Why Doesn't It Work? Voice-Driven Interfaces and Young Children's Communication Repair Strategies

Yi Cheng[1], Kate Yen[2], Yeqi Chen[3], Sijin Chen[4], Alexis Hiniker[1]

1. Information School, 2. Department of Sociology, 3. Department of Electrical Engineering, 4. Human Centered Design and Engineering

University of Washington, Seattle, WA, USA

{chengy23, yenk, yeqic, sijinc, alexisr}@uw.edu

ABSTRACT

In this study, we examine the conversational repair strategies that preschoolers use to correct communication breakdowns with a voice-driven interface. We conducted a two-week deployment in the homes of 14 preschoolers of a tablet game that included a broken voice-driven mini-game. We collected 107 audio samples of these children's (unsuccessful) attempts to communicate with the mini-game. We found that children tried a common set of repair strategies, including repeating themselves and experimenting with the tone and pronunciation of their words. Children were persistent, rarely giving up on the interaction, asking for help, or showing frustration. When parents participated in the interaction, they moved through four phases of engagement: first making suggestions, then intervening, then making statements of resignation, and finally pronouncing that the interaction could not be repaired. Designers should anticipate that in this context, children will borrow behaviors from person-to-person communication, such as pivoting strategies to probe the source of failed communication and structuring communication into turn-taking attempts.

Author Keywords

Voice interfaces; preschool; child-computer interaction; speech; communication repair

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

INTRODUCTION

Voice-driven interfaces—including conversational agents, such as Amazon Echo, Siri, and the Google Assistant—are increasingly common in end-user products. Analysts report that 30 million American homes already have a smart speaker with a conversational agent (such as the Echo or Google Home), with rapid adoption expected over the next two years [17]. Early work examining commercially available conversational agents has shown that they are used regularly by children [20], and many smart speakers include features designed specifically for these younger users (e.g., [33]).

Figure 1: A parent and child play a voice-driven mini-game together ("Cookie Monster's Challenge" by Sesame Workshop and PBS Kids). The app tells him, "Quick, say quack out loud!" and the child responds by leaning forward and saying "quack!" to the screen.

Designing a voice-driven interface that works well for children is a complex proposition and requires support for failed communication attempts. Although communication breakdowns are common in conversations among people of all ages, they are particularly common in conversations involving children age 5 and under, who are still acquiring the motor planning skills they need for clear pronunciation, learning the linguistic irregularities of their first language, and expanding their early vocabulary [4]. One study of toddlers' communicative attempts to their mothers showed that children's initial attempts to convey an idea were unsuccessful roughly two-thirds of the time and routinely required children to correct their mother's misunderstanding [8]. Thus, recognizing and correcting communication breakdowns is an essential part of young children's interactions with the world.

The act of working to correct a breakdown in communication is known as *communication repair* [29]. This can involve repeating a request to gain a listener's attention, highlighting a particular aspect of the original statement (e.g., "No, I said what are you *reading*"), or revising disfluencies while speaking (e.g., "What are you eating—I mean reading?"), among other techniques. Although young children's communicative

ACM ISBN 978-1-4503-5152-2/18/06…$15.00
https://doi.org/10.1145/3202185.3202749

attempts are often initially unsuccessful, prior research has shown that children are creative and persistent in using a variety of repair strategies to ensure their listener ultimately understands their intended meaning [8].

As communication repair is essential to young children's successful communication with other people, support for successful repair is likely to be important to young children's interactions with voice-driven interfaces. Prior work has shown that repairing misunderstandings between humans and machines is a fundamental part of human-computer interaction across many form factors, and successful interfaces provide mechanisms for recovery in the inevitable moment when the human and the interface fail to understand each other [25].

In this paper, we examine the repair strategies that young children use in the face of one type of communication breakdown, specifically, when a voice-driven interface does not respond to their communication attempt. We document the common in situ strategies children employ as course corrections, and we examine parents' role in supporting these repairs. To do so, we conducted a two-week field deployment with 14 preschoolers, in which each participant engaged with a voice-driven interface that was intermittently unresponsive due to an unanticipated bug. As a result, we inadvertently collected 107 audio samples of children attempting to engage with an interface that failed to notice or respond to their speech. Though we did not initially set out to conduct this experiment, we happened to collect a large sample of children's repair attempts in the wild. In approximately half of these instances, parents were present or engaging with the game with their child, giving us insight into how parents support children's attempts at conversational repair.

We found that children employed a small set of common strategies to repair the failed interaction, including repeating themselves, pausing for feedback between statements, and varying the volume and tone of their voice. We saw that children borrowed known conversational strategies that are common in person-to-person interactions and applied them to their attempted interaction with the app. And we saw that parents engaged in a predictable series of support activities which often ended with a "pronouncement" that the app was broken (and therefore, that the communication could not be repaired). Children accepted these pronouncements and only gave up on repairing the conversation when parents deemed it beyond help.

Just as touchscreens fundamentally changed the way children engage with technology, voice interfaces offer new possibilities to preliterate children and promise to open up child-computer interaction scenarios that have never before been supported. However, just as children's speech with other people demands routine repair, these interaction patterns are likely to require designed support for revising failed communication. This study demonstrates that children will work persistently with conversational agents, borrowing techniques that they use in person-to-person conversations, to ensure their communication succeeds.

RELATED WORK

Children and Voice-Driven Interfaces

A number of studies have looked at children's conversations with voice-driven interfaces, examining how they interact with robots, digital assistants, virtual agents, and other technologies. For example, Fridin showed that KindSAR, an interactive, social robot, was able to engage preschoolers in interactive story-telling and game-play experiences and could correctly respond to their suggestions [6]. Andrist and colleagues demonstrated the potential for voice-driven interfaces to structure turn-taking in conversations among groups of children [1], which other work has shown is one of the most difficult conversational skills for children under five to master [23]. Tamura and colleagues found that children preferred storytelling activities when they shared them with a listening robot that could respond to their statements and ask questions [27]. Across these and a number of other studies, research has shown that voice-driven interfaces have the potential to offer a powerful new mechanism for engaging, teaching, entertaining, and supporting children in daily life.

