The Last Decade of HCI Research on Children and Voice-based Conversational Agents

Radhika Garg
School of Information Studies
Syracuse University
Syracuse, NY, USA
rgarg01@syr.edu

Hua Cui*†
College of Computing and Informatics
Drexel University
Philadelphia, PA, USA
hua.cui@drexel.edu

Bo Zhang*
School of Information Studies
Syracuse University
Syracuse, NY, USA
bzhang49@syr.edu

Spencer Selingson*
School of Information Studies
Syracuse University
Syracuse, NY, USA
sseligso@syr.edu

Martin Porcheron
Computational Foundry
Swansea University
Swansea, UK
m.a.w.porcheron@swansea.ac.uk

Leigh Clark
Computational Foundry
Swansea University
Swansea, UK
l.m.h.clark@swansea.ac.uk

Benjamin R. Cowan
School of Information &
Communication Studies
University College Dublin
Dublin, Ireland
benjamin.cowan@ucd.ie

Erin Beneteau
Information School
University of Washington
Seattle, Washington, USA
benjamin.cowan@ucd.ie

ABSTRACT

Voice-based Conversational Agents (CAs) are increasingly being used by children. Through a review of 38 research papers, this work maps trends, themes, and methods of empirical research on children and CAs in HCI research over the last decade. A thematic analysis of the research found that work in this domain focuses on seven key topics: ascribing human-like qualities to CAs, CAs' support of children's learning, the use and role of CAs in the home and family context, CAs' support of children's play, children's storytelling with CA, issues concerning the collection of information revealed by CAs, and CAs designed for children with differing abilities. Based on our findings, we identify the needs to account for children's intersectional identities and linguistic and cultural diversity and theories from multiple disciples in the design of CAs, develop heuristics for child-centric interaction with CAs, to investigate implications of CAs on social cognition and interpersonal relationships, and to examine and design for multi-party interactions with CAs for different domains and contexts.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

*All four authors contributed equally to the paper

†Contributed to this work during her affliation with School of Information Studies, Syracuse University

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

KEYWORDS

conversational agents, voice agents, voice interfaces, voice user interfaces, smart speakers, virtual assistants, family, children, literature review, systematic literature review, parents

ACM Reference Format:

Radhika Garg, Hua Cui, Bo Zhang, Spencer Selingson, Martin Porcheron, Leigh Clark, Benjamin R. Cowan, and Erin Beneteau. 2022. The Last Decade of HCI Research on Children and Voice-based Conversational Agents . In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3491102.3502016

1 INTRODUCTION

Recent advancements in speech technology have made the use of voice assistants like Amazon's Alexa, Apple's Siri, and Google's Google Assistant more commonplace [27, 52, 102, 109]. Voice-based technologies, often called "conversational AI", "dialogue systems", "virtual assistants" or "conversational agents" [80, 99, 109], have increasingly been adopted by children. Voice-based Conversational Agents (CAs) are used by children for a range of purposes, including to conduct small talk or express emotions [52], to support learning and reading [51, 148], and to foster communication with other children and other members of the household [13, 125].

Designing CAs for children presents several challenges. For instance, rather than being static, children's needs and preferences change as they move through different stages of cognitive development [96]. Compared to those of adults, their interactions with CAs also are more difficult because of less precise articulation, a more limited vocabulary, and fewer strategies for modifying and adapting their language [15, 52]. Systematic and summative reviews in HCI and related fields have provided researchers with a clear mapping of trends, research gaps, and outstanding challenges in the recent rapid growth of topics such as speech interfaces (e.g., [33, 37, 41]) in support of directions for future research. Although it is clear that there are specific challenges to researching and designing CAs for children, with research focusing on various aspects of design and development over the last decade, there is no clear mapping or synthesis of themes and challenges in this area. Our work aims to fill this gap by providing a comprehensive systematic review of the child-related CA literature over the last decade, identifying key themes and trends. Furthermore, we use these insights to inform suggestions on directions for future research. In particular, we aimed to answer two key questions:

RQ1: What is the current state of HCI research on children and voice-based CAs?

RQ2: What research areas relating to children and voice-based CAs should be addressed in the future studies in this domain?

Our review identified 38 relevant articles published over the past decade. In this review, we identify the seven salient themes, such as (1) the use and role of CAs in the context of home and family, (2) supporting children's learning with CAs, (3) ascribing human-like qualities to CAs, and (4) CAs for children with differing abilities. Our analysis further revealed that over the last decade the research has been overwhelmingly done in the Global North, and only a minority of papers have focused on teenagers, specifically older teens (15- to 17-year-olds). With respect to future research agendas, our findings highlight immediate opportunities for work that focuses on diversifying participants and methods, designing multi-party interactions and culturally informed CAs, examining ethical challenges that emerge as children engage with CAs, and supporting the children through CAs in contexts such as mental well-being, online-safety, and games.

2 RELATED WORK

Before turning to our own study, we examine (1) how the HCI field has framed CAs and (2) previous literature reviews focusing on CAs in HCI.

2.1 Conversational Agents in HCI Research

Conversational agents (CAs) are systems capable of engaging in dialogue with users [73, 139, 143]. This definition includes both text-based and speech- or voice-based modalities for input and output. For voice-based CAs, the key components include the use of automatic speech recognition (ASR) and natural language understanding (NLU) components to recognize the intent of a user in response to commands, dialogue management to monitor the dialogue state, dialogue history, and the capability to select an appropriate dialogue action based on these elements. Language output is generated by a language generation component (NLG), which is then converted to speech using text-to-speech synthesis (TTS) components (further details can be found in [62, 112]). Text-based chatbots are similar in their design, yet do not include ASR and TTS components but instead use text as the primary modality of input and output.

In this paper, we use the term "conversational agent" to encompass both embodied and non-embodied CAs, which can be found in screen-based interfaces and in speakers that users can interact with through spoken utterances. While robots have verbal communication abilities similar to those of CAs, their affordances and capabilities are vastly different (e.g., physical movement, non-verbal expressions), which leads to the complication of speech by other factors. For example, children might personify the embodied voice interfaces such as a robot more as compared to non-embodied CAs. Since we do not currently understand the difference or if there is a difference, we took a conservative approach and focused our analysis on non-embodied CAs.

Therefore, we do not include robots in our use of the term CA–an approach followed as well by prior literature reviews on the state of speech in HCI research [27].

2.2 Related Systematic Literature Reviews of HCI Research

Systematic literature reviews (SLRs) aim to synthesize and summarize the findings on a research topic (e.g., [18, 27, 33, 66, 77, 116]). More specifically, SLRs serve as a tool that enables researchers to identify open research questions in a field and justify future research in specific areas [75]. Summaries of previous research generated through SLRs also help to inform funders wishing to support new research [22] and help editors judge the merits of publishing new studies [153]. SLRs also provide valuable insights for researchers into methods, findings, authors, and publication venues in specific fields [128]. While some researchers produce broad-ranging literature reviews using databases like ACM Digital Library (e.g., [37]), others search specifically through sub-fields and venues such as HCI, CSCW, and UbiComp (e.g., [46]). For this study, we followed the latter approach to understand the state of research on children's interactions with CAs in the field of HCI. We followed the SLR approach as suggested by Pittaway et al. [97], enabled us to identify existing gaps in the research field and to make suggestions for addressing them.

