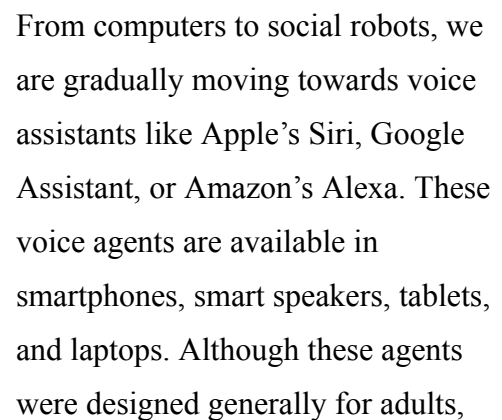


1



Dr. Sumitava Mukherjee, Faculty member, IIT Delhi

¹ Image by [ThinkGizmo on Flickr](#). Under Creative Commons Licence CC-BY-SA 3.0

Introduction

From computers to social robots, voice assistants like Apple's Siri, Google Assistant, or Amazon's Alexa require Natural Language Processing (NLP), voice recognition, and speech synthesis to recognize the voice commands and accordingly generate a response (Hoy, 2018; Terzopoulos and Satratzemi, 2020). These voice agents are available in smartphones, smart speakers, tablets, and laptops. Due to the very nature of voice-based interaction, one does not have to face the barriers of typing or spelling. Children do not have to be specifically fluent in reading or writing to use voice assistants (Yuan et al., 2019). But, although these agents were designed generally for adults, we are seeing children as additional emerging users. Are these agents capable of engaging in an interaction that is developmentally or age-appropriate for children who are driven by curiosity and (verbal) question-asking (Lovato & Piper, 2019)?

Social robots that can communicate through voice are quite different from voice assistants. This is because nonverbal elements of communication like facial expressions, body movement, posture, and eye contact are absent in voice assistants (embedded in smartphones or smart speakers). Even the variability of intonation in synthesized voices is less compared to the voice of humans (Aeschimann et al., 2020). In this case, there is a need to study these voice assistants separately from social robots and chatbots. This is reflected in a study conducted by Druga et. al (2017) where children interacted with Cozmo (an autonomous robot), Google Home, Amazon Echo Dot's Alexa, and Julie (a conversational chatbot). Children reported that they largely preferred and liked to interact with Cozmo because it could move, had feelings, and made eye contact, unlike the other conversational agents which were simply "just flat cylinders" (i.e. smart speakers). Therefore, the form of the voice agents does influence the interaction. Prior work in social robotics and human-machine interactions showed people ascribe different mental states to embodied Artificial Intelligence agents like robots - and the structure of such agents matters a lot. When the structure is more human-like (Anthropomorphic), they are seen as having human-like abilities and thinking. None of this can happen in voice assistants that are embedded in 'flat' cylindrical speakers or one's own mobile phone. The only human-likeness in digital

voice assistants is how human-like the speech or verbal interaction indeed is. Such interactions would vary depending on whether one is conversing with adults, children, or kids.

Thus, while there are extensive studies on how children interact with and perceive computers, social robots, and chatbots in the field of Human-Computer Interaction (HCI), we need a different and deeper understanding of what is going on cognitively among its users. Since voice assistants were developed and became available to most of the world's population a decade ago, there has been a few exploratory studies, only in developed countries on how children interact with and perceive these agents. Therefore, in this review, children's patterns of interaction with voice assistants will be discussed based on previous studies to understand what has been explored till now and what could be explored in the future.

Children's Usage of Voice Assistants

The first question is for what purposes do children use them? Lovato and Piper (2015) conducted an online survey with parents to understand how children use voice assistants. They broadly discussed the three purposes for which children used voice assistants in their study. First, children explored by asking questions like "Are you married?" or "Want to be my best friend?" to understand the voice assistant (**exploration**). Second, children used voice assistants to seek information about their areas of interest, weather, or restaurants (**information-seeking**). Third, children used it for functional purposes like commanding the voice assistant to call or send messages to someone (**functional needs**). Additionally, some children also used it for amusement or **enjoyment**. For example, in their study, a child aged 7 years found it funny to hear Siri's answers to their silly questions and to repeat the same questions for which they knew that Siri would give silly answers.

Interestingly, Druga et al. (2018) found that children in the age group of 4 - 8 years use play and experience to build the **meaning of an agent's intelligence**. But the way they perceived and developed their thoughts about the agent's intelligence was heavily influenced by their parents when children reached the age of 8 years. From this, it can be inferred that parents play a

critical role in early interventions for the child's understanding of the agent. These observations point out the importance of 'serious play' as a learning mechanism that could potentially also boost learning among children (Resnick, 2017). Further; play, entertainment, learning, information-seeking are heavily entwined in such interactions.