Voice-driven interfaces are also increasingly common in commercially available products, including smartphones, tablets, and smart speakers. Though this remains an under-explored space, early work examining adoption habits has shown that these interfaces are used regularly by children [20]. An analysis of YouTube videos depicting children using commercially available voice assistants found that children age 7 and under often struggled to be understood or to make use of the list of search results that they received in response to their speech input [14]. And work with adults has shown that detecting and recovering from breakdowns in communication remains a key challenge for human-robot interaction and other voice-based scenarios in HCI [32].

Though automated speech recognition and natural language processing are well-established fields, and human-robot interaction has explored many facets of children's responses to human-like interfaces, children's interactions with commercially available voice-driven interfaces is still largely unexplored from an HCI perspective. Here, we investigate one component of these interactions by examining children's responses to failed communication attempts and the repair strategies they attempt to use to recover.

Conversational Repair

When people conduct conversations, misunderstandings are common [9]. These breakdowns are corrected by a process of self-repair, wherein the speaker modifies their speech in some way that resolves the disfluency or misinterpretation. "Other-initiated self-repair" refers to self-repairs that are performed in response to a cue from the listener to the speaker that they have misunderstood [5]. Early work on organization of turn-taking for conversation has identified several repair techniques, such as repeating "one-word questions" like "what" and "who," as solutions for dealing with turn-taking errors and violations [23].

Such conversational repair is a common practice among speakers of all ages and abilities, but it is particularly useful for young children mastering their first language (age 0–8) [18]. Children who do not yet speak clearly are highly persistent in working to repair failed interactions, particularly when they want something from the person to whom they are speaking [7]. As children grow, their repair strategies mature [2]; for example, an infant gesturing for milk who receives a cracker might cry or throw the offered cracker to indicate the failed communication, a toddler might correct the misunderstanding by saying "No! Milk!" and an older child might say, "I didn't say *cracker*, I said *milk*." Prior research has identified three overarching categories of repair strategies in human speech: (1) repetition, (2) augmentation, and (3) substitution (where the speaker tries a new form to express the same sentiment) [8].

Repair is also a fundamental part of the interactions that occur between humans and computers [25], and many interfaces intentionally allow the user to correct mistakes, identify how the interface might have misunderstood, and repair discrepancies between the user's intention and the system's interpretation. Prior research has shown that adults intentionally make linguistic changes, such as simplifying their speech or asking questions, when a voice-driven interface provides feedback that suggests it does not understand them [13]. A central goal of research in natural language processing is to allow for conversational repair on the part of both the human and system [9]. Here, we examine how this plays out in practice in one particular context, specifically looking at repairs in child-computer interaction.

Joint Media Engagement

As part of our analysis, we also examined parents' participation in their child's use of this interface. This shared use of a tablet app is one form of Joint Media Engagement (JME) [26], the practice of parents and children engaging together with digital media. Past research suggests that preschoolers and school-age children enjoy engaging in joint experiences with parents more than solitary or parallel play [3], and preschoolers learn more from content when they discuss it with parents [21]. This mutual engagement benefits children of all ages, but may be particularly valuable for children age 3–5 [26].

The most valuable JME experiences involve rich shared engagement with digital content, and they foster shared behaviors like dialogic inquiry or co-creation [26]. However, parent scaffolding can be very useful even in scoped contexts that do not allow for expansive play and creation. Parents model routine interactions with technology for preschoolers and support them in troubleshooting [19]. They also instruct children on the use of specific interface components and help children understand the underlying purpose behind these interactions [10].

Here, we build on prior work examining the ways in which parents scaffold children's use of technology and the ways in which parents scaffold children's conversational repairs during person-to-person conversations. By considering these two bodies of work together, we focus on the ways in which parents may be likely to support their young children's use of voice-driven interfaces.

METHODS

The present study is part of a larger research project to evaluate Cookie Monster's Challenge (CMC), a tablet game created by Sesame Workshop and PBS Kids to support the development of self-regulation and executive function. We conducted a two-week deployment of the game in the homes of 14 preschoolers. Here, we focus on children's interactions during this two-week deployment with one voice-driven mini-game embedded within the larger app. This study was approved by the University of Washington Institutional Review Board.

Participants

Fourteen families participated in our study and completed all procedures. These participants were recruited through an institutional database of local families interested in research, who are recruited through regional birth records. All children were between the ages of 3 and 5 years old (inclusive). Our child participants included 10 boys and 4 girls and were overrepresentative of white, middle-class families; participant demographics are shown in Table 1. We asked parents about their child's average technology use, and all parents reported that their child uses a tablet or smartphone at least a few times a week. We did not explicitly ask about children's ex-

Demographic Variable		N
Child Gender	Male	10
	Female	4
Child Age (Mean = 3.86 years)	Age 3	6
	Age 4	4
	Age 5	4
Child Race	White	11
	Mixed	3
Household Income (US$)	<25 K	1
	25–75K	2
	75–100K	3
	>125K	6
	Prefer not to say	2
Parent Education	Some College	1
	Associate Degree	4
	Bachelor's Degree	4
	Master's Degree	4
	Prefer not to say	1
Parent Marital Status	Married/Partnered	12
	Divorced	2

Table 1: Participant Demographics

perience with voice-driven interfaces. Participants who completed all parts of the larger study received a $50 gift card and kept the tablet computer they used during the study as a thank-you for their participation.

Materials

Cookie Monster's Challenge and the "Quack" Bonus Game

Cookie Monster's Challenge (CMC) is a tablet app created by Sesame Workshop and PBS Kids. The description of CMC on the PBS Kids website [34] explains that the app contains:

> "*A series of brain-building games designed to challenge and engage children (ages 3–5) by practicing self-control, focus, memory, following directions and problem-solving skills that are essential for school readiness.*"

The game includes 10 unique mini-games that are repeated at increasing levels of difficulty. It also includes one "bonus game," which was the subject of this study. In the bonus game, a cartoon duck appears on screen (Figure 2). The first time the player encounters the game, Cookie Monster (the game's narrator), announces: "*Bonus game! When you see duck, say 'quack' out loud.*" If the player then says the word "quack," the duck flaps its wings and Cookie Monster praises the player, saying, "*Quacktastic!*" The voice-driven interface backing the game is forgiving and interprets almost any sound as a quack.