Recently, two systematic reviews of CAs in HCI research have been published. Rheu et al. [109] conducted a systematic review of 29 experimental studies that investigated the effects of CAs' and users' characteristics on trust. Their article search was conducted in the ACM Digital Library and Web of Science databases for articles published between 2000 and 2019; it found five key design themes that influenced users' trust: the "social intelligence of the agent, voice characteristics and communication style, look of the agent, non-verbal communication, and performance quality" [109]. The review highlighted the responsibility of designers to respond to stereotypes and the building of social norms in their work. With the aim of mapping trends, themes, and challenges within speech-based research in HCI, Clark et al. [27] conducted a systematic review of 99 articles across a range of HCI publication venues, focusing on non-embodied CAs. Searching the ACM Digital Library and the ProQuest and Scopus databases–an approach that we also follow in this paper—the work found that speech research in HCI, although fragmented, tends to coalesce around nine key topics: system speech production, modality comparison, user speech production, assistive technology and accessibility, design insights, experiences with interactive voice response (IVR) systems, the use of speech technology for development, people's experiences with intelligent personal assistants (IPAs), and the ways user memory *voice system, human system dialog*, spoken dialog system, natural language dialog* system, human computer dialog*, speech interface, intelligent personal assistant, spoken human machine interaction, conversational agent, natural language interface, conversational interface, conversational system, automated dialog* system, voice interface, interactive voice response system, conversational dialog* system, human machine dialog*, voice user interface, embodied conversational agent, voice assistant, personal voice assistant, personal voice agent, smart speaker, chatbot, conversational agent, conversational user interface, virtual agent, personal voice assistant, personal voice agent, smart speaker, voice system, digital home assistant, alexa, google home, siri*

Figure 1: Selected technology-specific keywords, with asterisks (*) denoting truncation to account for alternative spellings, e.g., dialog versus dialogue.

child, teen*, preschooler, infant, kid, tween, family, toddler*

Figure 2: Selected children-specific keywords, with asterisks (*) denoting truncation to account for alternative spellings, e.g., teen versus teenager.

affects speech interface interaction. From their analysis, the authors identified key challenges within the field, including the need to reach critical mass on a consideration of these themes, increase design-based insights, develop core theories for speech-based interaction, and broaden the scope of users and contexts within speech system research [27].

Though theirs is not technically a literature review, Xu and Warschauer [147] reviewed 535 voice apps on the market that support early literacy development, based on a framework of factors (i.e., learning content and goals, interactions and gamifications, and conversation design) that can influence the educational benefits young children receive from voice-based apps. They found that the apps covered a wide range of learning activities and that the majority used storytelling as a strategy. While almost half lacked ongoing interactions, the apps which did include them were interactive and included aspects of gamification to increase playfulness. Most utilized multiple-choice prompts rather than open-ended prompts. Finally, while many of the apps provided feedback based on users'/children's responses, the quality of the feedback varied, as very few apps provided elaborate feedback, adopted strategies to prevent conversation breakdown, or scaffolded interactions to ease children's learning and communication.

2.3 Research Aims and Contribution

While research on children's interactions with CAs has grown over the past decade, there remains a gap in current systematic reviews, in that they do not directly address or identify specific work related to this area. Our work contributes by reviewing the state of current research on CAs and children, identifying key themes, methods, and publication trends. Based on this investigation, we identify core challenges to be addressed by future research on child-based CA interactions.

3 METHOD

To address the research questions, we conducted a systematic review of the HCI literature on the topic of speech and voice technologies specific to children within the past decade (2011-2020). Our review process followed the method suggested by Pittaway et al. [97] and covers a decade of research, in line with other recent literature reviews (e.g. [66, 106]). In addition, this time period was selected because of the increased commercial availability of conversational agents in phones and Internet of Things (IoT) devices [123, 135] during these years, which has made the study of user interaction with these technologies increasingly important.

3.1 Search Procedure

In February 2020, following the lead of a recent systematic review of the HCI literature regarding speech interfaces [27], we searched for relevant publications in February 2020 in the ACM Digital Library (ACM DL) and the ProQuest (PQ) and Scopus (SP) databases. We also restricted our search to core HCI venues as indicated by Google Scholar, Thomson Reuters, and Scimago journal rankings. Our initial list of relevant keywords for the search was taken from the literature review conducted by Clark et al. [27]. The author team collaboratively expanded on and refined this list after reviewing author-defined keywords and additional terms that appeared in abstracts of the relevant literature. The final set of keywords that served as seeds for our final search are listed in Figure 1. As we focused on children's interaction with voice technologies, we also added terms pertaining to children, such as child, teen, infant, and kid (see Figure 2). The keywords were searched as exact phrases or modified to account for plural forms (e.g., teen/teens) or variation in spellings (e.g., dialogue/dialog, teen/teenager), depending on the database. The individual search strings in Figures 1 and 2 were combined using the Boolean operator `OR` and then the two final strings of terms were joined with the Boolean operator `AND`. The final search string was as follows:

> (("*voice system*" `OR` "*intelligent personal assistant*"....<keywords shown in Figure 1>) `AND` ("*teen*" `OR` "*child*"...<keywords shown in Figure 2>))

3.2 Inclusion and Exclusion Criteria

Searches were limited to publication keywords and terms appearing in titles and abstracts in journal articles and conference papers. This yielded 7,800 papers (ACM DL=6,789, PQ=106, SP=905). After removing duplicates, we had 1,311 unique entries (ACM DL=998, PQ=13, SP=300).

Several inclusion and exclusion criteria were employed to filter the papers further. Inclusions consisted of papers that fulfilled the following criteria; they had to be (1) full papers written in English that were published in conference proceedings and journals in the last ten years (2011 to 2020) (Notes from both CHI and IDC were included because of their similarity in status); (2) papers primarily

Table 1: Publication count by year and venue

Venue	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
Computer Speech and Language	1										1
Computers in Human Behavior										1	1
Conference on Human Factors in Computing Systems (CHI)	1			1			2	1	3	5	13
Designing Interactive Systems (DIS)								1	1		2
Interaction Design and Children (IDC)			1		2			2	2	3	10
Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)									1	2	3
International Conference on Multimodal Interaction (ICMI)			1								1
International Journal of Child-Computer Interaction									1		1
INTERSPEECH	1	1				1					3
Transactions on Computer-Human Interaction (TOCHI)		1						1			2
Transactions on Speech and Language Processing	1										1
Total	4	2	2	1	2	1	2	5	8	11	38

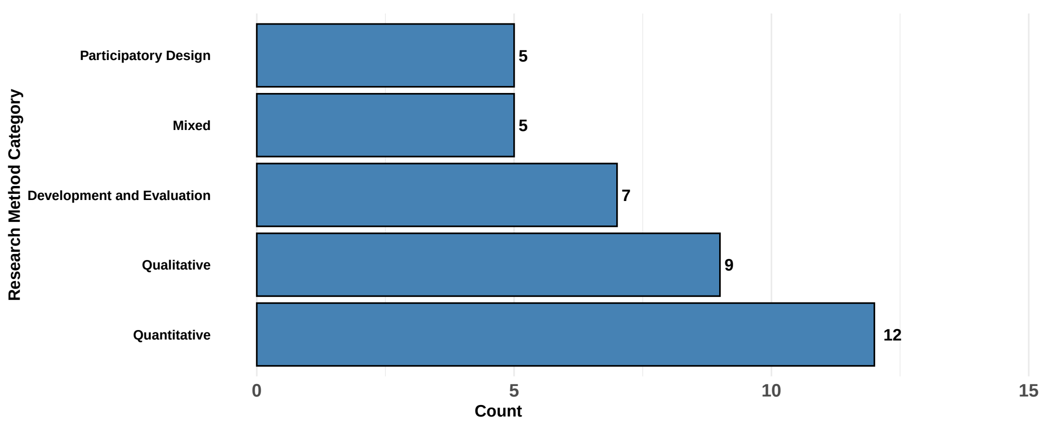

Figure 3: Applied research methods

investigating speech input, speech output, and/or dialogue; and (3) papers that included children as participants or were about child-oriented CAs.

We excluded (1) papers that did not use speech as the primary mode of user input or system output to the user or as the medium for a two-way dialogue (e.g., [92]), (2) as was the case with [27], papers focusing on robots (e.g., [42, 61], due to potential confounding effects (e.g., physical movement, non-verbal expressions), (3) non-full or non-peer-reviewed papers, such as works in progress, extended abstracts (e.g. [40]), workshop papers [118, 129]), reports from panels, and magazine articles (e.g., [59]).

After applying the above criteria to our search results, 38 papers were selected for the final analysis. We reviewed each paper and recorded: (1) the central goal or focus of the study, (2) the type of device that was the focus of the study and the mode of interaction supported by the device (i.e., user speech input, system speech input, or dialogue), (3) details about the sample population or participants, (4) the data collection methods used (e.g., interviews, questionnaires, observations) and the type of study (e.g., qualitative, quantitative, mixed), and (5) the key findings. We also extracted the list of authors, the publication year, title, publication venue, and geographic region where the research was conducted.

