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

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

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

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

Aisling Kelliher
Department of Computer Science, Virginia Tech
Blacksburg, VA, USA
aislingk@vt.edu

ABSTRACT

Children in international and cross-cultural families in and outside of the US often learn and speak more than one language. Challenges can arise for these children in terms of communicating with other children and being able to fully participate in school and society using the primary country language, in developing relationships with distant relatives in other languages, and with the lack of opportunities for practising additional languages within a small community of speakers. Recent research shows that some parents 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. We leverage the qualities of screen media in aiding children with language learning, and try to translate those qualities into the design of a CUI for children to explore the potential of designing conversational user interfaces which can double as assistive language aids. By reviewing the 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 present a preliminary list of objectives to guide the design of conversational user interfaces for young children's bilingual language acquisition.

CCS CONCEPTS

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

KEYWORDS

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

1 INTRODUCTION

Many international parents in the US raise children who speak more than one language [2, 5]. Challenges can arise for these children in terms of communicating with other children in the primary country language, developing relationships with distant relatives in other languages, and a lack of opportunities for practising additional languages within a small community of speakers [5]. Recent research shows that some parents 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, in the US, much 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.

Children's communication with conversational user interfaces (CUI) is a relatively nascent domain of research [2, 9, 12]. 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" [7]. Different terms appear in the literature for various types of CUIs including voice assistants [2, 19], intelligent personal assistants [6, 18], smart speakers [2], voice user interfaces [14], voice based conversational agents [12] and conversational agents [3, 6]. 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.

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 in 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 learn 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 a CUI for children.

2 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 [6]. Instead, children value the familiarity [10, 17] and interactivity [3] offered by screen media characters which are not a part of present day CUIs. To find out more about screen use and the lived experience of parents raising multilingual children, we conducted interviews with twelve parents in the US who belong to a subset of immigrant and non-US parents raising young bilingual children. We present some preliminary insights from those interviews in the form of CUI design opportunities and expand on the design ideas we generated in a subsequent ideation session [3].

3 CUI DESIGN IDEAS FOR CHILDREN

In the subsequent sections, we present a group of orientations for the design of CUIs as assistive interlocutors for young children's bilingual language acquisition.

3.1 Language Chunks

Children of four of the parents in our interview study experienced speech delay, which the parents described as being due to the cognitive load of understanding and communicating in more than one language. P3 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 P7 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" (P7)

Short and succinct utterances by intelligent personal assistants have been found to be helpful for non-native language speakers [18]. 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 [12]. 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.*

3.2 Familiar Language Scaffolding

Children of different ages and developmental delays need varying levels of language scaffolds [8]. While socially contingent video

viewing sessions can effectively promote preschoolers science learning [20], 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 [17] such as verbal scaffolding by mothers of 3 to 4 year old children, which can result in richer verbal interactions and better language skills [16]. One of the participant in our study mentioned how her husband provided verbal scaffolding during their daughter's secondary language learning sessions:

> "My husband puts on poems and verbal programs to get her more fluent with our native language. He has also installed an app 'Urdu Seekhiye' (learn Urdu) on his phone which has all Urdu alphabet and words, so when she sits with him they can swipe along to learn them. She has learned vegetable names from it, so she goes and brings them from the fridge and kitchen cabinet when we ask her to use our language. It is kind of a play and pretend thing where she goes and comes back with the stuff that she has seen on the phone screen." (P11)

P11 mentioned that they have minimal spoken communication with or in the presence of their child due to their busy schedules, and they do not have co-located extended family members, who can converse with the children using their native language.

> "We both don't speak a lot with her because we both were busy with our studies, so I feel she has also lacked behind because of that. But when i went to [home country], my father spent time with my daughter, and she could speak sentences in one month, so it was like really good thing for me that she learned so much." (P11)

Thus, we propose that training a CA with attributes of language such as familiar voice, talking style, and commonly used words from the child's native language could 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 suggest that CUI should be trained 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 the literature [10, 17] about the positive effects of children's para-social relationship with media characters on their learning due to the familiarity offered by them.

3.3 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. This phenomenon was acknowledged by P6 and P12, who considered age to be a primary factor in determining the engagement of the child with CA. P1 thought that young children's attention span was very limited to be engaged with someone who is not physically present with them. 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 parents in our study 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". (P3)

Screen media offers interactivity and learning encounters 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 CUIs to be beneficial for children's language learning. Thus for engaging young children in meaningful dialogue, it should be 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.

3.4 Social Communication Cues

Several interview 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*).

"This creature called 'Blippi', so there is this man and he interacts with the kids and my son kind of interacts with him too, when Blippi says 'hi' he (son) says 'hi', and when he says 'bye' my son says 'bye', like he tries to copy Blippi, and learns a lot from him too". (P3)

Children as young as two years old can learn new words by being a passive part of a social interaction, where they can hear two individuals conversing with each other [15]. Additionally, on screen social cues are reported to be associated with greater child engagement [17].

"Kids do learn from each other, and obviously our kids are not going to school or daycare so they can watch different videos and try to imitate according to that." (P11)

With respect to the ability of screen media content to successfully engage and acquaint children with new words, we envision a scenario where a CA could stage a show with different voice characters having conversations among themselves. In this way, the CA could also involve the children in language learning without assistance from family members.

3.5 Code Switching And Accent Variations

Young children tend to use unclear commands or silly voices when communicating with CAs as they find the encounter to be a fun activity [2], making it difficult for the CA to answer children due to incorrect syntax, limited vocabulary [13], mispronunciation [12], 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 [21]. While conventional screen media content provides age-appropriate and partially interactive communication

sessions [20], 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 orientation in order to maintain a seamless conversation flow.

4 FUTURE DIRECTIONS

Despite the popularity of CUIs in domestic settings [2, 12], 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 part of curricular requirements for older children. Due to the preliminary nature of the work and restrictions of scope, our proposed design orientations 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 [11]. Future work in this domain should investigate the impact of culture on the interaction of bilingual children with a CA [12]. 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] Kids Count Data Center. 11 January 2018. The Number of Bilingual Kids in America Continues to Rise. https://datacenter.kidscount.org/updates/show/184-the-number-of-bilingual-kids-in-america-continues-to-rise

[6] 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.

[7] Wikipedia contributors. 27 January 2021. Conversational user interface. https://en.wikipedia.org/w/index.php?title=Conversational_user_interface&oldid=1003133450

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

[13] Ivonne Monarca, Franceli L Cibrian, Angel Mendoza, Gillian Hayes, and Monica Tentori. 2020. Why doesn't the conversational agent understand me? a language analysis of children speech. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*. 90–93.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.