If the player does not quack audibly, the duck paces back and forth, moving on and off the screen (Figure 2, right). The player is successful if she quacks while the duck is visible on screen and unsuccessful if she quacks at a time when the duck is not visible.

During early levels of CMC, the app will prompt the player to say "quack" after several seconds. When the app delivers this prompt, the duck quacks twice and holds its wing up to its ear (Figure 2, left). At the same time, Cookie Monster says, "*Quick, say quack out loud!*" In later levels, there is no verbal prompt, and the player must remember on her own to quack when the duck appears. In early levels, the player has as long as 35 seconds to quack before the game progresses; in later levels, the opportunity passes after as little as 6 seconds.

Data Collection App

We created a custom logging app for Android devices that tracked when CMC was used. Beginning whenever CMC was launched and again at random times as children played, our logging app recorded ambient audio using the tablet's built-in microphone. Each sample collected five minutes of audio, which was then uploaded securely to the cloud using the Amazon Web Services S3 platform.

In addition to discussing this audio sampling with each family during the consent process, we also displayed a red dot on the screen any time the app was recording audio, and we informed them that they could look for this indicator. We also told families they could ask us to delete any recording they did not want us to access, although no families asked us to do so.

Because the data collection app took ownership of the microphone, the CMC bonus game was unable to hear the player say "quack" when logging was enabled. As a result, in all of our audio samples, the player is attempting to engage with a voice-driven interface that cannot hear them, yielding a data set of children's attempts to repair an irreparable conversation. The rest of the game experience was unaffected by our logging.

Procedures

All families participated in an initial lab session at the University of Washington. As part of this session, the child played CMC for 15 minutes together with a parent. During this initial visit, all families encountered and were able to navigate the voice-driven bonus game. At the end of this lab session, families received a new Galaxy Tab E Android tablet with CMC and our data collection app both pre-installed.

Families were asked to give their child an opportunity to play CMC for at least 10 minutes each day, at least six days a week, for a period of two weeks. They were told that they could use the app as much as they wished in excess of this lower bound. We also asked families to check regularly that our logging app was turned on. At the end of the two-week deployment, families returned to the lab for a follow-up session as part of the larger study; that data is outside the scope of this investigation and not analyzed here.

Data Analysis

Our data set was composed of the audio samples that we collected during the two-week in-home deployment. We collected 587 samples as part of our larger study. We identified 107 samples that were at least 30 seconds long and contained audio of the duck bonus game in the first three minutes of the recording. We then transcribed the first three minutes of each of these 107 clips into short, individual field notes, capturing verbatim quotes, and describing tone of voice, audio from the game, laughter, and other noises.

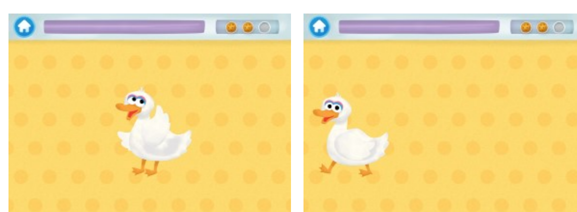

Figure 2. Screenshot of the "Quack" bonus game in Cookie Monster's Challenge. Left: The duck holds its wing up to its ear to imply that the player should speak (and specifically, that the player should say "quack"). Right: When the game begins, the duck starts pacing back and forth, walking on and off the screen. The speed at which the duck walks and the amount of time it spends on screen varies by level, such that the game becomes progressively more difficult as the level increases.

We then divided these samples across the research team and individually analyzed them for themes using an inductive approach. We then collaboratively discussed themes as a group over several sessions to develop a set of open codes. Early themes reflected adults' reactions to the bug, children's reactions to the bug, and children's repair strategies. We iteratively re-coded data, writing analytical memos about codes that had emerged across the research team. From this analysis, we developed a final codebook and re-coded the sample.

Examples of final codes included whether the child repeated their responses many times, whether the child changed their volume, pronunciation, duration, or words used, whether the child asked their parent for help, whether the child gave up (e.g. stopped trying and waited for the mini-game to end), whether parent tried themselves, and whether parent declared the app was broken. Two authors coded a randomly selected 10% of the samples using the final codebook. Cohen's κ was .78.

RESULTS

Children's and Parents' Repair Strategies

We observed that across families, children employed a common set of strategies to attempt to course-correct the broken interaction (see Table 2). Most children employed all of these strategies at various points throughout the two-week study and frequently combined multiple strategies at once. When one strategy failed, children often quickly shifted to another.

Parents encouraged or used some of the same strategies as well. Here, we report these systematic behaviors, as well as the ways in which children's and adults' responses differed.

Repetition

Children's most common repair strategy was to repeat themselves, saying "quack" again and again. All children used this strategy at least some of the time, five children used repetition in every sample, and across all participants, children used repetition 79% of the time.

In many cases, they did so in a way that allowed for a back-and-forth pattern of communication with the app, making space for the conversational turn-taking that occurs in human interactions. That is, after attempting to speak to the app, children would pause (as if waiting for a response) before speaking again, as illustrated in following field note:

> "*The child says 'quack quack,' then pauses for several seconds, then says 'quack quack' again. The child repeats this pattern four times*" (P10).

This stop-and-go pattern of speaking then waiting for feedback was quite common, and children often waited silently for as long as 10 seconds before trying again. Although children almost always attempted to repair the failed communication eventually, they typically would first pause and give the app the opportunity to demonstrate whether it had heard.

However, this was not always the case. In some instances, the child repeated his or her attempts at quacking without making conversational space for the app to reply. For example:

> "*The child quacks three times, and the app says, 'Say quack out loud!' The child then resumes quacking and continues to quack without pausing until the next mini-game begins*" (P13).

In these cases, children began with a stop-and-go approach but transitioned to continuous repetition.

We also observed that parents encouraged and used stop-and-go repetition, saying things like "*Do it again!"* (P10) or "*Keep doing it*" (P7). And parents not only suggested repetition as a strategy for children, they also enacted it themselves. For example:

> "*Dad says, 'quack quack quack!' and child repeats this; then they are both silent for several seconds. Dad spontaneously says, 'quack quack!' again, then pauses again*" (P1).