3.3 Data Analysis

We conducted an inductive, thematic analysis to identify topic themes in the literature [19]. Two authors began the process by coding the topical themes independently for every paper and then engaged in discussions to resolve inconsistencies and differences. Through this collaborative effort, the final themes, and, where applicable, sub-themes, emerged. In cases when the focus of a paper spanned more than one theme or sub-theme, the paper was categorized by its primary research theme, as deemed by the two coders. For example, one might argue that the theme "The Use and Role of

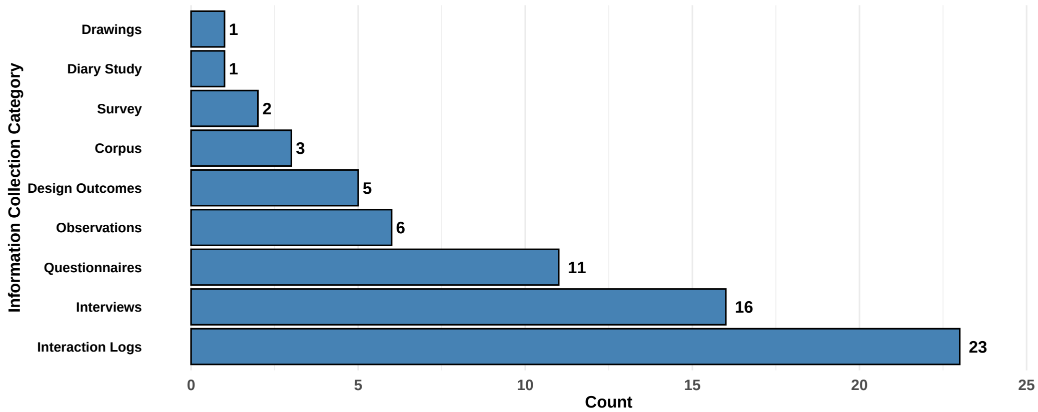

Figure 4: Approach to data collection

CAs in the Home and in the Family" can overlap with the theme "Supporting Children's Play", since children often play when they are at home and with their family. The coders faced this dilemma in regards to the work by Storer and colleagues Storer et al.[125]. We grouped only those papers in the former category which focused on the the impact CAs can have on family members and family dynamics and/or general use of device by family members while they are at home, whereas the primary focus of the papers in the latter category was on one specific activity, i.e., children's play. While [125] refers to opportunities for children and parents to play games together, the main focus of the paper was to understand how CAs get integrated in homes of mixed-visual-ability families. As a result, we allocated [125] under the theme of 'The Use and Role of CAs in the Home and in the Family." In total, we identified eight core themes (cf., Table 5), the results of which we present next.

4 FINDINGS

In this section we present the emergent research themes from our review, but first we cover a number of descriptive statistics.

4.1 Descriptive Statistics of the Papers Reviewed

4.1.1 Publication Trends and Venues. As shown in Table 1 the number of publications has increased overall, and especially in the last three years, rising from five in 2018 to 11 in 2020. The indexed publications appeared in 11 different venues. The most common venues were *CHI* (13 papers), *IDC* (10), *Ubicomp/IMWUT* (3), *INTERSPEECH* (3), *DIS* (2), and *TOCHI* (2). Four venues, including *International Conference on Multi Modal Interaction (ICMI)*, *Computer Speech and Language*, *Computers in Human Behavior*, *International Journal of Child-Computer Interaction*, and *Transactions of Speech and Language Processing* contributed just one relevant publication each.

4.1.2 Research Methodologies. As visualized in Figure 3, most ($N = 12$) of the reviewed studies were *quantitative* in nature; seven of these used an *experimental design* and were done in a lab setting, where the participants interacted with a mock system through a Wizard of Oz (WoZ) approach [100] or a working prototype and remaining five developed *quantitative models* – including machine-learning models or regression models – using some form of corpus (e.g., stories told by children) or logs of child interactions with a system. In WoZ approach a human "wizard" simulates an intelligent system and interacts with a user through a real or mock interface [142]. ($N = 9$) of the reviewed studies used qualitative methods (e.g., interviews, interaction logs). Five other studies used mixed-methods approaches to address their research questions. Seven papers *developed and evaluated* a prototype system through a user study, while a further five studies involved *participatory design* work with children, families, or parents. We note that qualitative and participatory design are not mutually exclusive. However, we separate them into categories to underline the focus on co-designing in the latter category.

Out of the total 38 papers, many papers included a combination of measures in conducting their research, as shown in Figure 4. Sixteen papers included interviews with the participants, which aimed to elicit children's preferences or experiences of interacting with a device (e.g., [52, 79]). Eleven used questionnaires involving scales (e.g., (e.g., [6]) and/or questions designed to elicit specific metrics (e.g., the preference of the agent [9]). In the former case, the studies used pre-existing scales (e.g., [6]), custom-scales adapted from pre-existing scales (e.g., [44]), or a combination of customized and pre-existing scales (e.g., [113]). Six studies used *observation* as a mode of data collection. In these studies, the researchers directly observed the participants as they interacted with a prototype/device or used a WoZ approach to elicit specific metrics. Another set of 23 studies collected and analyzed audio, video, and/or text-based interactions logs *interactions logs* (e.g., [15, 52, 154]). Five studies analyzed data collected from participatory design activities (e.g., design outcomes and audios/videos of design sessions) (e.g., [51, 95]). Other data collection methods included *drawings* ($N = 1$; [149]), surveys, where-in the questionnaires did not comprise of pre-defined or customized scales to measure a phenomenon ($N = 2$; [25, 78])), *diary studies* ($N = 1$; [50]), and a *corpus* for machine-learning modeling ($N = 3$; [24, 72, 151]).

4.1.3 Participants. Twenty-four articles specified the country in which the study was conducted and/or from which data were drawn,

Table 2: Location of the children on which the papers focused

Country	Relevant Papers	Total
USA	[13–15, 25, 50–52, 81, 85, 94, 104, 113, 125, 127, 144, 150, 154]	17
France	[6, 53, 119]	3
UK	[44, 98]	2
Korea	[95]	1
Italy	[122]	1
Not mentioned	[4, 9, 16, 24, 54, 72, 78, 79, 89, 117, 137, 148, 149, 151]	14

as seen in Table 2. Out of these 24 articles, the *United States of America* was the most common study site (N = 17). *France* (N = 3) and the *UK* (N = 3) were the next two most common countries where studies took place. Four studies [50, 95, 117, 125] did not collect data directly from children, but instead collected data about them from their parents.

We also collated the age groups of the children in the studies covered by the reviews (see Table 3). Most studies focused on children from more than one age group, based on the categorizations offered by the CDC (Centers for Disease Control and Prevention) [45]. For example, [13–15] all included families with children under 18. Only 19% of the studies (i.e., [16, 54, 94, 104, 119, 122, 127]) contained children from only one age group.

4.1.4 Engagement with Theories or Theoretical Frameworks. 14 out of 38 papers that were reviewed engaged with theories or theory-based concepts to ground their research questions, to frame their study in the designing of a prototype or working system, or in their data analysis. We elaborate on these types of engagement with theories or theoretical frameworks next:

Four studies used theories to arrive at their research questions. Aeschlimann et al. [4] posed research questions inspired by by *Social Agency Theory* [83], which argues that the "social cues of a computer (e.g., modulated intonation, human-like appearance) encourage people to interpret the interaction with a computer as being social in nature" [4]. Spitale et al. [122] investigated the role of *social embodiment*, defined as the "idea that the embodiment of a socially interactive agent plays a significant role in social interaction" [34] in children's linguistic assessment. Beneteau et al. [15] used the concept of *joint media engagement* [124] to frame their study and illustrate how communication breakdowns occur and are repaired as family members interact with smart speakers in the home. In another paper, Beneteau et al. [13] used two theoretical frameworks—*joint media engagement* [124] and *parental mediation* [30, 155]—to explore technology, parent-child dynamics, and parenting practices with smart speakers.

