57], we minimized the number of conditions. We omitted the non-personified, personalized condition as least natural and least interesting.

The Wizards were directed by the GUI as to the order of these interfaces and would switch speakers appropriately. The control interface automatically applied interface-specific wording and formatting to the dialog (e.g., adding the name to the greeting in the personalized condition). We chose to use the same voice for all three variations of the interface in order to avoid user preference of certain voices or dialects from influencing their selection.

Procedure

Potential participants were solicited by a research team member as they passed through the Driven2Discover Building. If they were still interested in the study after the brief pitch, the families were led to a table where the study was explained in more detail and a researcher answered their questions. The child wrote their first name on an assent form, if able. The parent filled out consent for himself or herself and for the child to participate, as well as a short demographic and background questionnaire.

The study represents a within-subjects design, with each participant asking at least one question to each interface, in counterbalanced order. It is important to note that we had specific ethical considerations and constraints that took priority over procedural consistency in certain cases. In our discussions with our IRB, it became clear that it was important to make sure that every child left the study feeling that they "succeeded" at the assigned task. This consideration led to four study design decisions. First, members of the family could sit next to the participant, potentially offering advice (this was used to allow siblings under 5 to serve as "special helpers" and get a toy at the end). Second, if a child was not making any progress towards a question (e.g., continuing to say it the same way), they were offered progressively more significant hints by the system or by a researcher. Third, if more than three minutes passed without a child arriving at an answer to a question or the child was getting increasingly dejected, the Wizard would "lower the bar," giving an answer in situations where the protocol would otherwise require the system to request additional clarification. Fourth, there were two "levels" of questions (with the second question labeled "bonus"). If a participant was able to arrive at an answer to the first question within one minute of using the system, they were presented with a harder bonus question for the same interface.

Table 1. Questions and "bonus" questions given as tasks.

Three Initial Questions
The biggest Pig in the history of the MN State Fair was Reggie the Pig in 2010. How much did he weigh?
500 thousand corn dogs were sold at the MN State Fair in 2016. Were more or fewer mini donuts sold there that year?
The oldest ride at the MN State Fair is "Ye Old Mill." What year was it new to the MN State Fair?
Three "Bonus" Questions
Where can you find the newest ride at the MN State Fair?
Did more people attend the MN State Fair in 2015 or 2016?
Which State Fair is older, the MN State Fair or TX State Fair?

The questions were presented to participants on a sheet of paper and they were asked to write a response once they arrived at an answer with the voice systems. The easier first question typically provided some context and then a question that referred to that context (see Table 1). To arrive at an answer, participants had to ask their question in a way that integrated the provided context. For example, for the second question in Table 1, the participants had to ask the system about the number of mini-donuts sold in a particular year and compare the amount to the one stated in the question. To answer the more difficult "bonus" questions, the participants had to decompose a given question into two parts. For example, for the first "bonus" question in Table 1, the participant had to first ask what the newest ride was and then ask where that ride was located at the fair. The six questions were always presented in the same order, however the order of the interfaces used was counterbalanced (Latin square). This was done to control for the role of both

Table 2. Descriptive statistics of the comparisons between conditions for children and adults. Average hints and exchanges were calculated across the first question in each condition (as that was the one that was completed by all participants).

		Single Condition Results			Grouped by Personification		Grouped by Personalization	
		Voice Search	Fraga	Swali	Non-Personified	Personified	Non-Personalized	Personalized
Children	Preferred by # (# expected if random)	17 (27)	34 (27)	31 (27)	17 (27)	65 (55)	51 (55)	31 (27)
	Avg. Hints	0.87	0.96	1.02	0.87	0.99	0.95	1.02
	Avg. Exchanges	2.73	2.85	3.10	2.73	2.98	2.79	3.10
	Got to Bonus?	33%	32%	26%	33%	29%	33%	26%
Adults	Preferred by (expected if random)	5 (7)	6 (7)	11 (7)	5 (7)	17 (15)	11 (15)	11 (7)
	Avg. Rating	3.63	3.40	3.80	3.63	3.60	3.52	3.80
	Avg. Hints	0.04	0.12	0.12	0.04	0.12	0.08	0.12
	Avg. Exchanges	2.00	2.46	2.23	2.00	2.35	2.23	2.23
	Got to Bonus?	85%	61%	75%	85%	68%	73%	75%

order effects and natural variations in question difficulty on participant preferences.

After each question, children were asked to rate the voice system they used on the smiley-o-meter scale [51]. As suggested in previous work, we used this scale as an opportunity for children to pause and reflect on the experience rather than as a reliable metric of preference [51,57]. Adults specified their ratings on a similar 5-point scale. After trying the three interfaces, we asked all participants to pick their favorite interface of the three (a three-way variation of the "This or That" method [57]), explain why they liked it, and ask it one other question on any topic. All questions and sections of the study were optional—for each question, we only report results for the subset of participants who answered it. At the end of the study, parents and children were given a choice of a university-branded drawstring pack or a small stuffed animal as compensation for their time.