Across families, parents' most common form of support was to encourage children to repeat themselves or to model this repetition themselves and as they did so, to pause for a response from the app.

Speaking Loudly

Children's second-most common repair strategy was to raise their voice. Children often began the interaction by speaking at a conversational volume but after one or more failed attempts began to say "quack" in a louder voice, at times even shouting. For example: "*The child says 'quack' but the app does not respond. After a short silence, the child says 'quack' again with a louder voice*" (P7). Or similarly, "*The child says quack several times. Each time, the child says 'quack' louder and louder*" (P8). In one sample, as the child shouts out her interaction attempts, an adult in the background can be heard saying, "*Oof! A little quieter, k?*" (P2).

In a few instances, children stated explicitly that they were intentionally raising their voices and that they expected this to be effective in helping the app understand them. For example, one child explained to his mother, "*You have to say it loudly*" (P12), implying that the failed communication was a result of speaking too quietly. These comments suggest that some children may be aware of using this linguistic strategy.

Parents also encouraged changing volume as a strategy, for example, praising the child by saying, "*that was a loud one; good job buddy!*" (P9). And they tried this strategy themselves: "*They both shout 'quack' loudly*" (P2), or "*Mom shouts, 'QUACK!'*" (P5).

Variation

In addition to trying more quacks and louder quacks, children employed a mix of other variations on the target word as they searched for something that worked. For example, in one instance, a child tried quacking several times, to no avail. The app—oblivious to these attempts—then prompted, "*Quick! Say quack out loud!*" The child responded to the prompt, but this time tried a new interpretation of the instruction, shouting, "*QUACK OUT LOUD!*" (P5).

In other instances, children experimented with some of the alternatives below as they sought one that worked:

- "*The child says 'quack' right away and then repeats several times, 'Qu…Qu…Quack'*" (P6).
- "*The child repeats, 'quack,' 'quack,' in different tones and volumes*" (P8).
- "*The child says 'quack' nine times and a long 'quaaaaaaaaack' at the end*" (P11).
- "*The child says 'quack' loudly. After a short pause, the child says 'quack' again. He waits a bit and says, 'quack-ack-ack-ack!*" (P12).
- "*Queek!*" (P7)

In all of these instances and many others, children continued quacking at the tablet, but varied their word choice, intonation, syllabic emphasis, and more. These enactments demonstrated both a wide range of variations on the correct answer and playfulness in children's approach to conversational repair.

Parents occasionally suggested and supported children's attempts to repair the conversation through variation. When P5-child stated, "*quack out loud,*" his father chuckled and said, "*You're smart!*" and his mother said in the distance, "*Did you say 'quack out loud?' Good job!*" In other instances, parents corrected the child's pronunciation, implying that correcting or changing the way they said the word might repair the communication. However, this was not a dominant theme or an approach that parents used themselves.

Technical Investigation

One strategy that parents attempted that was not reflected in data from children was troubleshooting or attempting to repair the communication through a technical fix. Parents said things like, "*We need to figure out how to get the microphone working*" (P6), or "*I can check if the mic is on*" (P11). We did not hear any evidence that children attempted to repair the conversation through technical means, and children did not make this suggestion to parents.

Strategy	Description	Frequency
Repetition	Child repeats the response many times, with or without pauses in between	79%
Volume	Child varies the volume of their voice	63%
Variation	Child varies word choice or pronunciation	34%

Table 2: Common repair strategies used by children.

Interaction Characteristics

In addition to common repair strategies, we also observed several other patterns in participants' interactions with the app. Here we describe some of the cross-cutting characteristics of their conversations.

Responsiveness to Prompts

We saw that children and adults were both responsive to conversational prompts from the app. The app periodically prompted the player to speak, by saying, "*Quick! Say 'quack' out loud!*" Children were highly responsive to these reminders and quickly replied. For example, in one instance, we heard:

> "*Cookie Monster says, 'When you see duck, say quack out loud!' and the child yells 'quack!' The app gives no feedback and after 2 seconds, the child repeats 'quack' again. There is no feedback and the child waits. After 15 seconds, Cookie Monster says, 'Quick! Say quack out loud!' and the child immediately shouts 'quack' again*" (P9).

Although the child had been waiting silently, he immediately spoke in response to prompting from the app.

Even when children appeared to recognize that the app was broken, they remained responsive to conversational prompts. For example, in another instance, we observed the following interactions:

> "*The child says 'quack' twice loudly. Mom asks, 'Did it work?' and child responds, 'no.' The app says, 'Say quack out loud,' and the child quacks again and sighs*" (P6).

Here, even though the child explicitly says that her interaction attempt failed and hints at frustration or resignation with a sigh, she chooses to persist and respond to the app's conversational prompt. Across participants, we saw that children were highly responsive to both the initial prompt from the app and follow-up prompts, mirroring the joint attention and coordinated back-and-forth that characterizes person-to-person conversations. Across all audio clips, the app presented an explicit prompt 91 times. Only twice did a child choose to ignore a prompt, despite the fact that in many of these instances the child had already made many unsuccessful attempts to repair the communication. That is, children continued replying to the app when it spoke to them, regardless of how unsuccessful the conversation had been up to that point.

Children's Independence

In six of our 107 samples, children proactively sought external support to repair the conversation. In these cases, they asked the parent to step in and attempt to communicate with the app in their place, saying things like, "*You say 'quack'*" (P9) or whining, "*I can't say quack!*" (P2). Although parents were actively involved or present in the background in more than half (54%) of our samples, children were far more likely to experiment on their own than they were to seek out help.

Parents did regularly step in and provide support spontaneously, and it is possible that children might have sought help more often if parents had not chosen to intervene. However, in these cases, children often continued to try to engage with the interface on their own and would attempt to quack at the app in between the parent's attempts.

Patience and Frustration in the Face of Failure

In 78% of samples, we detected no audible signs that children were frustrated with the experience, despite the fact that it was broken and entirely unresponsive. In most instances, children cheerfully attempted a variety of different repair strategies until time ran out and the game progressed.