Seven studies used theories in designing aspects of their prototypes or working systems. Alaimi et al. [6] used convergent- and divergent-thinking questions, which the CA asked of its child users; these were based on *Question-Answer Relationship* (QAR) [47], which defines the relationship between a question generated from the reading of a text and its answer. Ruan et al.[113] grounded the design of an agent for learning mathematics in a literary theory that promotes the benefits of narrative-based learning [3]. Tewari and Canny [127] built features into the conversational script of CAs using the elements from the framework of Hart and Risley [55] that are known to improve the quality of interactions between parents and children. Pantoja et al.[94] explored the design of CAs to support high-quality social play in terms of *Theory of Mind* [17]. In their design of a conversational reading partner, Xu and Warschauer [148] included three components of conversational guidance, including questions that open up conversations, contingent feedback, and language scaffolding; according to Snow [121], when these components are part of a parent's conversational guidance, they engage children in more cognitively and linguistically beneficial interactions. The design of ECHOES [98] was informed by and validated against the *SCERTS* model [103], which addresses the core challenges related to autism spectrum conditions. The ECHOES environment utilizes FAtiMA planning architecture *FAtiMA planning architecture* [36], which makes it an autonomous planning-based agent. It is based on the following componenets: (1) AI planning techniques [114], (2) an emotional model derived from the *OCC cognitive theory of emotions* [93], and (3) the *appraisal theory of emotions* [71]. To develop a conversational multi-media tutor, Ward et al. [137] applied the principles of *Questioning the Author*—an approach developed by Isabel Beck and Margaret McKeown, which challenges students to "think about and integrate new concepts with prior knowledge to construct enriched mental models that can be used to explain and predict scientific phenomena" [137]. The system was also informed by *Discourse Comprehension Theory* [69], *Social Agency Theory*, and by the work of Jean Piaget, Lev Vygotsky and Jerome Bruner that gave birth to *Social Constructivism* [136].

Three studies grounded their analysis in theories. Xu and Warschauer [149] used various frameworks when analyzing the properties that children attribute to CAs; these frameworks came from Kahn et al. [63], Kim et al. [68], van Duuren and Scaife [132], and Melson et al. [86]. Beneteau et al. [14] built on Rogers' concept of "*near peers*" to illustrate how families learn about smart speakers. Finally, to annotate child's interjections during storytelling Lejeune et al. [72] used *Dynamic Interpretation Theory* [21].

Another set of 11 papers out of the total 38 papers only mentioned theories or theory-based concepts only in passing in:

- *Related work (5):* Serban et al.[119] mentioned the *uncanny valley* [88], *persona effect* [74], and the *ZPD* (Zone of Proximal Development) [136]. Beneteau et al. [14] referred to *The Unified Theory of Acceptance* and *Use of Technology* [133] and *The Senior Technology Acceptance and Adoption Model* [108]. [50–52] referred to *parental mediation theory* [30, 131]. Garg and Sengupta [51] also referred to the *persona effect.*
- *Method (2):* Park and Lim [95] used *cultural dimension theory* [57] to justify the peculiarities of conducting research in

Table 3: Age group of the children on which the papers focused (age categories are those proposed by the CDC [45])

Age group	Relevant papers	Total
Infants (0–1 year)	[13–15, 78, 95, 117]	6
Toddlers (1–3 years)	[13–15, 25, 78, 95, 117]	7
Preschoolers (3–5 years)	[4, 9, 13–15, 25, 50, 52, 54, 78, 79, 94, 98, 117, 125, 127, 141, 148–150, 154]	21
Middle Childhood (6–11 years)	[4, 6, 9, 13–15, 24, 25, 50–53, 78, 79, 89, 98, 104, 119, 122, 125, 137, 141, 144, 148–151, 154]	27
Young Teens (12–14 years)	[6, 13–16, 24, 44, 51–53, 81, 89, 98, 125, 144, 150, 151, 154]	18
Teenagers (15–17 years)	[13–15, 44, 52, 81]	6
Age not specified	[72, 113]. Note: [113] worked with third- to fifth-graders and did not mention age, which varies a little across countries and regions	2

Table 4: Devices and types of interaction in the reviewed papers

	Mock systems through Wizard of Oz studies			Working prototypes			Design studies
Type of device	User speech input	System speech output	Dialogue	User speech input	System speech output	Dialogue	-
Computer-based	-	-	6 (e.g., [119, 137])	-	1 [16]	1 [53]	-
Mobile-based	-	-	-	-	-	1 [78]	-
Tablet-based	1 ([113])	-	-	-	-	2 [6, 54]	-
Smartspeaker-based	-	-	2 ([150, 154])	-	-	10 (e.g., [14, 15, 50, 52])	3 ([25, 81, 95])
Smart toys-based	-	-	-	-	-	1 [85]	-
Display or projection-based	-	-	1 [9]	-	1 ([98])	1 ([127])	-
Other	-	-	2 [4, 122]	-	-	-	4 ([44, 51, 94, 144])

a specific contexts and their selection of participants. Serban et al. [119] used the concept of the *uncanny valley* [88] to justify their decision of designing their own Wizard of Oz platform and interactive narration scenario.

- *Discussion of Implications (3):* Alaimi et al. [6] confirmed a theory-related concept, that a relationship exists between curiosity and children's mastery of question generation mechanics. Xu and Warschauer [148] used the concept of *ZPD* [136] to explain their design recommendation. Garg and Sengupta [51] referred to the *uncanny valley* [88] while discussing the ethical implications of their design recommendations.

4.2 The Devices Involved in the Reviewed Papers

4.2.1 Maturity of Devices. As shown in Table 4, 13 papers in our review used mock devices to conduct WoZ studies with children. Eighteen papers involved existing systems (e.g., Google Home [52], Amazon Echo [15]) or working prototypes developed by the researchers (e.g., [148]). In addition, seven papers focused on engaging children in co-design or participatory design work (e.g., [51, 95]) to understand their preferences concerning CAs. Rader et al. [104] explained how the design of a virtual peer could be informed by findings obtained from the analysis of a corpus of child-to-child language use; however, this paper did not fit into our categories and so has been omitted from Table 4.

Some of the studies spanned across categories listed in Table 4. A system developed Ward et al. [137] can be used autonomously, although the study was conducted using a WoZ approach. Although Chen et al. [25] and Luria et al. [81] both conducted mixed-method studies (noted as such in Figure 4), but also designed storyboards and used them as probes in surveys and interviews. Therefore, we listed these works under the category of "Design Studies" in Table 4. Porayska-Pomsta et al. [98] blended both a human conversation partner and an AI-driven CA; while the CA understood only touch-based user input, the children could engage in a dialogue with the human conversation partner. Hence we included this under the category of system speech output and working prototype. Finally, Tewari and Canny [127] compared the results from a human condition, i.e., when children interacted with researchers without a virtual agent, a WoZ condition, and a fully automated agent condition; we included this paper in Table 4 under "Working prototypes" only.

4.2.2 Types of Devices. In the 38 papers we reviewed, we identified various types of devices that mediated voice-based interactions (see Table 4). Fifteen papers looked into conversational agents that are (meant to be) deployed in smart speakers (e.g., [15, 52]). Nine papers were about computer-based systems where children interacted with systems deployed on a computer or laptop (e.g., [53, 119]). There were six studies that we could not categorize; we labeled "other" because they either did not mention any one specific device (e.g., [4, 51]) or focused on more than one device (e.g., [94, 122]). For example, in [4] human experimenter acted as voice assistant by being present via their voice during the experiments. Spitale and colleagues [122] were interested if embodiment matters, so they checked voice assistant embedded in a toy, an avatar that was displayed on a tablet screen, and a WoZ condition wherein a human acted as a voice assistant.

4.2.3 Direction of Communication. As shown in Table 4, the majority of studies (29 out of 38) studies looked into devices that supported two-way speech-based communication. Of the 29 studies, 12 explored the interactions of children with smart speakers

Table 5: Number of papers per research theme and sub-theme

Research themes	Research sub-themes	Count	Total
Ascribing Human-Like Qualities to CA	-	-	3
Supporting Children's Learning	Learning Through Question-Answering	2	8
	Language Learning or Development	3	
	Science or Math Learning	2	
	Design Insights	1	
Use and Role of CAs in Home and in Family Context	Use of CAs in the Home	9	13
	Design Insights	4	
Supporting Children's Play	-	-	2
Children's Storytelling	Storytelling Strategies	1	2
	Modelling the Roles of Storyteller and Listener	1	
Issues Around Information Collected and Revealed by CAs	-	-	3
CAs for Children with Differing Abilities	Language Training and Assessment of Children with Complex Communication Needs	2	5
	Interaction with Children with Autism	3	
Miscellaneous	-	-	2
			Total: 38

(e.g., [50, 117]), while only two studies examined monologic communication [16, 113].