RESULTS

In this section, we describe the analysis and discuss our findings to each of our three guiding research questions.

RQ1: How do children restructure informational queries towards a speech interface?

We reviewed the logs and audio recordings of all the participants, coding the exchanges initiated by each participant to answer each question. Since our IRB required that all children arrive at an answer by the end of each interaction, we had to use an alternative measure of effectiveness than answer accuracy. To estimate this effectiveness, we coded the number of hints and/or prompts participants required to reach an answer (which could come from the system, the researcher, or the parent when the child was stuck). We also noted when a child attempted the bonus question (meaning that they arrived at the answer to the first question within one minute) as a signal of success with the interface. Table 2 provides the descriptive statistics of each of these measures across conditions. Generally, children struggled with the reformulation task, requiring one or more hints and with less than half of the children reaching the "bonus" question. These difficulties reduced with age, with the chance of getting to the bonus question significantly increasing (*$r = 0.30$, $p = 0.006$*) and the number of needed hints significantly decreasing (*$r = -0.33$, $p = 0.004$*).

A team of four researchers reviewed the recorded audio of each interaction, taking notes, transcribing (example transcripts available in supplementary materials), and describing each instance of a reformulation. Thus four researchers transcribed, memoed, and open coded each transcript individually following the process described by Lofland et al [35]. Then the four researchers took part in a workshop led by the lead author to arrive at clusters of codes through multiple rounds of constant comparison, using "abductive" analysis [40] to develop our categories. Once our codebook was developed, we each reviewed all the transcripts again, applying those codes to the original dataset. The following categories of reformulations emerged through this data-driven inductive process:

Table 3. Prevalence of categories of query reformulation by age. We excluded 3 children whose audio data was incomplete (one or more condition was inaudible) and 5 children who did not make any independent reformulations (all reformulations were prompted). In this study, starred categories helped participants get closer to an answer; others were not effective.

	Children, 5–7 (N = 19)	Children, 8–12 (N = 58)	Adults (N = 27)
Off-Course	5%	9%	0%
Restate	53%	64%	30%
Substitute Words	53%	52%	63%
Reorder	37%	53%	37%
State Context	26%	22%	30%
Expand Pronouns*	32%	62%	70%
Add Context*	58%	66%	44%

Off-Course – changing the question to something relevant to the topic that the system *can* answer, but that does not get the asker closer to the target answer. Examples:

> *Child repeatedly asks "Was Reggie the Pig the fattest pig in 2010 at the MN State Fair?" receiving "Yes" as a response from the system but not knowing how to follow up about the pig's weight.*

> *Child gets frustrated with repeatedly getting the response "That question is too complicated for me" and changes the topic, asking the system: "Who is the strongest superhero?"*

Restating or Repeating – changing how a question is pronounced or emphasized, without changing any words or structure of the question. Examples:

> *System asks the child to "Please, ask the question in a different way." Child sings the question to the system.*

> *System asks the child to "Please, ask the question in a different way." Child repeats question louder.*

Substituting Words – changing a word or phrase in a question without adding any additional information or removing any complexity from the previously stated question. Examples:

> *Child rereads the whole question about which State Fair is older, substituting "which is younger" for "which is older."*

> *Child rephrases question as "How many pounds was Reggie the Pig?"*

Reordering – reordering the components of a question without adding any additional information or removing any complexity. Examples:

> *Child rephrases question as "True or False: the MN State Fair is older than the Texas State Fair."*

> *Child rephrases question as "How many fewer mini-donuts were there sold than corn dogs in 2016?*

Stating Context – adding additional information before asking a question, but phrasing this addition as keywords or statements (not integrated into the question). Examples:

> *Child states "Reggie was the fattest pig in 2010 at the MN State Fair. How many pounds was he?"*
>
> *Child states "The Great Big Wheel is the newest ride at the MN State Fair. Where is it located?"*

Expanding Pronouns in Question – replacing the non-specific pronoun in a question with a specific noun. Examples:

> *Child expands the pronoun "he" to "How much did Reggie the Pig weigh?"*
>
> *Child expands the pronoun "it" to "What year was Ye Old Mill first introduced?"*

Adding Context Phrases – adding context following or preceding the noun in a question to narrow to the specific case of interest. Examples:

> *Child asks the questions one at a time until the system is able to answer: "How much did Reggie the pig weigh?" "How much did Reggie the pig from the State Fair weigh?" "How much did Reggie the pig ... from 2010 ... from the fair ... from the MN State Fair weigh?"*
>
> *Child first asks about the number of mini donuts and corn dogs without specifying a year or location. System asks him to be more specific. Child expands: "At the MN fair were there ... in 2016 ... were there more or less mini donuts sold than corn dogs?"*

It is important to note that in this study, the first five types of reformulations did not help the participants arrive at an answer. However, when interacting with other speech interfaces, some of these may be helpful strategies. Table 3 reports the percent of children who employed each reformulation. The most common unsuccessful reformulations were restating without changing the question (53% and 64%, among 5–7 year-olds and 8–12 year-olds respectively) and changing the words in the question without changing its structure (53% and 52% in the two age buckets). The most common reformulation strategies that got the participant closer to the answer included adding context into the question (58% and 66% in the two age buckets) and expanding the pronouns in a question (32% and 62% in the two age buckets). Most children tried a number of different strategies (on average, three or more), but many were stuck in a single strategy until a hint or prompt was provided.