However, frustration was not entirely absent from our sample. In a non-trivial minority of instances, children showed overt or subtle hints of frustration as they engaged with the broken interface, sometimes indicating that they wanted to give up on their repair attempts. For example, at one point we observed:

> "*The child says 'quack' several times. Cookie Monster says, 'Say quack out loud!' and the child responds, saying 'quack' several more times. The app does not respond and the child announces, 'I don't wanna play this game!'*" (P8).

Across all instances, children's frustration is detectable in 23% of audio clips, suggesting that this is routine occurrence but less disruptive than one might expect from a fundamentally broken interaction. There were 18 instances in which the child eventually gave up and stopped trying to repair the conversation; in ten of these the child simply waited patiently for time to run out, and did not express obvious frustration.

Stages of Parent Support for Repairs

When parents were aware of the communication failure, we observed that they often worked to evaluate the source of the problem and to determine whether the child or the app was responsible for repairing the breakdown. As they did so, they engaged in behaviors that clustered into a hierarchy of four stages. We labeled these: "suggesting," "intervening," "resigning," and "pronouncing" (see Table 3).

Suggesting: First, parents sometimes made repair suggestions and prompted the child to correct the communication, saying things like, "*Keep going, try again*" (P1). When parents made suggestions, there was an implicit assumption that the child was responsible for the communication breakdown, or at the very least, had the power to repair it. At times, parents made this assessment of responsibility explicitly and said things, "*Why can't you say it?*" (P3) or "*What do you say?*" (P9). Whether or not the parent stated that the child was responsible, these instances were characterized by the parent offering advice or prompting the child to perform a certain repair behavior without enacting any repair behaviors themselves.

Intervening: Second, we observed that parents would sometimes intervene and try to perform the repair by quacking at the app themselves. This often occurred after the parent had first attempted to offer suggestions to the child. Occasionally, children would request this intervention, and in other cases, parents asked if they could help (saying things like, "*Let me know when the quack shows up, ok?*" (P11)). However, in most cases, parents just spontaneously began quacking, either joining in with their child as he or she quacked, or quacking after their child had tried to do so and failed to elicit a response.

Stage	Description	Example
Suggest	Parent suggests possible repair strategies	*Mom says, "Hold it up; now try."*
Intervene	Parent attempts possible repair strategies	*The child says, "I don't know what's going on." Mom says "quack" out loud and waits for a moment.*
Resign	Parent runs out of ideas for repairing the conversation	*Dad says, "There's that duck again. I think you're supposed to quack at it, but I don't know why it doesn't do anything."*
Pronounce	Parent tells the child that the conversation is beyond repair	*Mom says, "The duck one's just not gonna work."*

Table 3: Stages of parents' reaction to the communication breakdown. Over time, parents' reactions shifted from earlier phases (suggesting and intervening) to later phases (resigning and pronouncing).

Resigning: Third, we observed parents eventually express resignation and indicate that they were out of ideas as to how the communication might be repaired, usually explaining that they "*don't know*" what is wrong. They most often made these statements after first suggesting strategies to the child and/or intervening and trying themselves. Once they shifted from acts of repair to statements of resignation, they said things like, "*(in an exasperated voice) What does that duck want?*" (P5), or after quacking at the screen, describing his own attempt by saying, "*Why doesn't that do anything?*" (P1).

When expressing resignation, parents no longer implied that the child was responsible and instead began to validate children's repair attempts. They said things like, "*You're doing the right thing*" (P6), or "*That's a good quack!*" (P9). Once they moved to expressing resignation, parents often implied that the child had upheld his or her end of the conversation sufficiently, implicitly suggesting that the app may be responsible for allowing the communication to fall apart.

Pronouncing: Finally, we observed parents would occasionally go beyond resignation and pronounce the app broken. These definitive statements declared the conversation beyond repair. They said things like, "*The duck one's just not gonna work for a while*" (P6) or "*Just sit and watch; the duck's gonna go back over here and then it will be the next game*" (P2), implying the child should give up on attempting to restore the communication. In these cases, parents went beyond subtle implication and squarely placed ownership for the breakdown on the app.

Parents moved among all four of these stages, often within a single interaction. For example, in one instance:

> "*Mom says, 'quack' out loud and waits for a moment. The child cheerfully says, 'I don't know.' Mom replies, 'I don't know either.' After a pause, Mom says, 'Try it one more time.' The child says 'quack.' The app says, 'quick, say quack out loud!' and the child says 'quack' loudly one more time. Mom says, 'quack!' immediately after. Both pause, then Mom says, 'I don't think sound works on this thing. Both wait in silence until the mini-game ends'*" (P12).

In this example, the parent tries quacking herself (intervenes), explains that she doesn't know what to do (resigns), prompts the child to try again (suggests), tries one more time herself (intervenes), and finally declares it broken (pronounces). Unsurprisingly, parents were more likely to suggest and intervene early on and became more likely to resign and pronounce later, as they gained more experience with the broken app. However, as shown above, they did not move linearly through these stages and sometimes optimistically returned to suggesting and intervening even after they appeared to give up. We saw all four stages across families, and we saw that parents used these as a way of working to assess the source of the breakdown.

DISCUSSION

Our broken interaction served as a kind of technology probe [11], that is, a simple, potentially incomplete, piece of technology used to explore a tool's uses and appropriations and to probe users' needs. Providing families with a broken voice-driven interaction to navigate allowed us to examine their reactions to encountering this type of problem, their strategies for recovery, the systematic ways in which children's and parents' responses differed, and the ways in which these reactions changed as users' gained more experience with the failure. Though prior work has examined adults' reactions to miscommunications with voice-driven interfaces (e.g., [13]), our probe gave us access to the naturalistic repairs that much younger users employ.

Borrowing Lessons from Voice Interactions with People

The ways in which children engaged in this limited conversation with the Quack mini-game reflected several known ways in which children converse with people. First, prior research has shown not only that young children use repetition and word-variation as conversational repair strategies in person-to-person conversations, but also that they selectively employ these strategies in context-specific ways. For example, Tomasello and colleagues showed that toddlers are more likely to use repetition to correct repairs in conversations with familiar adults but more likely to reformulate their statements in conversations with unfamiliar adults [28]. This differentiated response pattern reflects children's awareness that a familiar adult is likely to understand their speech habits, making any misunderstandings a result of the adults' failure to hear or attend to the child. In contrast, unfamiliar adults are less likely to understand the child's specific idiosyncratic patterns of speech, making it more effective for children to try alternative forms of their statement.

