Conversational User Interfaces As Assistive interlocutors For Young Children's Bilingual Language Acquisition

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

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

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

ABSTRACT

Children in a large number of international and cross-cultural families in and outside of the US learn and speak more than one language. However, parents often struggle to acquaint their young children with their local language if the child spends majority of time at home and with their spoken language if they go to daycare or school. By reviewing relevant literature about the role of screen media content in young children's language learning, and interviewing a subset of parents raising multilingual children, we explore the potential of designing conversational user interfaces which can double as an assistive language aid. We present a preliminary list of objectives to guide the the design of conversational user interfaces dialogue for young children's bilingual language acquisition.

CCS CONCEPTS

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

[1] Wikipedia defines Conversational User Interfaces (CUIs) as "a user interface for computers that emulates a conversation with a real human [...] (which) provides opportunity for the user to communicate with the computer in their natural language rather than in a syntax specific commands" [6]. Different terms appear in the literature for various types of CUIs including voice assistants [2, 17], intelligent personal assistants [5, 16], smart speakers [2], voice user interfaces [12], voice based conversational agents [11] and conversational agents [3, 5]. In this paper, we use the terms CUI and CA interchangeably while referring to voice based conversational agents which are commonly used in domestic settings such as Google Home, Amazon Alexa, Echo Dot, Siri etc.

KEYWORDS

conversational agents, child-computer interaction, conversational user interfaces, language aid

INTRODUCTION

A large number of international parents both in and outside of the US raise children who speak more than one language. These parent(s) usually have minimal spoken communication with or in the presence of their children, and they usually do not have co-located extended family members, who can converse with the children using their native language. Additionally, such children feel challenged to communicate with their age mates in a language different from the one spoken at their home. Research shows that parents often use screen media content to acquaint their children with their parent's native language and to also help them become proficient in the language of communication in the country that they reside in[2]. However, screen media content with animated characters (popular examples:*Dora the Explorer, Mickey Mouse Clubhouse*) as well as human characters (popular examples: *Blippi, Mother Goose Club, Blue's Clues*) provide a form of interactivity and incidental language learning for young children primarily in English language only.

Children's communication with conversational user interfaces (CUI)[1] is a relatively nascent domain of research [2, 8, 11]. Conventional conversational partners for children (both in the form of voice interfaces and screen media characters) are unresponsive, do not give the children the feeling of being understood, and are incapable of maintaining an uninterrupted and seamless conversational flow [3]. Children also struggle communicating with conversational agents (CA) due to grammar and language complexity [2]. Although parents seek opportunities for using CUIs to expand their children's communication skills and learning another language [2], the functionality of switching languages is not available in state-of-the art CUIs. We explore this understudied domain by investigating the capability of CUIs as an interactive language aid for young children (2-5 year old) that can help children learn a secondary language in the absence of an interlocutor. We leverage the qualities of screen media in aiding children with language learning, and try to translate those qualities into the design of CUI for children.

APPROACH

Having a conversation partner in a true sense has a different meaning to children than adults, as attributes of conversation which are valued by adults such as trustworthiness and establishing common ground are not relevant to young children [5]. Instead, children value familiarity [9, 15] and interactivity [3] offered by screen media characters which are not a part of present day CUIs. Therefore, in the pursuit of proposing design goals for CUI which can act as an assistive language aid for children, we adopted a three-pronged approach while focusing on the aspects of dialogue that need

to be *different* from the conversational flow with adults. We first reviewed relevant literature on children's interaction with voice agents. Next, we conducted interviews with twelve mothers in the US who belonged to a subset of immigrant and non-US parents raising young bilingual children. Finally, a group of three researchers met to discuss and encapsulate the design goals of a CUI for young children in an ideation session [3] based on the literature, requirements of the user population, and the experience of first author as a mother as well as a keen observer of her bilingual preschooler's screen media watching sessions.

DESIGN GOALS

In the subsequent sections, we present a list of objectives (not arranged by importance) for the design of CUIs as assistive interlocutors for young children's bilingual language acquisition.