After the children had the opportunity to use each interface, they could ask one more question on any topic to the interface of their choice. We categorized these based on the structure and assumed intent of the question. While children were told that they could ask a question about anything at all, many seemed to be quite influenced by the previous tasks (this was expected given previous preliminary work [53]). 32% of their questions were quantitative questions about other state fair topics (e.g., "*What year did the Ferris wheel open?*") and another 34% were quantitative questions unrelated to the state fair (e.g., "*Are there more dogs in the world than cats?*"). Qualitative questions (e.g., "*How do staplers work?*") accounted for 8% of the dataset. Two other interesting categories emerged. In 18% of the questions, children wanted to learn more about the experiences and interests of the interface (e.g., "*What is your favorite football team?*"). In 9% of the questions, children tried to test the interface (e.g., "*What color bucket is under you?*" or "*What is zero divided by zero?*").

RQ2: How does a speech interface's personification and naming personalization affect children's experience?

We asked each child to interact with all three of the interfaces, thereby gauging their response to personified and personalized systems. Table 2 provides descriptive statistics of metrics gathered in each of the conditions. We compared children's preferences for personified vs. non-personified conditions using Chi-Square Goodness-of-Fit test, finding that children did indeed prefer personified interfaces at a greater and statistically significant rate ($p = 0.015$). A similar comparison between personalized and non-personalized conditions was not statistically significant. We did not observe a relationship between the child's age and their interface preference, though it is possible that one could emerge in a study with more participants of each age.

Effectiveness measures were similar across the conditions. When compared using a Repeated Measured ANOVA, the difference in the number of hints required and exchanges made across conditions was not statistically significant. Similarly, comparing the number of children getting to the bonus question in each condition versus an even distribution across conditions using the Chi-Square Goodness-of-Fit test did not show a statistically significant difference. Therefore, we conclude that personification and naming personalization do not influence children's effectiveness with speech interfaces in a statistically significant way.

We asked each child about why they liked a particular interface best. Of those that could give an answer beyond the tautological (e.g., "I liked it better"), 62% said that they thought that their favorite interface understood them better or was less confused about their questions (this was mostly an order effect—children developed more effective reformulation strategies by the third condition). Eleven percent of the answers were related to irrelevant surface consideration (e.g., favorite color), however removing these responses from the analysis did not change the direction or magnitude of the result. Two children mentioned liking the personalized interface best because of its personality ("*more friendly*" and "*polite*") and another four children explicitly mentioned liking the interface that "*knew my name.*" On the other hand, another four children were explicitly turned off by the naming personalization saying that it was "*creepy*." One of these children loudly exclaimed, *"Are you stalking me?!"* when the system referred to him by name and age. However, it is important to interpret this reaction in context—this information was solicited by the researcher not directly by the robot, so this reaction may be moderated in other arrangements. Fewer children provided interface-specific reasons for liking the non-personified one, though one did mention liking it because it "*spoke faster and was more efficient.*" While the qualitative experience for this individual was more efficient with the non-personified interface and there was a descriptive difference between the conditions in this direction, it was not statistically significant so it does not generalize across our sample.

When given the opportunity to ask their favorite interface any question they wanted, it was interesting to note that questions about the interests and experiences of the interface and qualitative questions were directed almost exclusively (all but one) towards the personified conditions.

RQ3: How do the experiences and preferences of adults compare to those of children?

As can be expected, adults faced fewer struggles with question reformulations, with only a few requiring hints and with the majority reaching the bonus question (see Table 2). We did not see any examples where adults went so off-course in their reformulations that they asked questions that did not get them closer to the answer (though in two cases, adults asked additional irrelevant questions after completing their task). Fewer adults (30%) than children (53% or 64%, depending on age) focused on the unsuccessful strategy of simply restating the question with a different emphasis or pronunciation. Adults employed some successful strategies more often than children. For example, older participants were more likely to expand pronouns (32% of 5–7-year-olds, 62% of 8–12-year-olds, and 70% of adults). They were less likely to reformulate the question with additional context, but only because many included all the necessary context phrases from the first formulation.

Adult preferences for naming personalization and personification were not statistically significantly different from random when tested with the Chi-Square Goodness-of-Fit test ($p = 0.095$). There was no statistically significant difference between their ratings of each condition on a 5-point Likert-type scale when compared with a Repeated Measures ANOVA. Like the children's results, there was no statistically significant differences between conditions for adults' effectiveness with the systems in terms of the number of hints required, number of exchanges, or the likelihood of getting to the bonus question. While descriptively, adults seemed to prefer the naming personalization condition at greater rates than children, the difference in preference distributions between adults and children was not statistically significant when compared with a Chi-Square test.