We saw that children regularly used both variation and repetition with the app. This suggests that they explored treating the breakdown as a possible misinterpretation of their specific word choice *and* as a possible failure of attention. Thus, not only did children borrow the repair strategies they use in conversation with adults and apply them to their interaction with the interface, they also appeared to explore different contexts of failure and probe the underlying source of the problem: Does the app know I'm saying quack? Is this the type of quack it expects to hear?

If children flexibly shift among strategies and apply the strategy that best fits the contextual cues they have access to (as they do when speaking to another person), app designers may be able to leverage this tendency. Rather than providing a single type of prompt, Cookie Monster could respond differently when waiting for a response and when confused by input. The app could also detect the type of strategy the child is using (repetition, variation, etc.) and attempt to extrapolate the child's impression of the situation. A child who is shouting loudly may have a different mental model of the breakdown than a child who is changing pronunciation and reformulating the request by adding information.

Second, it is well-established that during person-to-person conversations with children, parents support children in systematic ways in performing repairs. They will, for example, reframe their partial understanding instead of asking an open-ended question. That is, a parent might ask, "*You want* **what** *to eat?*" rather than simply saying, "*What?*" [15] to highlight the aspect of the child's communication that is unclear. As our results are consistent with the idea that children borrow from their knowledge of person-to-person conversations when speaking to a voice-driven interface, it is possible that they would also benefit from interface scaffolding that borrows from person-to-person conversations. Designers might, for example, mirror parents' scaffolding and provide feedback that reframes the child's statement and reveals the system's partial understanding in order to highlight the part it fails to understand (example sketch in Figure 3).

Third, we saw that children applied the norms of conversational turn-taking to their interactions with the Quack mini-

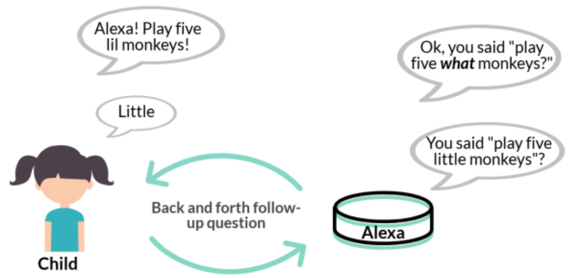

Figure 3. Wireframe of one possible design concept, in which an Echo device replies to speech it doesn't understand ("Play five lil monkeys") by reformulating the statement to highlight what it does ("play five…monkeys") and does not ("lil") understand.

game. It is hard to imagine a user patiently waiting for 10 seconds after clicking a mouse for an interface to respond, and yet children gave the app this leeway before repeating their statement. They also took turns themselves and almost always responded to conversational prompts from the app. Communication researchers consider turn-taking to be the most salient feature of human conversation [30], and parents intuitively begin scaffolding this behavior in infants even before they can speak [4]. Our results suggest that designers of voice-driven interfaces can expect this behavior. That is, designers should expect that children will consistently respond to follow-up prompts (even after experiencing failure) and work to engage in the conversational turn-taking that characterizes human speech.

Differentiating Children from Adults

In many ways, parents' responses mirrored those of children (Figure 4). Adults replied to prompts from the app, repeated themselves, and increased the volume of their voices (although they did not shout, as children did). However, we also saw systematic ways in which adults and children behaved differently. Parents did not vary the way they said the word "quack" or suggest that children do so, they were more likely to give up on the interaction, and they were the only ones to attribute the failure to technical causes. This is consistent with prior work showing that as children grow older, they are less likely to view an anthropomorphic interface as a social and moral other [12]. That is, adults may be more likely to keep in mind the technical infrastructure underpinning the interface, while children may be more likely to consider only the human-like front-end. Future work might explore how the developmental arc of children's beliefs about the social standing of a speech-driven interface move together with the types of utterances they give as input to the system.

Some of the differences we saw between children and their parents may have arisen from the fact that the child was the primary user while the parent, if present, was merely a supporter of the child's interaction. It is possible that adults would be more likely to repeat themselves many times or shout at a voice-driven interface if they were seeking information or entertainment for themselves. Prior work has shown, for example, that adults do make linguistic changes to their voice input—such as speaking more clearly and more slowly—when they receive feedback that an interface has not understood them [13]. Thus, it is possible that adults would have varied their quacks and shouted loudly if they were invested in the communicative act for its own sake.

However, there is also theoretical justification for predicting that, in some respects, young children will engage in systematically different patterns than adults when conversing with voice-driven interfaces, and our results were consistent with this idea. Prior research has shown that children are persistent in their communicative attempts with people; in a lab study of children's follow-up requests, children tried to repair the conversation more than 99% of the time when a researcher showed evidence of misunderstanding [7]. If children's immature speech abilities pose a challenge for natural language processing, one mitigation strategy might be to leverage the persistence children already display with people as they work to master their first language. Designing for collaborative repair between child and interface may be at least as productive as improving automated speech recognition to avoid failure in the first place.

Parents' Stages of Support

Finally, we saw that parents moved through common stages of support as they witnessed their child's repair attempts. Parents appeared to anticipate the linguistic strategies that might work, nudge children toward these behaviors, verify their own intuitions by testing out the interface themselves, and assess the source of communication breakdown by attributing it either to the child or to the device. Although parents often began by assuming the child was responsible for the breakdown, through experimentation and follow-up they shifted to the belief that the device was responsible and explained this to the child.