Language Chunks

Children of four of the mothers we interviewed experienced speech delay due to the cognitive load of understanding and communicating in more than one language. One of the four mothers repeatedly expressed how screen media assisted her child's speech more than sessions with his speech therapist or play-dates with his age mates, while another mother shared her child's speech therapist's strategy which seemed to be effective for her son:

> "I was having a lot of difficulty because we were using very long sentences to communicate with our child, like we used with each other, so as a result he (child) did not answer us. She (speech therapist) showed me how to use smaller words and sentences with the child so that he understands them and responds"

Short and succinct utterances by intelligent personal assistants have been found to be helpful for non-native language speakers [16]. The importance of short sentences and common, easily graspable words has also been reported in literature, as lengthy and complex utterances by CA can make children vary of the responses [11]. Hence, design of CA as assistive interlocutor could benefit from responding to a child differently than it does to an adult by *communicating in simple, decomposed language*.

Familiar Language Scaffolding

Children of different ages and developmental delays need varying levels of language scaffolds [7]. While socially contingent video viewing sessions can effectively promote preschoolers science learning [18], research suggests the insufficiency of communicative social cues to support toddler's word learning via video presentation in the absence of a co-present communication scaffold [15] such as verbal scaffolding by mothers of 3 to 4 year old children, which can result in richer verbal interactions

and better language skills [14]. Thus, we believe that training the CA with attributes of language such as familiar voice, talking style, and commonly used words from the child's native language can help it mimic the child's family members including parents or older siblings. We take into account that parents of such children who have not lived in their native place for an extended period of time might not have good grip on the language themselves. Hence we also consider training the CUI in a remote location with familiar language scaffolding before being deployed in the home. Despite the critiques of mimicry objective [1], we argue for familiar language scaffolding due to evidence in literature [9, 15] about the positive effects of children's para-social relationship with media characters on their learning due to the familiarity offered by them.

Contextual Appropriateness

Young children have a limited attention span and are more prone to distractions than adults; for instance, disruptions in the flow of conversation are likely to shift their attention elsewhere. The state of the art screen media shows do a good job of grabbing children's attention through interactive chatter and questioning. As one of the mothers mentioned:

> "Blippi is kind of like Dora, except that instead of waiting after asking question, he just goes and plays by himself. So I think they (children) get attracted to him because it's not that he's always asking questions, he just asks a question or says hi and waits for the response, and then he does something or starts playing. I think asking questions is kind of their way of grasping the attention of the viewer".

As screen media offers interactivity and learning for children with human and animated characters attracting children's attention by asking them questions and performing activities that can keep them engaged, we envision this asynchronous interactivity in CUI's to be beneficial for children's language learning. Thus for engaging young children in meaningful dialogue, it is important for a conversational agent to be designed with a goal of being *contextually appropriate*, where context can be a child's age, mood, or developmental milestones.

Social Communication Cues

Several interviews participants mentioned their children learning English as a secondary language by watching static content (rhymes, cartoons) and shows with characters which urge children to respond to them (*Blippi, Dora, Mickey Mouse Clubhouse*). Children as young as 2 year old can learn new words by being a passive part of a social interaction, where they can hear two individuals conversing with each other [13]. Additionally, on screen social cues have been reported to be associated with greater child engagement [15]. In consideration of attributes of screen media and normal social interactions which can successfully engage and acquaint children with new words, we envision a scenario where

DESIGN GOALS AT A GLANCE

(1) **Language chunks:** communicating in simple, decomposed language

(2) **Familiar language scaffolding:** training the conversational agent with familiar voice, talking style, and commonly used words from their native language

(3) **Contextual appropriateness:** contextually appropriate, where context can be a child's age, mood, or developmental milestones

(4) **Social communication cues:** secondary language speaking interlocutor in a multi-person dialogue

(5) **Code switching and accent variations:** attuning the conversation to children's code switching and accent variations

CA can stage a show with different voice characters having conservation among themselves while also involving children to enable language learning without assistance from family members. This phenomenon was also acknowledged by the first author, who co-watched *Dora the Explorer* with her preschooler, where the titular character would occasionally rub off Spanish words in a primarily English show by giving the impression that certain characters in the show don't understand English. The portrayal of one of characters as a *secondary language speaking interlocutor in a multi-person dialogue* urges the children to understand and communicate in a different language. Apart from helping the child learn the language, it can also enable them to understand multi-person social cues.

Code Switching And Accent Variations

Young children tend to use unclear or goofy tone when communicating with CAs as they find it to be a fun activity [2], making it difficult for the CA to answer children due to incorrect syntax, mispronunciation [11], or mixed utterances [4]. Multilingual children may have an accent that is typically different from accents supported by the CA. They also tend to mix words from different languages (code-switching) as they are unable to have a handle on switching between languages due to their young age [19]. While conventional screen media content provides age-appropriate and partially interactive communication sessions [18], contemporary voice agents are not attuned to child friendly conversations [2]. Thus, we propose *attuning the conversation to children's code switching and accent variations* as a design goal in order to maintain a seamless conversation flow.