So, the existing literature, in both directions, is not sufficient to suggest that children view voice assistants as agents with moral and social characteristics similar to humans. Interactions (see the section below) point out ways in which children expect the voice assistant to answer as a human (or rather, elder) would, but it does not do so at the moment. Despite this, children do apply similar conversational repair strategies of human communication to improve their interaction with voice assistants as well.

Challenges and repair strategies in Interaction

Children face several challenges when interacting with voice assistants. These include wrong transcriptions, incorrect or complex responses, and reformulating queries. In a study conducted by Lovato et al. (2019), children aged 5-6 years were encouraged to ask questions to Google Mini Home installed in their homes for two weeks. Out of the total number of questions asked (1,577), 89% (820) were transcribed correctly by the agent. But out of this 89%, only 50% (463) of the questions were answered correctly. Some questions were refused by the agent due to incorrect pronunciation and wording, lack of context, broad questions (like "Where do babies come from?"), and advice-seeking questions (like "How to make friends?"). Sometimes the correct responses given by the voice assistants did not contain the answers in a direct way and, therefore, required further interpretation by children. But the interpretation of complex responses was difficult for children who rather needed simplistic phrasings to understand the correct responses. Despite these challenges, children employ repair strategies to improve the interaction.

In order to understand repair strategies employed by children during interaction with a voice agent, Cheng et al. (2018) provided 14 pre-schoolers with an app having a voice-driven

interface that contained a bug to make it occasionally unresponsive. It was observed that children often applied the repair strategies employed in human-to-human communication to an interaction with a voice assistant as well. Repetition, speaking at a high volume (sometimes even shouting) and variations in word choice or pitch were the frequently used repair strategies by children. External help from parents for attempts at repairing the conversation was less preferred by most children. Instead, they preferred to explore on their own. Cheng et al. (2018) proposed a four-stage model of parental support at conversational repair strategies. First, parents usually suggest to keep going on with the conversation. At this stage, children are indirectly seen as responsible for the breakdown in communication. Second, parents intervened or themselves tried by directly engaging in the interaction to elicit a response from the voice assistant. Third, parents resigned from repairing the conversation if the first and the second stage hadn't worked or if they lacked any new ideas. At this stage, the app, not the children, was held responsible for the breakdown in communication. Lastly, parents pronounced or declared that the app couldn't be restored for communication. Though sometimes, parents still returned to the first two stages after giving up.

In case of an incorrect response elicited by voice assistants, children primarily engaged in phonetic reformulation, like changing the speed or pitch of the voice. However, phonetic reformulations and sometimes, threats made to Siri by children did not result in better recognition of their voice (Lovato & Piper, 2015). In this case, could prolonged exposure to voice assistants teach patience to children while they might learn to ask specific questions? Children also couldn't reformulate their query correctly if the agent transcribed incorrectly or gave an incorrect answer (Aeschimann et al., 2020). Choosing keywords and dissecting a complex query into a simple query was difficult for children. Future studies could focus on how semantic reformulations (like rephrasing questions) could be employed by children as an effective conversational repair strategy.

Concerns and Design recommendations

Voice assistants have been found to be useful for seeking information or developing a companionship but there are few concerns that need to be addressed when it comes to children. Voice assistants elicit answers immediately after the child's question which could impact the child's development of reasoning skills. One problem is to move from machine-type interactions to human-type interactions. This can be done in two ways: One - more natural language processing research to understand non-structured statements, entertain questions, skip pauses (like “mmmm”), etc. and Two - emulate more human-like feedback or conversations with them (meaning, the voice assistants also behaves more human-like). Further, when the assistants become more human-like, there could be roles to play as an adult (with more adult voices) or as a peer (with more kid voices).

Meanwhile, parents and teachers often engage in conversations so that children could also partially contribute to the answer (Lovato & Piper, 2019). Lovato and Piper (2015) also analyzed 40 Youtube videos that covered the interaction between children and Siri. In order to get humorous responses, children in eight videos used bad language with Siri. Though, Siri recognized the abusive language and rebuked with phrases like “That is not nice” or “I’ll pretend I didn’t hear that”, there are concerns regarding negative developmental outcomes on children. So, future research could focus on how children could develop a more cooperative relationship with voice assistants.

There are additional digital privacy issues that need to be considered as well since the recordings are sent to the companies’ servers. Terzopoulos and Satratzemi (2020) showed that there was underutilization of privacy settings or controls since the voice recordings were manually deleted by only 3% of the 116 owners of the smart speakers who took part in their study. Even the mute button was used by only 4% of the owners. So, users of voice assistants need to be made aware of the control settings that could be utilized to protect their privacy.