Fourteen of the 22 adults who had a preferred interface provided a reason for that selection. Interestingly, the most common (36%) reason explicitly referred to liking the "*personality*" of the naming personalization interface best, describing it as *"like a friend,"* "*witty,"* and "*more personal.*" However, one adult did mention being "*creeped out*" and disliking the same interface for saying their name (even though they understood that they provided their name to the system). One adult also mentioned that they liked the non-personified interface best because of its efficiency (*"not a lot of extra talking"*). Finally, adults generally recognized that all three interfaces understood them equally poorly and were more likely to reflect on the role of order in their system preferences, acknowledging that they became better at asking questions by the end of the study.

While almost all children took the opportunity to ask their favorite interface another question, only 13 out of the 27 adults did so. As with the child participants, 38% of these were quantitative questions related to the state fair and another 38% were quantitative questions on other topics. None of the adults asked qualitative questions or attempted to learn more about the experiences and interests of the interface. However, as children did, 23% of the adults did try to ask questions that tested the interface, usually by asking a question where they already knew the answer (e.g., *"Who was the first president of the United States?"*).

DISCUSSION

In this section, we discuss the implications of our findings on design and research in speech interfaces for families.

Limitations and Future Directions

All study designs have inherent limitations. We had to make specific decisions regarding the wording of questions, responses, and specific operationalization of concepts like "personification" and "naming personalization." Children may have personified the non-personified condition despite the language used by the system, as previous work shows that personal pronouns are not necessary for perspective-taking [2]. Similarly, personalization can be much more nuanced and useful than merely referring to a person by their first name [16]. For example, it would be a useful personalization to use the participants' location to provide context for the questions or to tailor the systems responses based on a child's ability. Given that many of the concerns children expressed against personalization were privacy-based, the privacy theory of "proportionality" [23] suggests that some of those concerns would be ameliorated if the personalization provided a functional benefit in interpreting and answering questions. Also, it may not have been transparent to the child how the systems learned their name, since these were entered by the researchers. But, this parallels the real-world situation where a parent would generally provide information for a child's voice assistant account. Certainly, we do not consider our investigation to be the final word on personalization of speech interfaces, but rather a point of evidence towards the idea that a parent entering a child's name into a speech system does not support a personalization that provides measurable value to children.

We made study design choices that allowed us to collect data from a large, diverse sample in a field setting. While this has many advantages, it also introduced several limitations. First, our study was not tightly controlled or in a lab setting. This may mean that there were times when siblings, parents, etc. distracted the children from their tasks. However, this also represents a more ecologically valid situation. Second, we provided most of the questions asked by the participants. We selected questions that were challenging and that would necessitate reformulation (similar to the approach of previous work on children's online search, e.g., [9]). We picked questions that were relevant to the setting and similar in style to the kinds of questions children may be asked to answer on a homework worksheet or similar assignment. However, we do not know to what extent these compound complex questions appear in the lexicon of children's interactions with speech interfaces in the wild.

Finally, each participant had a relatively brief interaction with a speech interface that belonged to the researcher. There are two possible biases that this introduced in the dataset. First, we could not observe long-term learning effects. We did observe that participants learned from earlier interactions and their performance improved even in the short time of the study. It is important to conduct a follow-up study with a home or mobile speech device in the wild, to understand which reformulations remain problematic in the long run and which are quicker adaptations. Second, privacy concerns may have been exacerbated by interacting with a speech assistant that belonged to the researcher, rather than a personal device. It is possible that some of the "creepiness" factor of naming personalization may have been ameliorated if participants interacted with their own personal speech device. This was not possible in the context of our controlled investigation, but would be an important and promising follow-up study if participants could be given personal devices to keep and interact with long-term.

Considerations for User-Friendly Speech Interfaces

Observing the struggles and relatively low success rate of our child participants, we reflect that current speech interfaces may be considered poor designs from a classic design perspective (e.g., [46]). They provide no constraints on the kinds of statements that may be directed to them, leading to the observed high number of exchanges before arriving at an answer. They provide little feedback regarding what they heard and understood, leading most children to spend significant time on phonetic reformulations. They give almost no visibility as to why the system may be struggling with a particular question (e.g., Amazon Alexa simply answers "I don't know" if it is confused by any part of the question). Many of our participants struggled but took reasonable paths—first assuming that a question was misheard, then assuming that a word was not understood, before moving on to more effective strategies. However, the process was undeniably frustrating and many of the children required help to get past the first two strategies and to arrive at questions that would yield an appropriate answer.