4.3 Research Themes

This section introduces the results of our thematic coding of the papers. An overview of these is given in Table 5. We identified seven main themes (plus "Miscellaneous"), and some of these themes were comprised of multiple sub-themes (of which there were ten in all).

4.3.1 Ascribing Human-Like Qualities to CAs. Three papers examined if and how children distinguish between CAs and people. Through an interview- and drawing-based study, Xu and Warschauer [149] found that children recognize the cognitive properties of CAs most frequently through behavioral references, the psychological properties of CAs most frequently through references to reciprocity (e.g., when the CA "liked" a child because she was nice to the agent), and the behavioral properties of CAs through biological references, mechanical causality, and fantasy reasoning. While the six-year olds taking part in the study described CAs as technological objects, a considerable number of younger children (i.e., 3- to 5-year-olds) attributed both human- and artifact-like qualities to them. As a consequence, this study suggested developing a "new ontological category" [149] that cuts across prototypical categories of animate and inanimate to describe such agents.

Aeschlimann et al. [4] found that children between five and six years old are able to distinguish between CAs and humans. More specifically, while the type of information shared (e.g., experience-based or knowledge-based) impacted the way children interacted with people, this pattern was not observed in their interaction with CAs. Furthermore, children shared less information with CAs than they did with people, because they believed that CAs do not care much about progressing the conversation. This echoes other work on challenges to progressivity in CAs [43]. Overall, assumptions of the *social agency theory* [83] did not hold true as children children do not impose the same expectations on voice assistants as they do on humans even when voice assistants have human-like features.

Woodward et al. [144], through participatory design sessions, found that children want intelligent user interfaces to be able to handle human-like conversations. For example, children expected these devices to accurately understand input, intentions, and social contexts and then make decisions about how to respond appropriately. The children also expected such interfaces to elicit, recognize, and respond to emotion through speech and other input/output modalities (e.g., gestures, motion). Finally, in cases of error detection, they expected such devices to admit their shortcomings by saying "I do not know" – a human-like trait.

4.3.2 Supporting Children's Learning. In our review, eight studies examined the role CAs can play in children's learning. Two papers in this category looked at how answering questions, a skill that drives curiosity and constructs academic knowledge [105], can be supported by CAs. Using a CA embedded in a web application called the Curiosity Notebook with 10- to 12-year-olds, Alaimi et al. [6] showed that such agents have the ability to foster question-asking skills in children that leads them to ask more complex questions. Tewari and Canny [127], in their work with preschoolers, also designed and evaluated a question-answering pedagogical agent, Spot, using common parenting styles in agent dialogues. The authors found that children under both a WoZ condition and a fully automated agent condition asked more questions and solved more problems than if they played with a familiar person. The agents did not deviate from the task at hand and did not need significantly more hints or explanations under the agent condition than they did under the human condition.

Three papers examined the role CAs can play in the language learning or development of children. Using insights from the analysis of a child-child interaction corpus, Rader et al. [104] suggested using virtual peer CAs rather than using a tutor CAs for teaching "school English" and "school-ratified science talk" to children who speak African American Vernacular English (AAVE). They argued that a CA designed as a peer would avoid creating an "oppositional culture" between the child's home and school environments. In turn, the authors argue this would enable children to learn with a peer who can model, scaffold, and evoke the target language and nonverbal behaviors, i.e., teach them to "codeswitch" among verbal and nonverbal styles in the way that children sometimes do for one another, albeit not reliably. Gómez Jáuregui et al. [53] used an e-learning platform for learning English that supports the describing, building, and playing of a conversational scenario between a learner and a virtual CA that speaks English and recognizes English sentences. The study suggested that inter-ocular distance (the distance between the centers of the pupils of a person) can be used to accurately measure the approach and avoidance behaviors of teenagers reacting to summative evaluations provided by CAs in a classroom. The authors further concluded that these automatically detected approach-avoidance behaviors are negatively correlated with the anxiety levels of the users while learning with CAs. Xu and Warschauer [148] developed and deployed a CA that acted as a joint-reading partner, designing it to boost a child's language development. They found that a combination of open-ended questions and multiple-choice questions, combined with feedback from the CA, kept children emotionally engaged. Even though younger children (3- and 4-year-olds) had some difficulties in communicating with the agent (e.g., difficulties with language production and flow maintenance), they were more interested in the reading partner than the older children (5- and 6-year-olds).

Two of the papers looked into the role that CAs can play in engaging children in science or mathematics learning activities. Ruan et al. [113] designed a tablet-based interactive system in three forms — with narratives, narratives with hints, and narratives with a tutoring CA (to which users could give input by typing or using a speech-to-text button). In this WoZ study, they tested the effectiveness of each approach for engaging third- to fifth-graders in mathematics-learning activities. The authors found that combining narratives with conversational feedback positively influenced the learning outcomes of their participants, but that a passive hint-based system had a negative impact. Ward et al. [137] described the design and architecture of a conversational multi-media virtual tutor for elementary school science, using the principles from *Questioning the Author*. While the study did not formally collect data on learning gains, the authors reported positive impressions on the part of teachers and students after conducting a WoZ procedure to test the implemented system.

Finally, Garg and Sengupta [51] engaged in co-design work with children aged 7 to 12 and their parents to understand their preferences for CAs for in-home learning. The authors found that children expect such agents to possess a personality and an advanced level of intelligence, to support multiple content domains and learning modes, and to engage in human-like conversations. Parents indicated their desire for such devices to include them in their children's learning activities, to enable them to monitor their children's use of the systems, and to foster social engagement.

4.3.3 The Use and Role of CAs in the Home and in the Family. The next theme that emerged in our analysis was the role CAs (can) play in a family context. Nine studies investigated how children use smart speakers in their homes and the challenges they face while doing so. Beneteau et al. [14] reported that families (i.e., children and parents) try to learn about the smart speaker's functionalities through the device itself and through their social circle (family, friends, and acquaintances). Consequently, they positioned the smart speaker as a near-peer [111] to the user and suggested that it should describe its functionality using just-in-time learning (e.g., by providing relevant information based on a user's past failed attempts). Garg and Sengupta [52] found that children (5- to- 7-year-olds) invest CAs with human-like qualities, develop an emotional attachment to them over time, and engage in social exchanges with them. They also showed that children's use over the long-term often shifted from engaging in small talk or to functional tasks [e.g., setting alarms]. Sciuto et al.. [117] found that while parents appreciate children's being able to talk to CAs, they had concerns about the impact these systems may have on children's social interactions. However, Sciuto et al. [117] and Garg and Sengupta [52] found that, initially, younger (5- to 7-year-old) face more difficulties than older children in interacting with CAs because of their verbal intonations and cadences. Beneteau et al. [15] showed that children who belonged to middle- to lower-income families, with the help of parents, employ several repair strategies (e.g., prosodic changes, over-articulation, semantic adjustments and modification) to be able to converse with smart speakers at home. Another study by Beneteau et al. [13] showed that despite the challenges presented by smart speakers, democratic access to them in households impacts family dynamics in terms of: (1) fostering communication (e.g., engaging in speech and language practices that help the device understand the child, expanding communication skills by learning new language), (2) disrupting access to the device (e.g., interrupting one another while interacting with device, regulating use), and (3) augmenting parenting (e.g., leveraging the device as a neutral third-party mediator or as a complement to parenting tasks, and using the device to increase a child's autonomy). Garg and Sengupta's work [50] with Asian Indian and White parents showed that the (non-)use of CAs built into smart speakers by parents around and with children, (non-)use by children, and parental mediation strategies vary by race and socioeconomic status. In preliminary work, Lovato and Piper [78] found that children's use of devices with CAs primarily involved exploring the devices (e.g., for fun or relational purposes), seeking information (e.g., curiosity, location seeking), and performing various functions (e.g., making a phone call). In a later paper, Lovato et al. [79], based on their finding that children used these devices to ask questions out of curiosity, suggested that CAs could be used for children's self-directed learning. However, the authors suggested that current smart speakers could better support question-asking behavior by tailoring their responses to specific users (e.g., children versus adults); by simplifying and decomposing the answers, as children had difficulty understanding complex and nested information; and by developing contextual understanding based on sequences of interactions Storer et al.. [125], through their work with mixed-visual-ability families, found that smart speakers prove to be appropriate tools for supporting more accessible parent/child interactions. Examples include parents' and children's bonding while using smart speakers, including playing games together, sending each other jokes, and learning to use the devices. However, it also created new anxieties about children possibly gaining access to adult content. Therefore, the authors suggested that universal usability should account for the fact that age, like ability, presents different access needs.