FUTURE DIRECTIONS

Despite the popularity of CUIs in domestic settings [2, 11], their potential to be used as assistive interlocutors to learn a secondary language has not been fully realized. With our initial explorations in the domain of CUIs as assistive interlocutors for young children, we present a preliminary, yet prioritized list of high-level objectives which can guide the design of CUIs to aid bilingual children's secondary language learning. The application of such a CUI is beyond language learning by bilingual children, and can branch out to learning a language as a part of curricular requirement for older children. Due to the restrictions of scope and page limit, our proposed design goals are relevant to CUIs which are used exclusively by children, hence we do not account for situations which can have multiple people from the same household interacting with them [10]. Future work in this domain should investigate the impact of culture on the interaction of bilingual children with a CA [11]. Further research can also investigate the prospects of using CUI in conjunction with screen media to promote a greater level of engagement of children with the virtual interlocutor.

REFERENCES

[1] Matthew P Aylett, Benjamin R Cowan, and Leigh Clark. 2019. Siri, echo and performance: You have to suffer darling. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–10.

[2] Erin Beneteau, Ashley Boone, Yuxing Wu, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2020. Parenting with Alexa: Exploring the Introduction of Smart Speakers on Family Dynamics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[3] 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*. 221–223.

[4] Katja F Cantone. 2007. *Code-switching in bilingual children*. Vol. 296. Springer.

[5] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. 2019. What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[6] Wikipedia contributors. 27 January 2021. Conversational user interface. https://en.wikipedia.org/w/index.php?title=Conversational_user_interf

[7] Susan E Dieterich, Mike A Assel, Paul Swank, Karen E Smith, and Susan H Landry. 2006. The impact of early maternal verbal scaffolding and child language abilities on later decoding and reading comprehension skills. *Journal of School Psychology* 43, 6 (2006), 481–494.

[8] 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.

[9] Alice Ann Howard Gola, Melissa N Richards, Alexis R Lauricella, and Sandra L Calvert. 2013. Building meaningful parasocial relationships between toddlers and media characters to teach early mathematical skills. *Media Psychology* 16, 4 (2013), 390–411.

[10] Yue Huang, Borke Obada-Obieh, and Konstantin Beznosov. 2020. Amazon vs. my brother: How users of shared smart speakers perceive and cope with privacy risks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[11] Silvia B Lovato. 2019. *"Hey Google, Do Unicorns Exist?": Children Finding Answers through Conversational Agents*. Ph.D. Dissertation. Northwestern University.

[12] Christine Murad and Cosmin Munteanu. 2019. " I don't know what you're talking about, HALexa" the case for voice user interface guidelines. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. 1–3.

[13] Katherine O'Doherty, Georgene L Troseth, Priya M Shimpi, Elizabeth Goldenberg, Nameera Akhtar, and Megan M Saylor. 2011. Third-party social interaction and word learning from video. *Child development* 82, 3 (2011), 902–915.

[14] Lori Skibbe, Michelle Behnke, and Laura M Justice. 2004. Parental scaffolding of children's phonological awareness skills: Interactions between mothers and their preschoolers with language difficulties. *Communication Disorders Quarterly* 25, 4 (2004), 189–203.

[15] Georgene L Troseth, Gabrielle A Strouse, Brian N Verdine, and Megan M Saylor. 2018. Let's chat: On-screen social responsiveness is not sufficient to support toddlers' word learning from video. *Frontiers in psychology* 9 (2018), 2195.

[16] Yunhan Wu, Daniel Rough, Anna Bleakley, Justin Edwards, Orla Cooney, Philip R Doyle, Leigh Clark, and Benjamin R Cowan. 2020. See What I'm Saying? Comparing Intelligent Personal Assistant Use for Native and Non-Native Language Speakers. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–9.

[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. 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.

[19] W Quin Yow, Jessica SH Tan, and Suzanne Flynn. 2018. Code-switching as a marker of linguistic competence in bilingual children. *Bilingualism: Language and Cognition* 21, 5 (2018), 1075–1090.

This figure "sample-franklin.png" is available in "png" format from:

http://arxiv.org/ps/2103.09228v1