Currently available systems rarely (if ever) provide hints or clarification. Based on each of the common reformulation types and the types of hints and prompts that were most useful to the participants in our study, we recommend that systems integrate the following types of clarifications:

- Restate what was heard if the system fails to meet a confidence threshold. Children are used to being misheard by speech interfaces and most of them focused on phonetic reformulations. With feedback that the system heard the question correctly, they may proceed to semantic reformulations.
- If a particular piece of context is needed but missing (e.g., year, identity of a particular pronoun), follow up requesting this information (e.g., "What is "it" in your question?). It may also be reasonable to make a "best guess" about missing information from previous questions or by assuming current location, year, etc.
- If a part of the question is known but the entire question is too complex, provide some information on what is known and request clarification for the rest (e.g., "I know that the newest ride at the fair is the 'The Great Big Wheel,' what would you like to know about it?").
- If a particular style of question is difficult for the system, clarify this and provide guidance (e.g., "Comparisons are hard for me. Can you ask about each part of your question separately?").

Additionally, sometimes a significant amount of contextual information is needed to provide the correct answer to a question. It can be difficult and awkward for children to construct complex questions that integrate all of the necessary components. While many children (58%-66%) eventually attempted integrating contextual information directly into the question, they struggled with the compound sentences that resulted, pausing after each informational element. We also saw that stating information up-front rather than integrating it directly into the question was a strategy employed by both adults (30%) and children (22%-26%). Speech systems could become more usable in that regard by allowing contextual information for a particular question to be provided in the form of a sentence or multiple sentences. For example, instead of having to ask a system "What will the weather be like this Thursday and Friday in Trondheim?" it may be easier for users to give upfront context ("I am traveling to Trondheim on Thursday.") followed by the particular question ("What will the weather be like?").

Costs and Benefit of Personifying & Personalizing

Previous work in the field of speech interfaces had provided divergent findings regarding the potential roles of personification and personalization. This investigation helped address some aspects of this open question.

None of the participants had specific objections to personification of interfaces, though several participants did note that non-personified interfaces covered less non-task content and thus were more efficient at quickly providing an answer. Children responded to personification, preferring personified conditions to the non-personified. Children also seemed more willing to direct a broader variety of questions, including qualitative ones, to personified interfaces. Whether because of actual confusion or as a playful act, children asked the personified interfaces question about their experience and preferences (echoing previous studies [3]). This may not be entirely positive, as it is difficult to balance engaging in the play to provide a "fun" answer and answering questions honestly without misleading the child as to the nature of the interaction (e.g., if asked about its favorite food, should the system lie or should it say that it does not eat?). It is worth noting that the same consideration need not be extended to adults—we saw no examples of adults asking these kinds of questions. While not stated as a concern by any participants, researchers in other domains have noted that increased personification of interfaces may create a "robotic moment," in which children may become confused the role and agency of humans compared to machines [54]. This may be a trade-off of personification if empirically confirmed in future research.

We did not find statistically significant evidence favoring naming personalization as operationalized in our study. One common objection cited by several children and one adult was that it was a violation of privacy expectations to have the speech device know specifics like their age and name, even though they themselves provided this information to the researcher. This objection complements previous investigations of children interacting with robotic agents [34]. On the other hand, some children and adults did find the personalized system to be more friendly, witty, and polite. This may be an individual difference. However, it is also worth noting that even the kind of surface naming personalization mentioned in this study may not be easy to accomplish in the field. First, it may require interfaces to distinguish between users based on voice, which is possible but not trivial [43]. Second, it is increasingly common for children to have names that are non-traditional, unusually spelled, or names from non-western cultures. This increases the likelihood that a speech system will mispronounce a child's name, which may reverse perceptions of friendliness and politeness. Given that benefits are not evidenced in this investigation, naming personalization should not be a high priority feature. However, it is again important to reiterate that there are many other possible types of personalization beyond "naming," which may provide a different set of costs and benefits to the participants.

Building for the Whole Family

One of the surprising descriptive findings of our study is that 93% of children had used one or more speech interfaces prior to the study and the plurality of these children used such interfaces multiple times every day. While we thought of this interface modality as "emerging," it became clear that it was already well-entrenched in the lives of children.

There are three ways children typically interact with speech interfaces. Some children may have their own smart phone or tablet. In this case, it may make sense to default to a personified interface, letting the child control their level of personalization. Other children may experience so-called "pass-back"—the opportunity to interact with speech interfaces on their parents' smart phone. It may be useful for such devices to detect children's voices (e.g., [44]) and remove any personalization features aimed at the parent while retaining personification. Third, children may live in a home that has a household speech information appliance (e.g., Google Home). Such devices need to account for the specific preferences of multiple users [2]. Our study complements previous investigations showing that interface personification preference may be more related to individual differences than age [56]. We did not see specific age effects found in other previous work (e.g., older children considering personification to be too childish) [20]. There was a substantial split regarding personalization and there were also a fair number of participants who preferred the non-personified condition (i.e., a family of four is not unlikely to have one person in this category). It is an open question whether it would provide the best experience for the whole family to adapt to individual preferences or whether families would prefer to have a consistent experience even if some individuals' preferences are violated.

