The Interactive Show: A Conversational Companion for Young Children and Childcare Assistant for Parents

Neelma Bhatti
Department of Computer Science
Virginia Tech, Blacksburg, USA
neelma@vt.edu

Timothy L. Stelter
Department of Computer Science
Virginia Tech, Blacksburg, USA
tstelter@vt.edu

D. Scott McCrickard
Department of Computer Science
Virginia Tech, Blacksburg, USA
mccricks@vt.edu

ABSTRACT

Advances in the fields of natural language processing, machine learning and speech recognition have led to the increasing adoption of conversational agents in a variety of domains. However, the use of conversational agents as a childcare assistant for parents of young children has not been realized. We propose the use of conversational agents into children's programs to produce an interactive show, which can be employed as conversational companion for children, and a childcare assistant for parents of young children who have limited access to childcare options. Key features of the interactive show include understanding, responsiveness and maintaining an uninterrupted conversation flow. We argue that having a truly interactive show can engage children for a longer period of time without being distracted, which can give parents some time to attend to their needs in the absence of help.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; *Child-Computer Interaction*; Natural language interfaces.

KEYWORDS

conversational agents, child-computer interaction, childcare, interactive show

ACM Reference Format:
Neelma Bhatti, Timothy L. Stelter, and D. Scott McCrickard. 2020. The Interactive Show: A Conversational Companion for Young Children and Childcare Assistant for Parents. In *Proceedings of the 8th International Conference on Human-Agent Interaction (HAI '20), November 10–13, 2020, Virtual Event, NSW, Australia.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3406499.3418753

1 INTRODUCTION

Young children use screen media for two hours on average everyday [7, 18], giving the parents some time to attend to their house chores or work related tasks. Many animated programs and cartoons such as Mickey Mouse Clubhouse (Disney+), Word Party (Netflix) and Dora the Explorer (Nickelodeon) are designed to interact with children, with the characters asking questions from them as they go along. These shows are designed to provide entertainment as well as education to children by incorporating various concepts and ideas that the cartoon characters convey during the course of the show, and engage children by having them answer questions. Some of the characters say the correct answers when a brief period elapses after asking the question from the children, giving children a reassuring feeling about their replies. Few others (such as Dora the Explorer) do not sync well with children's responses, making the children feel interrupted or mistaken [4]. This feeling sometimes leads to frustration on the children's part for not being acknowledged for their knowledge. Children also feel that the character is unable to listen to them, making them yell louder. Other times, they feel like the character's replies make no sense. The character's inability to provide flexible time gaps for garnering child's response and the lack of scaffolding mechanisms to drive the conversation along the context [18] creates a desire among children to have an interactive show in which they can really *talk* to the character, and the character can understand what they are saying instead of just listening.

HAI '20, November 10–13, 2020, Virtual Event, NSW, Australia

ACM ISBN 978-1-4503-8054-6/20/11.
https://doi.org/10.1145/3406499.3418753

2 BACKGROUND

Almost 30 million American homes have been reported to have a smart speaker with a Conversational Agent (CA) such as such as the Echo, Google Assistant, Siri and Amazon Alexa [4]. Children have been found to increasingly interact with conversational agents, making them a good learning partner for young children [17–21]. Advances in natural language processing have enabled conversational agents to hold dialogues with children such as asking questions according to the developmental level of the child [17]. An animated PBS KIDS science show named "Elinor Wonders Why" planned to debut in September 2020 leverages intelligent CA to enable verbal interactions between children and an on-screen character [18]. Another study explored the child-agent verbal interaction by developing a story-telling CA for engaging children in book reading activities [19]. However, the use of conversational agents integrated into a show which can act as a childcare assistant for parents has not been realized.