Another four studies in this category involved design work aimed at achieving an understanding of the expectations of children and parents towards CAs within the context of home and family. Park and Lim [95], after conducting a participatory user study, suggested that devices should be able to develop an understanding of family dynamics and rituals and adapt their features to social versus private situations (e.g., avoid revealing personal and private information in shared spaces). In a similar vein, Luria et al. [81] conducted speed dating and group interviews with families containing children who were aged twelve or older. The authors proposed that the information people should not receive, the actions they should be able to perform through the agent, and whether they should have control over another individual's data should be based on the users' social roles (e.g., outsiders, children). Fitton et al. [44] conducted a design study with teenagers, who first scripted hypothetical exchanges with smart homes on paper, then prototyped at a higher level of fidelity, using a tablet app with speech output. They found that teenagers wanted agents to help them most to find lost objects and decide what to eat. While the designs assumed parent-like roles, most depicted smart homes in servile roles that are known to have negative impacts on a child's development [64]. Chen et al. [25] used mixed-method interviews and surveys with the parents of 2- to 6-year-olds in eight families that used storyboards as prompts, and examined barriers to the adoption of technology during mealtimes. They found that the parents preferred screen-based technology over voice-based devices and smart objects. In the case of voice-based devices, the personification of CAs led them to believe these devices could intrude upon their relationship with their children (e.g., by disrupting interpersonal interactions and displacing parenting relationships).

4.3.4 Supporting Children's Play. While most of the studies reviewed examined the role that CAs can play in children's structured activities (e.g., question and answer-based learning), two studies investigated developmentally appropriate activities that were less structured. Pantoja et al. [94] conducted 24 design sessions with eight preschool children and compared the roles different CAs can adopt in children's social play. Different configurations of CAs were researched: researcher-controlled CAs, which required researchers to type text to control what the CAs say; portable and screen-based CAs, with speech controlled through an app that could be used by children or researchers; and a "turned-off" portable, tangible agent with no speech capability. The authors found that tangible and portable CAs were the most effective for keeping children socially engaged in play. Consequently, CAs have the potential to support the Tools of the Mind [17] style of play. Furthermore, while being able to control the CA can distract the user from play, and while CAs that cannot take advantage of contextual information can be counterproductive, personification is critical in supporting social play. Gordon and Breazeal [54] designed a virtual agent for tablets that supports a parent-driver by supervising entertainment for children in the car, so that the parents can focus on driving. An exploratory assessment with 4- and 5-year-olds revealed that the challenges of using such a system included speech recognition in a very noisy automotive environment. Children struggled to understand the non-verbal cues of the tablets (e.g., lights and sounds), and the parents were unaware of any problems because they had to focus on driving.

4.3.5 Children's Storytelling. Two papers addressed storytelling with CAs. One paper looked into storytelling strategies that voice-based technologies could adopt. More specifically, Serban et al. [119], a WoZ study, found that a virtual avatar with prosody and facial expression modalities improved children's engagement in interactions during the storytelling sessions. For example, the children responded more often to the agent with full modalities than to the agent with no prosody. Crucially, prosody was found to be less important than facial expressions as a modality. A study by Lejeune et al. [72] proposed a model for understanding and predicting the behaviors and roles of storytellers and their listeners. They found that predictions about whether or not a child is likely to react during storytelling by a CA can be made using a combination of psychological features: the informative or active intent of the speaker, the communicative contributions of the speaker to the conversation, and pattern-mining techniques (frequent closed patterns were computed using an algorithm proposed by Ukkonen [130]).

4.3.6 Issues Involving Information Collected and Revealed by CAs. The review found three papers that identified issues relating to information collected by CAs. Following a demonstration and the interaction of children with connected interactive toys, McReynolds et al. [85] conducted interviews with parent-child pairs to understand privacy expectations and concerns regarding the parental controls. The authors found that children are often unaware that their conversations with a toy might be heard by others and that they often expect flexible interactions and integrated instructions. In their discussion of the implications of the study, the authors raised points about the need to boost parental knowledge about the privacy and security properties of connected toys and the need to increase their willingness to allow their children to play with them, and about the fact that parental controls should be included only when potential ethical matters and issues of trust are recognized. Yarosh et al. [150] and Yuan et al. [154] found that children struggle to reformulate questions and suggested that voice-based devices can help children by providing feedback on what the devices hear, by asking for missing context, by clarifying what the device knows, and by pointing out formulations that are difficult for the system to comprehend. They also found that children (aged 5-12) prefer personified interfaces, although when a device knew the age and name of a child, it was taken as a violation of privacy.

4.3.7 CAs for Children with Differing Abilities. Two papers assessed the role of speech-based systems used by children with complex communication needs. Black et al. [16] developed and evaluated a system for children with complex communication needs (CCNs) to assist children in talking about their school day. The authors noted, "CCNs describe individuals who, due to motor, language, cognitive, and/or sensory perceptual impairments (e.g., as a result of cerebral palsy), have not developed speech and language skills as expected" [16, p. 2]. The system created a narrative using sensor data, voice recordings of teachers, and annotations from children. When the system worked well, it enabled the children to engage in storytelling, control the conversation, initiate topics, provide relevant information, and respond to their communication partner. Spitale et al. [122] explored whether CAs can support the linguistic assessment and training of children with a diagnosed language impairment. The authors compared the use of three forms of CAs – a physical toy, an avatar on a tablet, and a human on a tablet – by 6- to 8-year-old neuro-typical [122] children and children with a diagnosed language impairment. They found that while the linguistic performance of neuro-typical children was independent of the CA mode [122] and children with a diagnosed language impairment tended to perform better in the toy mode than in the other two modes.

Three papers looked into the interactions of children with autism using voice-based systems. Porayska-Pomsta et al. [98] designed and evaluated the educational efficacy of a technology called ECHOES, which opportunistically blends human and AI support for autistic children aged 4-14 years old. The ECHOES AI agent was designed as a responsive social partner, i.e., it is pro-active, reactive, and has social ability. The system also includes the capability to have a human practitioner as an interaction partner. In the study, in which each child played in the ECHOES environment for 10-20 minutes, several times a week over a six-week period, the number of initiations of joint-attentional behaviors increased with both the virtual character and the human practitioner. Also, while the children's responsiveness to the human practitioners increased, their responsiveness to the CA decreased. The authors observed that this was perhaps due to continuous reciprocal interactions that only human practitioners could facilitate and that the CA could not. Essentially, the CA was not aware that the child had responded unless the response was related to, and was expressed through, touch actions recognizable by ECHOES. Chaspari et al. [24] explored the link between Electrodermal Activity (EDA), a physiological signal indicative of a person's arousal, and the duration of verbal response latency, as children with autism interacted with an Embodied Conversational Agent named *Rachel*. They found that EDA patterns conveyed information about a child's inner state, since they differed according to the duration of verbal response latency with respect to Rachel's turns. In other words, EDA provides a complementary view of a child's state in cases where there may be no obvious (audible/visible) signs of arousal. The authors suggested this information could be used to dynamically personalize CAs that are targeted for children with autism. Mower et al. [89] found that while parents' interactions differ when interacting in parent-led or Rachel-led sessions, the differences were not seen in the interactions of children. From the data, the authors claim that if Rachel were used for diagnostic or interventional interaction, the communicative data from children with autism could be considered representative of their communication in a naturalistic setting.

4.3.8 Miscellaneous. Two papers could not be categorized under any of the other emerging themes. Andrist et al. [9] showed that flexibly employed verbal and nonverbal behaviors used by a CA (e.g., gestures, gazes and gestures, or gazes, proxemics, and gestures and speech), when passing the conversational turn in a group of children, can equalize participation, reduce overlapping speech, and improve speech recognition. Through a WoZ dialogue corpus collected from 103 children (aged 7-14 years) playing a voice-activated computer game, Yildirim et al. [151] investigated the automatic detection of "frustrated", "polite", and "neutral" attitudes in children's speech communication cues. They showed empirically that lexical information has more discriminative power than acoustic and contextual cues for the detection of politeness, whereas context and acoustic features serve best for frustration detection. However, a combination of the three cues provided better classification results overall. In their work, the accuracy of the classification also varied based on the age and gender of the children—"politeness" was detected with higher accuracy for females and 10- to 11-years-olds, compared to males and other age groups, respectively.