CONCLUSION

Most children in our study used voice assistants and many of them used such interfaces multiple times every day. Informational queries are a common use case, but prior to this investigation little was known about how children formulate questions towards speech interfaces. Our work addresses this gap by observing 87 children asking questions of three variations of Wizard-of-Oz speech interfaces. We found that children struggled with reformulating questions, with most of them requiring hints to complete the task. Though most children eventually tried effective reformulations such as substituting objects for pronouns and providing context within the question, many children first began with surface reformulations such as repeating the question or substituting synonym words. Older children and adults were more effective than younger children at informational query reformulation. By comparing variations of the interface, we discovered that children preferred personified interfaces, but showed no preference towards the interface that was personalized with their name and age. We suggest several considerations for future speech interfaces. First, personified interfaces are well indicated, while naming personalization (especially, when that name is provided by others) is not. Second, we point to five strategies to support more effective reformulations: providing feedback on what was heard, asking for missing context, clarifying what is known, specifying formulations that are difficult for the system, and allowing context to be provided as a statement.

ACKNOWLEDGMENTS

We gratefully acknowledge our funding from a 2017H2 Mozilla Research Grant and a 2017 Google Faculty Research Award. We also thank the D2D State Fair team.

SELECTION AND PARTICIPATION OF CHILDREN

This study was reviewed and approved by a University IRB. We invited families with children between the ages of 5 and 12 to participate at the MN State Fair event. A researcher explained the study and its risks to both the child and parent and answered any questions. If interested, parents signed informed consent and a parental permission form. A researcher read a paper assent form out loud to each child. If assenting to the study, the child wrote their name on the form (as able). Both parental permission *and* the child's explicit assent were necessary for the study to proceed and either person could ask us to stop the study at any time (the child still received a toy for their help).

REFERENCES

1. Nadia Aram. 2017. Hey, Alexa: What's New in Children's Privacy?... FTC Updates COPPA Guidance. Retrieved August 28, 2017 from http://www.wcsr.com/Insights/Alerts/2017/June/Hey-Alexa-Whats-New-in-Childrens-Privacy-FTC-Updates-COPPA-Guidance
2. Tad T. Brunyé, Tali Ditman, Grace E. Giles, Amanda Holmes, and Holly A. Taylor. 2016. Mentally simulating narrative perspective is not universal or necessary for language comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 42, 10: 1592–1605. https://doi.org/10.1037/xlm0000250
3. A. J. Bernheim Brush and Kori M. Inkpen. 2007. Yours, Mine and Ours? Sharing and Use of Technology in Domestic Environments. In *Proceedings of the 9th International Conference on Ubiquitous Computing* (UbiComp '07), 109–126. Retrieved December 27, 2016 from http://dl.acm.org/citation.cfm?id=1771592.1771599
4. 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
5. Theodora Chaspari, Samer Al Moubayed, and Jill Fain Lehman. 2015. Exploring Children's Verbal and Acoustic Synchrony: Towards Promoting Engagement in Speech-Controlled Robot-Companion Games. In *Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And infLuence* (INTERPERSONAL '15), 21–24. https://doi.org/10.1145/2823513.2823518
6. Yun-Nung Chen, Ming Sun, Alexander I. Rudnicky, and Anatole Gershman. 2015. Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (ICMI '15), 83–86. https://doi.org/10.1145/2818346.2820781
7. Christopher K. Cowley and Dylan M. Jones. 1993. Talking to Machines (Abstract). In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (CHI '93), 522–. https://doi.org/10.1145/169059.169512
8. Dan Bohus and Alexander I. Rudnicky. 2003. RavenClaw: Dialog Management Using Hierarchical Task Decomposition and an Expectation Agenda. Retrieved from http://repository.cmu.edu/compsci/1392/
9. Steven P. Dow, Manish Mehta, Blair MacIntyre, and Michael Mateas. 2010. Eliza Meets the Wizard-of-oz: Blending Machine and Human Control of Embodied Characters. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10), 547–556. https://doi.org/10.1145/1753326.1753408
10. Allison Druin, Elizabeth Foss, Leshell Hatley, Evan Golub, Mona Leigh Guha, Jerry Fails, and Hilary Hutchinson. 2009. How Children Search the Internet with Keyword Interfaces. In *Proceedings of the 8th International Conference on Interaction Design and Children* (IDC '09), 89–96. https://doi.org/10.1145/1551788.1551804
11. Allison Druin, Elizabeth Foss, Hilary Hutchinson, Evan Golub, and Leshell Hatley. 2010. Children's Roles Using Keyword Search Interfaces at Home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10), 413–422. https://doi.org/10.1145/1753326.1753388
12. Sergio Duarte Torres, Djoerd Hiemstra, and Pavel Serdyukov. 2010. An Analysis of Queries Intended to Search Information for Children. In *Proceedings of the Third Symposium on Information Interaction in Context* (IIiX '10), 235–244. https://doi.org/10.1145/1840784.1840819
13. Sergio Duarte Torres, Ingmar Weber, and Djoerd Hiemstra. 2014. Analysis of Search and Browsing Behavior of Young Users on the Web. *ACM Trans. Web* 8, 2: 7:1–7:54. https://doi.org/10.1145/2555595
14. Carsten Eickhoff, Pieter Dekker, and Arjen P. de Vries. 2012. Supporting Children's Web Search in School Environments. In *Proceedings of the 4th Information Interaction in Context Symposium* (IIIX '12), 129–137. https://doi.org/10.1145/2362724.2362748
15. Pedro Fialho and Luísa Coheur. 2015. ChatWoz: Chatting Through a Wizard of Oz. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (ASSETS '15), 423–424. https://doi.org/10.1145/2700648.2811334
16. FitzGerald Elizabeth, Kucirkova Natalia, Jones Ann, Cross Simon, Ferguson Rebecca, Herodotou Christothea, Hillaire Garron, and Scanlon Eileen. 2017. Dimensions of personalisation in technology-enhanced learning: A framework and implications for design. *British Journal of Educational Technology* 49, 1: 165–181. https://doi.org/10.1111/bjet.12534
17. D. Giuliani and M. Gerosa. 2003. Investigating recognition of children's speech. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)*, II-137–40 vol.2. https://doi.org/10.1109/ICASSP.2003.1202313
18. Michal Gordon and Cynthia Breazeal. 2015. Designing a Virtual Assistant for In-car Child Entertainment. In *Pro-*