3 METHOD

The researchers, two of which have young children, encapsulated the desired features of an interactive show as being *understanding*, *responsive*, and maintaining an uninterrupted conversational *flow* in an ideation session. These characteristics would distinguish it from the conventional conversational shows for children by creating a symmetry in the conversation. These features combined with the problem domain of a child feeling misunderstood with current child-focused programs allowed us to have an extended brainstorming session to make real our design thinking [8]. Since

we envisioned to make use of the interactive show as a childcare assistant, engagement and reduced distractions were our primary point of concern. We believe that minimizing distractions would ensure greater engagement, giving parents some hands-free time to attend to their work.

3.1 Conversational Goals

Some of the goals of using CA to provide interactivity include:

(1) *Understanding* or giving the impression of being understood.
(2) Recognizing the verbal and emotional cues to provide relevant *response*.
(3) Maintaining an uninterrupted conversational *flow* by picking up the voice cues.

4 THE INTERACTIVE SHOW

We present our proposition for an interactive show for children between 2-5 years of age, as children in this age range tend to naturally interact with agents [6]. Our approach is similar to one used for Netflix interactive shows [2, 14, 16], except that instead of a remote control, here the child would use their voice to control the output as "kids in general don't hesitate to talk to the screen" [16].

4.1 Understanding

Since the developmental milestones for children in this age range vary largely [15], the CA needs to have a preset of age appropriate answers for the targeted audience to *feel understood*. In case the child digresses and talk about random things according to their mood, or for testing the system, the CA needs to respond in a way which makes the children feel that it is keeping up with them [4]. In this case, the CA can listen to whatever the child is saying, and answers with a relevant default answer like "Can you say that again, I seemed to have missed it" or "That's nice, can you tell me more?". The first default response gives the child a sense of the CA owning their inability of getting what they said (instead of vice versa), whereas the second default response gives them a reassuring feeling of being heard, both fallback mechanisms [19] embodying the goal 1 for an interactive show. For younger children who cannot speak coherently, a reassuring reply to their "gibberish" can also do the trick. That way, the child can still feel like the system is listening to them, without feeling misunderstood [4]. Overtime, the CA could learn from the gibberish to provide personalized learning.

4.2 Responsive

For a show to be truly interactive, we believe that it should go beyond the conventional shows asking context-specific questions, and expecting predetermined answers. For a true interactive element, the questions and replies also need to be two-dimensional: incorporating the elements of interactive story-telling [19] in existing shows with a CA. A real talking and interactive show would listen to a child speak to their hearts content, and give relevant comprehensible replies. With advancements in recognition of emotions and affects through voice [11], the proposed show can also recognize a child's verbal and emotional cues and react appropriately to meet goal 2 [1, 3, 13]. In event of complex queries such as "Which Mousekatool should we use now?", the CA could limit the responses to two to four choices, running the response by the child through all the choices to see if it matches any to respond by combining fallback intent with scaffolding mechanism [18].

4.3 Flow

Maintaining a uninterrupted conversational flow by picking up the voice cues (i.e. start and end of a conversation), and shaping the conversation flow to the children's behavior and attention level is the key to avoid scattered conversational session with children, which can result in reduced engagement of the child. Scaffolding by providing accessible questions have also been known to avoid breakdowns in communication [19]. Some of the noted prompts include tone and pitch of the child's voice, emotional state or level of attention (or the lack thereof). Character portraying the CA visually can make appropriate pauses to hear the child's response, and appreciate or politely correct them depending them on their answers to successfully meet goal 3. Specific feedback provided in such way can engage children seamlessly, as has been suggested in similar studies [19].

5 OPPORTUNITIES AND CHALLENGES

The interactive show can simultaneously act as a conversational companion for children and childcare assistant for parents of young children, who have little or no help while they perform house chores or work from home, especially in situations like the recent outbreak of the COVID-19 disease. The fact that the show is conversational can help in addressing concerns about screens inducing slow language development and speech delay in children [9]. Integration of CA in a children's show does not need a technological leap to be implemented, since solutions like Alexa and Google Home already are popular among children [5, 10, 12].