Our results suggest that in some contexts, targeting parent-child dyads as a collective "user" could be a productive strategy for designing voice-driven interfaces. Adults' corrections and modeling are the primary mechanism by which children acquire their first language [4]. Adults intuitively

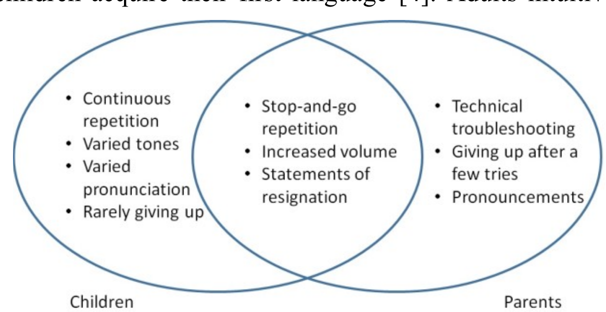

Figure 4: Differences and commonalities between children's and parents' repair strategies and patterns of interacting with the interface

adjust their speech to accommodate children's current level of language acquisition, meeting them where they are, regardless of whether they are just starting to babble, beginning to form two-word combinations, or speaking in full sentences [24].

Given adults' innate skill in scaffolding children's communication, devices might do well to expect that parents or other adults will play a critical role in mediating young children's interactions with voice interfaces. It is easy to envision a parent supporting a toddler attempting to speak to an interface, sharing the interaction and scaffolding the child's vocalizations. It is certainly important for interfaces to be trained on children's speech, but it might also be useful to train them on scaffolded speech, such that the training data includes both children's attempts and accompanying child-directed speech from parents.

Limitations and Future Work

We conducted a very narrowly scoped investigation: a single interaction within a single game, driven by a single word. Though the consistency of the scenario allowed us to hold constant many parameters and examine what was common across children, here we do not begin to explore the vast space of potential communication that might occur between a child and a voice-driven interface or the many types of broken communication that a child might seek to repair. An interface that is backed by artificial intelligence, has a user-specific understanding of the child, engages in conversation, or provides access to a variety of functions and information sources would unquestionably surface many patterns of behavior and speech that are not reflected here. Future work remains to examine the space of conversational breakdown broadly by considering other common misunderstandings, such as an interface misinterpreting one word as another, or an interface recognizing a user's speech correctly but failing to understand the user's intention.

Although we collected general information about participants' technology usage, we did not examine parents' or children's prior experience with voice interfaces. Technical knowledge or experience with this interaction pattern may affect how parents or children interact with voice interfaces and the repair strategies they use. Future work remains to examine the stability of these strategies over time.

We also conducted this study only with English-speakers and with a small qualitative sample that is over-representative of white, middle-class families. Human speech, even within a single language, is culturally specific, and future work remains to examine repairs among diverse children, languages, and vernaculars.

Despite the fact that natural language processing and speech recognition are robust disciplines, there are many unexplored questions with respect to voice interfaces from the perspective of human-computer interaction and child-computer interaction. Future work might explore asking children to design feedback and scaffolding that will help them repair conversations together with an interface (much like existing work in user-defined and child-defined gestures [22,31]). Other work might probe whether children are aware of the linguistic strategies they use to repair conversations with a device and whether they can articulate these strategies to others. Or future work might explore whether these are mindless social scripts [16]; we saw that children consistently responded to prompts from the interface, even after repeated failures. Does the prompt lead children to believe their input will now be effective? Or does it simply elicit a mindless, trained response? Future work might also explore combining multiple modalities with voice. During in-person interactions, children rely on gestures, gaze, and non-word verbalizations to communicate their intentions. Perhaps voice-driven interfaces can improve children's ability to express themselves by detecting mixed input, such as touch and speech combined.

CONCLUSION

We explored the communication repair strategies that preschoolers use when a voice-driven interface failed to understand them. And we examined the ways in which their parents supported them as they worked to correct the conversation. Our results show that children were highly responsive to prompts from the app to engage in conversation, even though this conversation deterministically failed. Children used a small set of common repair strategies to attempt to get the conversation back on track, and they attempted to structure the interaction into the conversational turn-taking that characterizes person-to-person speech. It is well-understood that children—novice speakers, relative to adults—are persistent in repairing communication with other people. We saw evidence of this same diligence as they worked to communicate with the broken interface.

Voice-driven interfaces have become common and are increasingly embedded in children's lives, with every indication that adoption will continue to grow. Our results show that much of what is known about children's linguistic behaviors in person-to-person communication applies to their interactions with these technologies and can serve as inspiration for designers seeking to create positive experiences for children and families.

ACKNOWLEDGMENTS

This work was supported by a dissertation award from Sesame Workshop to Alexis Hiniker. We would like to thank Yiran Ni, Willa Yang, Jesse Van Hoy, and Julia Hon for their help with data preparation, as well as all of the families who participated in our study.

SELECTION AND PARTICIPATION OF CHILDREN

Families in this study were recruited through an institutional participant pool of local people who are interested in participating in research. We were given a list of contact information of parents of children in our target age range. Parents who expressed interest were given more details about the study by phone or email and were sent a copy of the consent form describing all procedures. Parents who chose to schedule a study session went through the consent process when they arrived at our lab. Before beginning each session, we asked the child participant if he or she would like to play some games. We told both the child and the parent that the child did not have to participate and could stop at any time.