ceedings of the 14th International Conference on Interaction Design and Children (IDC '15), 359–362. https://doi.org/10.1145/2771839.2771916

19. Tatiana Gossen, Thomas Low, and Andreas Nürnberger. 2011. What Are the Real Differences of Children's and Adults' Web Search. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR '11), 1115–1116. https://doi.org/10.1145/2009916.2010076
20. Tatiana Gossen, Rene Müller, Sebastian Stober, and Andreas Nürnberger. 2014. Search Result Visualization with Characters for Children. In *Proceedings of the 2014 Conference on Interaction Design and Children* (IDC '14), 125–134. https://doi.org/10.1145/2593968.2593983
21. A. Hagen, B. Pellom, and R. Cole. 2003. Children's speech recognition with application to interactive books and tutors. In *2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No.03EX721)*, 186–191. https://doi.org/10.1109/ASRU.2003.1318426
22. Hannaneh Hajishirzi, Jill F. Lehman, and Jessica K. Hodgins. 2012. Using Group History to Identify Character-directed Utterances in Multi-child Interactions. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (SIGDIAL '12), 207–216. Retrieved from http://dl.acm.org/citation.cfm?id=2392800.2392838
23. Giovanni Iachello and Gregory Abowd. 2005. Privacy and proportionality: adapting legal evaluation techniques to inform design in ubiquitous computing. In *Proc. of CHI*, 91–100.
24. Pradthana Jarusriboonchai, Thomas Olsson, and Kaisa Väänänen-Vainio-Mattila. 2014. User Experience of Proactive Audio-based Social Devices: A Wizard-of-oz Study. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia* (MUM '14), 98–106. https://doi.org/10.1145/2677972.2677995
25. Jiepu Jiang, Wei Jeng, and Daqing He. 2013. How Do Users Respond to Voice Input Errors?: Lexical and Phonetic Query Reformulation in Voice Search. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR '13), 143–152. https://doi.org/10.1145/2484028.2484092
26. Oliver Jokisch, Horst-Udo Hain, Rico Petrick, and Rüdiger Hoffmann. 2009. Robustness Optimization of a Speech Interface for Child-directed Embedded Language Tutoring. In *Proceedings of the 2Nd Workshop on Child, Computer and Interaction* (WOCCI '09), 10:1–10:4. https://doi.org/10.1145/1640377.1640387
27. Yvonne Kammerer and Maja Bohnacker. 2012. Children's Web Search with Google: The Effectiveness of Natural Language Queries. In *Proceedings of the 11th International Conference on Interaction Design and Children* (IDC '12), 184–187. https://doi.org/10.1145/2307096.2307121
28. 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
29. Sawit Kasuriya and Alistair D. N. Edwards. 2009. Pilot Experiments on Children's Voice Recording. In *Proceedings of the 2Nd Workshop on Child, Computer and Interaction* (WOCCI '09), 13:1–13:5. https://doi.org/10.1145/1640377.1640390
30. Lin-shan Lee, James Glass, Hung-yi Lee, and Chun-an Chan. 2015. Spoken Content Retrieval: Beyond Cascading Speech Recognition with Text Retrieval. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.* 23, 9: 1389–1420. https://doi.org/10.1109/TASLP.2015.2438543
31. Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2010. Receptionist or Information Kiosk: How Do People Talk with a Robot? In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (CSCW '10), 31–40. https://doi.org/10.1145/1718918.1718927
32. Jill Fain Lehman and Iolanda Leite. 2017. Turn-Taking, Children, and the Unpredictability of Fun. *AI Magazine* 37, 4: 55–62. https://doi.org/10.1609/aimag.v37i4.2685
33. Iolanda Leite, Hannaneh Hajishirzi, Sean Andrist, and Jill Lehman. 2013. Managing Chaos: Models of Turn-taking in Character-multichild Interactions. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (ICMI '13), 43–50. https://doi.org/10.1145/2522848.2522871
34. 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
35. John Lofland, David A. Snow, Leon Anderson, and Lyn H. Lofland. 2005. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. Cengage Learning, Belmont, CA.
36. 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* (IDC '15), 335–338. https://doi.org/10.1145/2771839.2771910
37. 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* (CHI '16), 5286–5297. https://doi.org/10.1145/2858036.2858288
38. Matt Marx and Chris Schmandt. 1994. Putting People First: Specifying Proper Names in Speech Interfaces. In *Proceedings of the 7th Annual ACM Symposium on User*