Recent news that might be relevant:

<https://www.wired.com/story/google-assistant-parental-controls-kid-friendly-features/>

Previous studies have also suggested few design recommendations that could be implemented in voice assistants. For example, agents could be programmed to recognize and differentiate the voice of children and adults and respond accordingly (Lovato et al., 2019). Recently, technologists are also undertaking efforts to build voice assistants that could be children friendly. Systems could be designed to engage in interactions that reflect contextual understanding based on children's prior responses or questions (Yuan et al., 2019). Or, additional information could be requested by the voice assistants to respond to a query (Lovato & Piper, 2019). Children's difficulty in understanding complex responses could be overcome by generating simplified responses and more in a human-like role. In 2022, Google Assistant announced "children friendly" features such as: simplified, age-appropriate dictionary to answer in a child-appropriate manner; added multiple voices including gender-neutral ones, parental control to make the service unavailable (downtime) or stop certain kinds of downloads. According to a prominent magazine WIRED, it cited the Sissie Hsiao, the vice president of Google Assistant as saying "They're growing up in an era where voice assistance and voice as an access point for technology is only going to get better. We really want to make sure that those first engagements are positive, delightful, and helpful to parents and kids." Other companies like Amazon (Echo) launched a dedicated kids version with similar aims. We still have to do behavioral studies to see how exactly the interactions pan out and what could be the psychological underpinnings of such interactions.

Conclusion

Voice assistants have potential uses for children in educational settings, for the elderly as a companion, and for visually impaired individuals as a better form of screen readers. We need to now see what can be done for voice assistants to infuse curiosity and learning among children. This opens a space for behavioral sciences to experimental study the use of voice assistants with the aim of promoting knowledge, curiosity, and learning among children, including potential integration in homes and schools; with designers and companies in the loop.

References

- Aeschlimann, S., Bleiker, M., Wechner, M., & Gampe, A. (2020). Communicative and social consequences of interactions with voice assistants. *Computers in Human Behavior*, 112, 106466.
- Biele, C., Jaskulska, A., Kopec, W., Kowalski, J., Skorupska, K., & Zdrodowska, A. (2019, February). How might voice assistants raise our children?. In *International Conference on Intelligent Human Systems Integration* (pp. 162-167). Springer, Cham.
- Cheng, Y., Yen, K., Chen, Y., Chen, S., & Hiniker, A. (2018, June). 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* (pp. 337-348).
- Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017, June). "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* (pp. 595-600).
- Druga, S., Williams, R., Park, H. W., & Breazeal, C. (2018, June). How smart are the smart toys? Children and parents' agent interaction and intelligence attribution. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (pp. 231-240).
- Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Medical reference services quarterly*, 37(1), 81-88.
- Kahn, P. H., Friedman, B., Perez-Granados, D. R., & Freier, N. G. (2006). Robotic pets in the lives of preschool children. *Interaction Studies*, 7(3), 405-436.
- Lovato, S., & Piper, A. M. (2015, June). "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* (pp. 335-338).

- Lovato, S. B., Piper, A. M., & Wartella, E. A. (2019, June). Hey Google, do unicorns exist? Conversational agents as a path to answers to children's questions. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (pp. 301-313).
- Lovato, S. B., & Piper, A. M. (2019). Young children and voice search: what we know from human-computer interaction research. *Frontiers in psychology*, 10, 8.
- Resnick, M. (2017). *Lifelong Kindergarten: Cultivating Creativity through Projects, Passions, Peers, and Play*. MIT Press.
- Terzopoulos, G., & Satratzemi, M. (2020). Voice Assistants and Smart Speakers in Everyday Life and in Education. *Informatics in Education*, 19(3), 473-490.
- Yuan, Y., Thompson, S., Watson, K., Chase, A., Senthilkumar, A., Brush, A. B., & Yarosh, S. (2019). Speech interface reformulations and voice assistant personification preferences of children and parents. *International Journal of Child-Computer Interaction*, 21, 77-88.
-

Disclaimer

The trademarks are registered by respective organizations. This report makes no judgment implicitly or explicitly about any company, product, or brand. No liabilities are imposed upon the authors of any form. The content is for information purposes only and do not hold any bearing whatsoever regarding its usage.

Authors

Jeeva Muthukkumar worked as a summer intern to create an early draft which has formed this report when she was a M.Sc. student in Cognitive Science at IIT Gandhinagar. Her work during the internship is duly recognized and credited.

Dr. Sumitava Mukherjee is a lead researcher in Digital Behavior and a faculty member at Department of Humanities and Social Sciences at IIT Delhi. He is a cognitive and behavioral scientist specializing as a decision researcher who also leads the Decision Lab.

About Humans & Technology

We aim to work as a non-partisan academic research group and an information tank in India that engages and informs the public about key aspects of human behavior or cognition in the context of human-technology interactions. Our insights are relevant to people, businesses, society, and the nation. We work with individuals and organizations to create Reports for easy, free, public dissemination of knowledge. Contact: Dr. Sumitava Mukherjee [sm1@hss.iitd.ac.in]