REFERENCES

1. Sean Andrist, Iolanda Leite, and Jill Lehman. 2013. Fun and fair. In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 352–355. https://doi.org/10.1145/2485760.2485800
2. Bonnie Brinton, Martin Fujiki, Diane Frome Loeb, and Erika Winkler. 1986. Development of conversational repair strategies in response to requests for clarification. *Journal of Speech and Hearing Research* 29, 1: 75–81. https://doi.org/10.1044/jshr.2901.75
3. Cga. Let's CoPlay! 2016 A Developer's Guide to Family Cooperative Mobile Play.
4. Eve V Clark. 2009. *First language acquisition*. Cambridge University Press.
5. Michael A. Forrester and Sarah M. Cherington. 2009. The development of other-related conversational skills: A case study of conversational repair during the early years. *First Language* 29, 2: 166–191. https://doi.org/10.1177/0142723708094452
6. Marina Fridin. 2014. Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. *Computers & Education* 70: 53–64. https://doi.org/10.1016/J.COMPEDU.2013.07.043
7. Tanya Gallagher. 1977. Revision behaviors in the speech of normal children developing language. *Journal of Speech and Hearing Research* 20, 2: 303–318.
8. Roberta Michnick Golinkoff. 1986. "I beg your pardon?": the preverbal negotiation of failed messages. *Journal of Child Language* 13, 3: 455–476. https://doi.org/10.1017/S0305000900006826
9. Philip J. Hayes and D. Raj Reddy. 1983. Steps toward graceful interaction in spoken and written man-machine communication. *International Journal of Man-Machine Studies* 19, 3: 231–284. https://doi.org/10.1016/S0020-7373(83)80049-2
10. Alexis Hiniker, Bongshin Lee, Kiley Sobel, and Eun Kyoung Choe. 2017. Plan & play: supporting intentional media use in early childhood. In *Proceedings of the 2017 Conference on Interaction Design and Children - IDC '17*, 85–95. https://doi.org/10.1145/3078072.3079752
11. Hilary Hutchinson, Heiko Hansen, Nicolas Roussel, Björn Eiderbäck, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, and Helen Evans. 2003. Technology probes. In *Proceedings of the conference on Human factors in computing systems - CHI '03*, 17. https://doi.org/10.1145/642611.642616
12. Peter H. Kahn, Takayuki Kanda, Hiroshi Ishiguro, Nathan G. Freier, Rachel L. Severson, Brian T. Gill, Jolina H. Ruckert, and Solace Shen. 2012. "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental Psychology* 48, 2: 303–314. https://doi.org/10.1037/a0027033
13. Manja Lohse, Katharina J. Rohlfing, Britta Wrede, and Gerhard Sagerer. 2008. "Try something else!" — When users change their discursive behavior in human-robot interaction. In *2008 IEEE International Conference on Robotics and Automation*, 3481–3486. https://doi.org/10.1109/ROBOT.2008.4543743
14. Silvia Lovato and Anne Marie Piper. 2015. "Siri, is this you?" In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 335–338. https://doi.org/10.1145/2771839.2771910
15. Ernst Moerk. Principles of interaction in language learning. *Merrill-Palmer Quarterly of Behavior and Development 18*, 229–257. https://doi.org/10.2307/23084609
16. Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues* 56, 1: 81–103. https://doi.org/10.1111/0022-4537.00153
17. Dave Paresh. 2018. Which voice in your fridge? Makers pick virtual assistants. *Thomson Reuters*. Retrieved January 16, 2018 from https://www.reuters.com/article/us-tech-ces-assistants/which-voice-in-your-fridge-makers-pick-virtual-assistants-idUSKBN1F12OQ
18. Karen Petty. *Developmental milestones of Young Children, Revised Edition*. Retrieved April 2, 2018 from https://www.redleafpress.org/Developmental-Milestones-of-Young-Children-Revised-Edition-P1411.aspx
19. L. Plowman, O. Stevenson, J. McPake, C. Stephen, and C. Adey. 2011. Parents, pre-schoolers and learning with technology at home: some implications for policy. *Journal of Computer Assisted Learning* 27, 4: 361–371. https://doi.org/10.1111/j.1365-2729.2011.00432.x
20. Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. 2017. "Alexa is my new BFF." In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*, 2853–2859. https://doi.org/10.1145/3027063.3053246

21. Eric E Rasmussen, Autumn Shafer, Malinda J. Colwell, Shawna White, Narissra Punyanunt-Carter, Rebecca L. Densley, and Holly Wright. 2016. Relation between active mediation, exposure to *Daniel Tiger's Neighborhood* , and US preschoolers' social and emotional development. *Journal of Children and Media* 10, 4: 443–461. https://doi.org/10.1080/17482798.2016.1203806

22. Karen Rust, Meethu Malu, Lisa Anthony, and Leah Findlater. 2014. Understanding childdefined gestures and children's mental models for touchscreen tabletop interaction. In *Proceedings of the 2014 conference on Interaction design and children - IDC '14*, 201–204. Retrieved from http://dl.acm.org/citation.cfm?id=2593968.2610452

23. Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50, 4: 696. https://doi.org/10.2307/412243

24. D. N. Stern, S. Spieker, R. K. Barnett, and K. MacKain. 1983. The prosody of maternal speech: infant age and context related changes. *Journal of Child Language* 10, 1: 1–15. https://doi.org/10.1017/S0305000900005092

25. Lucy A Suchman. 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.

26. Lori Takeuchi, Reed Stevens, and others. 2011. The new coviewing: Designing for learning through joint media engagement. In *New York, NY: The Joan Ganz Cooney Center at Sesame Workshop*.

27. Yumiko Tamura, Mitsuhiko Kimoto, Masahiro Shiomi, Takamasa Iio, Katsunori Shimohara, and Norihiro Hagita. 2017. Effects of a Listener Robot with Children in Storytelling. In *Proceedings of the 5th International Conference on Human Agent Interaction - HAI '17*, 35–43. https://doi.org/10.1145/3125739.3125750

28. Michael Tomasello, Michael Jeffrey Farrar, and Jennifer Dines. 1984. Children's Speech Revisions for a Familiar and an Unfamiliar Adult. *Journal of Speech Language and Hearing Research* 27, 3: 359. https://doi.org/10.1044/jshr.2703.359

29. Amy M Wetherby and Barry M Prizant. 1993. *Communication and symbolic behavior scales: Manual*. Riverside Publishing.

30. John M. Wiemann and Mark L. Knapp. 1975. Turn-taking in Conversations. *Journal of Communication* 25, 2: 75–92. https://doi.org/10.1111/j.1460-2466.1975.tb00582.x

31. Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, 1083. https://doi.org/10.1145/1518701.1518866

32. Selma Yilmazyildiz, Robin Read, Tony Belpeame, and Werner Verhelst. 2016. Review of Semantic-Free Utterances in Social Human–Robot Interaction. *International Journal of Human-Computer Interaction* 32, 1: 63–85. https://doi.org/10.1080/10447318.2015.1093856

33. Amazon.com: The Magic Door: Alexa Skills. Retrieved January 16, 2018 from https://www.amazon.com/The-Magic-Door-LLC/dp/B01BMUU6JQ

34. Cookie Monster's Challenge Mobile Downloads | PBS KIDS. Retrieved January 14, 2018 from http://pbskids.org/apps/cookie-monsters-challenge.html