Interface Software and Technology (UIST '94), 29–37. https://doi.org/10.1145/192426.192439

39. Jack Mostow and Steven F. Roth. 1995. Demonstration of a Reading Coach That Listens. In *Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology* (UIST '95), 77–78. https://doi.org/10.1145/215585.215665
40. Michael Muller. Curiosity, Creativity, and Surprise as Analytic Tools: Grounded Theory Method. In *Ways of Knowing in HCI*, Judith S. Olson and Wendy A. Kellogg (eds.). Springer, 25–48.
41. S. Narayanan and A. Potamianos. 2002. Creating conversational interfaces for children. *IEEE Transactions on Speech and Audio Processing* 10, 2: 65–78. https://doi.org/10.1109/89.985544
42. Clifford Nass and Li Gong. 2000. Speech Interfaces from an Evolutionary Perspective. *Commun. ACM* 43, 9: 36–43. https://doi.org/10.1145/348941.348976
43. R. Nisimura, A. Lee, H. Saruwatari, and K. Shikano. 2004. Public speech-oriented guidance system with adult and child discrimination capability. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, I-433–6 vol.1. https://doi.org/10.1109/ICASSP.2004.1326015
44. R. Nisimura, A. Lee, H. Saruwatari, and K. Shikano. 2004. Public speech-oriented guidance system with adult and child discrimination capability. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, I-433–6 vol.1. https://doi.org/10.1109/ICASSP.2004.1326015
45. Don Nix, Peter Fairweather, and Bill Adams. 1998. Speech Recognition, Children, and Reading. In *CHI 98 Conference Summary on Human Factors in Computing Systems* (CHI '98), 245–246. https://doi.org/10.1145/286498.286730
46. Donald A. Norman. 1990. *The Design of Everyday Things*. Basic Books, New York.
47. A. Oulasvirta, K. P. Engelbrecht, A. Jameson, and S. Moller. 2007. Communication failures in the speech-based control of smart home systems. In *2007 3rd IET International Conference on Intelligent Environments*, 135–143. https://doi.org/10.1049/cp:20070358
48. Sharon Oviatt, Courtney Darves, and Rachel Coulston. 2004. Toward Adaptive Conversational Interfaces: Modeling Speech Convergence with Animated Personas. *ACM Trans. Comput.-Hum. Interact.* 11, 3: 300–328. https://doi.org/10.1145/1017494.1017498
49. Aasish Pappu and Alexander Rudnicky. 2013. Deploying Speech Interfaces to the Masses. In *Proceedings of the Companion Publication of the 2013 International Conference on Intelligent User Interfaces Companion* (IUI '13 Companion), 41–42. https://doi.org/10.1145/2451176.2451189
50. A. Potamianos and S. Narayanan. 2003. Robust recognition of children's speech. *IEEE Transactions on Speech and Audio Processing* 11, 6: 603–616. https://doi.org/10.1109/TSA.2003.818026
51. Janet C. Read and Stuart MacFarlane. 2006. Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proc. of IDC*, 81–88.
52. Ben Shneiderman and Pattie Maes. 1997. Direct Manipulation vs. Interface Agents. *interactions* 4, 6: 42–61. https://doi.org/10.1145/267505.267514
53. Lisa Stifelman, Adam Elman, and Anne Sullivan. 2013. Designing Natural Speech Interactions for the Living Room. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '13), 1215–1220. https://doi.org/10.1145/2468356.2468574
54. Sherry Turkle. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other*.
55. J. G. Wilpon and C. N. Jacobsen. 1996. A study of speech recognition for children and the elderly. In *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, 349–352 vol. 1. https://doi.org/10.1109/ICASSP.1996.541104
56. Maria Wolters, Kallirroi Georgila, Johanna D. Moore, and Sarah E. MacPherson. 2009. Being Old Doesn'T Mean Acting Old: How Older Users Interact with Spoken Dialog Systems. *ACM Trans. Access. Comput.* 2, 1: 2:1–2:39. https://doi.org/10.1145/1525840.1525842
57. Bieke Zaman, Vero Vanden Abeele, and Dirk De Grooff. 2013. Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That method. *International Journal of Child - Computer Interaction* 1, 2: 61–70. http://dx.doi.org/10.1016/j.ijcci.2012.12.001