The solution has scalability considerations in terms of content to be presented to the children, as it has to stick to a limited number of options to avoid confusion. The content also needs to be adapted exclusively for children out of the target age range. Furthermore, privacy and trustworthiness are primary concerns when dealing with the broader implication of technology. Specifically, parents should feel comfortable letting their children interacting with the show without fear of data being collected of about their child(ren).

6 CONCLUSION

We present a proposition for an interactive show which utilizes CAs to act as a childcare assistant for parents by engaging their children. Future research in this area can look into the type of activities and content that can be presented to children in this age group according to their attention span, physical and mental (dis)abilities and parent preferences with different levels of active supervision. Strategies can be developed for responding to multiple children who are interacting with the show by prioritizing each child's responses. Specific concerns for CAs as childcare assistants for differently-abled children is also a notable future research prospect.

ACKNOWLEDGMENTS

We would like to thank Kejun Pan (University Montessori School) whose brief submitted in Research Design competition (ACM IDC 2020) inspired our design.

REFERENCES

[1] Haojun Ai, Liangliang Han, Yifeng Wang, and Liang Liao. 2018. Accurate acoustic based gesture classification with zero start-up cost. In *International Conference on Algorithms and Architectures for Parallel Processing*. Springer, 44–58.

[2] Edgar Alvarez. 2017. Kids control the story in Netflix's new interactive shows. https://www.engadget.com/2017/06/20/netflix-interactive-shows/

[3] Rafael A Calvo and Sidney D'Mello. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing* 1, 1 (2010), 18–37.

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

[5] Statista Research Department. 2020. Number of smart speakers in households in the United States from 2017 to 2019. https://www.statista.com/statistics/967402/united-states-smart-speakers-in-households/

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

[7] Shalom M Fisch. 2014. *Children's learning from educational television: Sesame Street and beyond*. Routledge.

[8] Rex Hartson and Pardha S Pyla. 2012. *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier.

[9] Donna Hermawati, Farid Agung Rahmadi, Tanjung Ayu Sumekar, and Tri Indah Winarni. 2018. Early electronic screen exposure and autistic-like symptoms. *Intractable & rare diseases research* 7, 1 (2018), 69–71.

[10] Rodrigo Ibanez, Alvaro Soria, Alfredo Teyseyre, and Marcelo Campo. 2014. Easy gesture recognition for Kinect. *Advances in Engineering Software* 76 (2014), 171–180.

[11] Yelin Kim, Tolga Soyata, and Reza Feyzi Behnagh. 2018. Towards emotionally aware AI smart classroom: Current issues and directions for engineering and education. *IEEE Access* 6 (2018), 5308–5331.

[12] Shanhong Liu. 2020. Smart speakers - Statistics Facts. https://www.statista.com/topics/4748/smart-speakers/

[13] Catherine Marechal, Dariusz Mikołajewski, Krzysztof Tyburek, Piotr Prokopowicz, Lamine Bougueroua, Corinne Ancourt, and Katarzyna Węgrzyn-Wolska. 2019. Survey on AI-Based Multimodal Methods for Emotion Detection. In *High-Performance Modelling and Simulation for Big Data Applications*. Springer, 307–324.

[14] Casey Newton. 2017. Netflix's interactive shows arrive to put you in charge of the story. https://www.theverge.com/2017/6/20/15834858/netflix-interactive-shows-puss-in-boots-buddy-thunderstruck

[15] US Department of Health, Human Services, et al. 2014. Birth to Five: Watch Me Thrive! A compendium of screening measures for young children.

[16] Peter Rubin. 2018. How the Surprise New Interactive Black Mirror Came Together. https://www.wired.com/story/black-mirror-bandersnatch-interactive-episode/

[17] Ying Xu and Mark Warschauer. 2019. Young children's reading and learning with conversational agents. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.

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

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

[20] Ying Xu and Mark Warschauer. 2020. What Are You Talking To?: Understanding Children's Perceptions of Conversational Agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

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