5 DISCUSSION

Our categorization of research topics revealed that HCI work on children and CAs developed around seven main themes over the past decade, focusing in large part on *ascribing human qualities to CAs, supporting children's learning, the use and role of CAs in the context of home and family, support for children's play, children's storytelling, issues around information collected and revealed by CAs, and CAs for children with differing abilities*. Based on our observations from the review and its findings, we derive key directions for future HCI work at the intersection of children and CAs.

5.1 Diversifying Participants for Culturally Sensitive CAs

Our review points to the fact that scholars are currently paying little attention to working with children with diverse backgrounds and identities. Except for a few studies (e.g., [15, 50]) that specifically targeted diverse families, based on race and socioeconomic status, our review found that CAs are often studied, designed, and developed only with samples of children who typically have infrastructure support (e.g., access to the Internet), are familiar with technology, and/or belong to specific sections of society (e.g., children who live in the Global North and belong to the middle or upper-middle class). Furthermore, only a small number of studies (e.g., [50, 104]) have taken into account aspects of children's intersectional identities (e.g., identities based both on race and socioeconomic status) while designing or trying to understand children's interaction with CAs.

As mentioned by Dell and Kumar [33], speech is of growing importance in the HCI4D (human-computer interaction for development) community. However, many families in the Global South (e.g., low-income families), as well as in the Global North (e.g., ethnic minorities), tend to engage less with voice-controlled smart devices. Although limited prior exposure to CAs is important, one of the key barriers to use also lies in the failure of CAs to support a user's native language, dialect, or accent [15, 50, 110]. This lack of "linguistic flexibility" in CAs has driven recent research on adult users who have to use a non-native language to interact with a CA [145]. In our review sample, one study also highlighted the difficulties of a Spanish-speaking family who felt they had to adapt their phonological processes to emulate English pronunciation when using the CAs embedded in smart speakers at home [15]. Multilingual adult users tend to strongly prefer chatbots (text-based CA) that can code-mix or code-switch [11] because multilingual adults often code-switch themselves while speaking, i.e., they alternate between two or more languages [20]. This code switching is also common in bilingual and multilingual children, influenced by their language proficiency, language preferences, and social identity [84]. Currently, the research on bilingual and multilingual children's CA use is scant. While several scholars have looked into the potential of using CAs to help children learn English (e.g., [53, 148]), a study by Rader et al. [104] is the only one to focus on children who a different dialect. They identified how CAs can help children who speak African American Vernacular English (AAVE) at home to code-switch between school-ratified English and AAVE in a school setting. Future avenues for research should focus on a wider range of linguistic capabilities and backgrounds, to ensure that a broader range of children's voices is heard in the CA design process.

Similar to considering linguistic diversity, ensuring cultural diversity is also critical when researching child-oriented CA interaction and design. The relationship between culture and technology, as well as how design can support cultural preferences and interactions (e.g., [49, 65, 120]) has been a topic of research within the HCI field. Within design, working across cultural differences is known as a way to seek stronger plurality [5]. Our review showed that current research is heavily skewed towards Western-based views, cultural norms, and communicative practices, as most of the studies were conducted with participants who identify with Western culture. Engaging with more culturally diverse participants in child-based CA research would enable researchers to develop a deeper and broader understanding of how CAs can be culturally tailored from a child's perspective. In this way, CAs can become aligned with the values and ideas that diverse children draw upon in their development. Researchers could also focus on the cultural dimensions of CAs' communication styles and the implications of using conventional social expectations or norms in the design of CAs. Current approaches and theories, such as Hofstede's *Dimensions Theory* [56] (often critiqued for offering a deterministic view [60]), transnational HCI (e.g. [76]), postcolonial computing (e.g., [146]), and Swidler's work on cultural sociology [65, 126] could be used as foundations to inform culture-based CA work.

Overall, CA researchers should aim to focus on the challenges faced by a wider and more diverse group of child users. This would enable scholars and designers to examine the degrees and forms of penalties and privileges that children experience in their lives lives [35, 116] and the impact these have on their interactions with CAs and their preferences in regard to them.

5.2 Considering Developmental Stages and Complex Communication Needs

Children's social, cognitive, and communication skills change as they grow from infancy to adolescence. For example, younger children tend to have shorter attention spans than older children [82] and toddlers syntactical development and vocabulary is more limited than the language skills of older children [12]. Furthermore, compared to other age groups, teenagers tend to have a heightened sensitivity to emotional and social scenarios and stronger responses to rewards [48].

Although it is clear that children vary significantly across developmental stages, our review suggests that relatively few CA research papers target infant and toddlers (0-3 years old) and older teenagers (15-17 years old). While limited research on infants and toddlers could be attributed to the limited conversational capacities of the children in these groups, there is nevertheless a need for the community to engage more widely across the age spectrum, particularly with older teens (as was also reported in a recent review of research on children and broader interaction design [67]). Increased engagement would allow for a more holistic view of children's experiences with CAs across developmental stages, leading to more detailed insights into designing CAs for these age groups.

Along with these considerations of variations in developmental stages, researchers also need to focus more explicitly on the diversity of children's needs. Our review shows a clear gap in this area. Only two of the reviewed papers explored children with complex communication needs (e.g., those with language impairments), with a further three papers examining interactions with CAs for children with autism (see Section 4.3.7). These papers exemplify the importance of understanding CA interactions with children who have differing abilities and, therefore, have specific needs and requirements, and of designing appropriate interactions for their needs. Work in this area can build on adult-based CA research by looking at adults with diverse capabilities and needs. These include non-sighted individuals (e.g. [1, 2]), people who stammer/stutter or those who have other diverse speech patterns [26], and people with motor disabilities [31]. While this research can act as a starting point, children's interactions cannot be assumed to be the same. Future work in this area should focus on other complex needs that specific cohorts of children may have and should investigate how these impact CA interactions and how CAs can be designed to be sensitive to such needs.

5.3 Increasing Design-Based Research

Our systematic review also revealed a lack of design-based research in the area of children's interactions with CAs. Only five studies focused on this topic, engaging with children in design through methods such as cooperative inquiry (e.g., [51]) and co-design [44]. While most of these design works included parents in the design sessions, there is a need to include other relevant stakeholders based on the context under consideration (e.g., teachers for learning-domain, therapist for the context of mental-health) Furthermore, rather than on practical, wide-ranging outcomes, these papers tended to focus on design recommendations and implications specific to the situation being researched, limiting their application to wider design practices [8]. In keeping with previous speech-based reviews in HCI [27], more widely applicable, design-based insights are needed to support the design of CAs for children. This could include the development of robust heuristics to support the design process, which are useful for facilitating the design and evaluation of products. Recent efforts have been made to generate heuristics for speech-based interfaces and conversational agents [70, 90, 138] that can aid designers in CA development and evaluation. However, these do not focus on specific needs and issues that may arise in the types of child-based interactions that are emphasized within the research covered by our review. Therefore, researchers should use the above mentioned heuristics as a starting point for developing heuristics to support the child-centric design of CAs. As the design of child-based speech devices and interactions becomes more commonplace, such heuristics and wider design research could significantly benefit the research and design community.

5.4 Further Engagement with Theory

Fourteen papers engaged with theories or theory-based concepts to hypothesize, to design a study or artifact, or to analyze data. Our review showed engagement with a number of theories/sets of concepts for understanding children's language choices in their use of CAs. We also observed that papers focusing on work with developmental communication patterns through the design of CAs are lacking in the literature. To this end, we see the need for scholars to continue engaging with literature from across disciplines on education, speech-language pathology, and children's social and communication development. Engaging with this literature would be valuable for designing interventions that support children's everyday needs and account for linguistic, developmental, and socio-emotional differences among children of different age groups.

Additionally, researchers and practitioners may need to consider theoretical approaches emerging from spoken interactions with CAs. Research has argued that our assumptions about communicative abilities (*partner models*) can be a key driver in how we interact with these devices, though debate on its significance continues [32]. Design choices around a system's voice and language can impact user's language production and perceptions of CA partners [28, 29]. Recent research has identified three core concepts that affect people's partner models of speech-based CAs: *perceptions of partner competence and dependability, assessment of CA human-likeness, and the perceived cognitive flexibility of systems* [38]. Our review especially highlighted the human-like attributes children of different ages may perceive in CAs [149] and how their expectations and interactions may be shaped by human-like features built into CA systems [4, 144]. Consequently, further research should examine how these concepts may map onto children's interactions with CAs at various stages of a child's development. Partner models could be developed in relation to complementary theories on CA interactions and child development.

5.5 Widening the Scope of Domains and User Contexts

Our research outlines a number of key topics that have been covered by studies of interactions with children. Many of these echo topics that have been widely covered in the research on the adult CA-user experience. For instance, the role of human likeness and the attribution of human qualities, while being researched in child-CA contexts, have also been considered in research on adult-based interactions with speech (e.g. [39, 91]) and text-based CAs [107]. Similarly, a focus on learning as a context of use [107] can be seen strongly in the child-based CA literature we reviewed. As seen in recent thematic reviews, in comparison with speech [27] and chatbots [107] research, there is a striking lack of overlap in many of the topics, concepts and clusters identified. Up to 2019, much of the research in the speech domain focused on researching system speech production, design-based insight, and modality comparison, and was limited to usability, theory, and specific system-based insights [27]. A recent review of chatbot research over the past ten years revealed an emphasis on the concepts of satisfaction, empathy, engagement, and trust, and highlighted a wide range of specific domains of use, such as mental health, art exhibitions, and e-commerce [107]. In comparison, in light of our review, it seems that CA research with children tends to primarily focus on learning-based interactions and the use of CAs within the family home; it is clear that there is a broad scope for introducing new domains, concepts, and user contexts to child-based CA research. For instance, focusing on contexts such as healthcare, mental health, online-safety, games, and entertainment would lead to significant insights that could inform the design and development of child-based applications for these domains, as well as the user experience requirements of child users. Recently, scholars have started to explore how text-based chatbots can help young adults (16- to 21-year-olds), by providing social support [10] and help children to develop emotional intelligence, including the ability to formulate alternate actions to address their own negative behavior [115]. In addition to widening contexts and domains, it is also important for future work to focus more on a wider range of concepts, such as engagement, children's perceptions of trust, privacy, and the usability of CAs. Finally, currently most of the work focuses on children's interactions with CAs embodied in smart speakers (as seen in Table 4), and therefore there is a need to increase efforts to further investigate voice-based interactions mediated through other devices, specifically mobile-phones, tablets, and toys, which are perhaps more extensively used by children.

5.6 Exploring Multi-Party Interactions

Another key aspect that we found across the literature was the notion of *multi-party* interactions, specifically involving children. The HCI literature highlights the multi-party nature of CA technologies, as they become "embedded" within home environments [101], and the topic of multi-party CA interactions with children appears in studies of the use and role of CAs in home and family contexts. Recording interactions in the home and interviewing families, Beneteau et al. [14] demonstrated how household members learned from each other how to use devices. Using similar methodology, Garg and Sengupta's work [50] showed that parents use of smart speakers change around children. Furthermore, Chen et al. [25] found that some parents preferred screen-based interactions to voice-based interactions, due to a multitude of factors, including a lack of control and intrusion into family life. Luria et al. [81] proposed that designers adopt a more nuanced consideration of family dynamics, shifting away from treating interactions as person-agnostic. Gordon and Breazeal [54] highlighted an approach by which a system is designed to use different modalities in working with both children and parents. In sum, although it is rarely the focus of attention, the topic of multi-party interactions remains prominent throughout a number of papers.

While the above literature adopts a somewhat broad focus on interaction with CAs, we saw the need also to consider specific interactional sequences and to inquire into how to avoid conflicts between users through design. Andrist et al. [9] indicate the importance of scaffolding their conversational turns to ensure fairness and avoid conflicts between users, when CAs are used in collaborative gaming. This raises important considerations for designers and makes the case for a deeper understanding of how to equip such systems to respond to interactional dynamics through their design. For example, researchers may want to explore how power dynamics between parents and children intersect with CA use, and how CAs enable such dynamics to play out, or not. Our review also covered related research showing disparities between genders in recognition accuracy [151], which, when one considers the role CAs might play in educating children (cf. [149]), precipitates the exigency of how to ensure that CA interactions are equitable and supportive of learners and how to assess and minimize the impacts of such biases in machine-learning systems [58]. Through a more generalizable approach to designing CAs, by which they can be adapted to the various multi-party dynamics they become embedded within, they could overcome some of the barriers to adoption and address concerns highlighted by parents across the board.

5.7 Investigating Ethical Challenges

Both within and outside the HCI research, there is evidence that suggests technology can shape children's behaviors and attitudes. For example, a body of work has examined whether social robots can impact adults' and children's social and conversational behaviors (e.g., [134, 140]). In regard to CAs, recent non-HCI researchers have developed CAs that can foster expressive language [23] and informal conversations [7]. In fact, some of the studies reviewed also showed that CAs can encourage children's verbal communication and language skills (e.g., [104, 122]). While this work should continue, there are also unanswered questions pertaining to the impact of CAs on the social-emotional development, linguistic skills, and attitudes of children that researchers should examine. These questions include: How do interactions with intelligent human-like CAs impact the development of children's social emotion and social cognition? Do children transfer linguistic patterns of conversing with CAs to interactions with people? What impact does conversing with CAs have on children's interpersonal relationships?

Finally, our literature review also found that children who interact with CAs may reveal personal information to CAs and form personal attachments to them [52]). Parents also have concerns regarding their children's interactions and attachment to CAs [52, 117]. Therefore, designers, researchers, and developers should strongly consider the ethical implications involved in fostering personal relationships between children, particularly young children, and technologies with which children may feel encouraged to share personal information. The concept of "stranger danger" [87] is a well-recognized concept taught to young children in some Western parts of the world (particularly the U.S.); it conveys the idea or warning that any stranger can be dangerous, and therefore children should avoid forming personal relationships with unfamiliar humans [87]. The designers of CAs could consider if "stranger danger" is an appropriate concept to incorporate into children's interactions with CAs, specifically for children who live in Western countries and identify with this concept. Prior work has reported that children deem technologies to be creepy based on their degree of deception and transparency, ominous visual appearance, inability to be controlled, mimicry, and unpredictability [152]. Therefore, future research can investigate how to make CAs less intrusive and more trustworthy and how these concepts are embedded in the cultural and societal landscapes of child users.

6 LIMITATIONS

The aim of this research was to review studies of children's interactions with voice-based CAs, published in core HCI venues. We focused on a comprehensive list of the top HCI publication venues, based on Google Scholar, Thomson Reuters, and Scimago rankings. While this process has been followed by prior reviews of speech and HCI work [27], this led us to exclude papers published in related non-core venues (e.g., ACM SIGCHI In-Cooperation conferences, like the Human-Agent Interaction (HAI) and Conversational User Interfaces (CUI) conferences). While future research may wish to consider venues outside the domain of HCI, this research, in line with previous reviews on CAs, conceives of HCI as a specific field.

7 CONCLUSIONS

This paper reports the results of a systematic review of 38 papers on child-based CA research, published in core HCI venues over the last decade (2011-2020) and thereby fills an important gap left by the only other major review on the state of speech-based research in HCI [27]. Our review calls attention to important research trends, such as CAs for supporting children with differing abilities, CAs supporting children's learning, using CAs in home and family settings, and understanding children's reasons for ascribing human qualities to CAs. The results of this review point to a number of short- and long-term opportunities for child-based HCI research, notably the needs to account for children's intersectional identities and engage with children who have differing needs and are linguistically and culturally diverse to inform the design of CAs, to increase design-based research, specifically to develop heuristics for child-centric design of CAs and involve relevant stakeholders (e.g., teachers, therapist) in the such research, to increase engagement with older teens and theories from various disciplines including education and speech-language pathology, to investigate ethical implications of CAs such as its impact on children's interpersonal relationship and social emotion and social cognition, to examine role and design of CAs for different domains and contexts such as mental health, play, online-safety, public spaces such as museums, and to examine the design and implications of multi-party interactions (e.g., power dynamics between parents and children) with CAs.

ACKNOWLEDGMENTS

We would like to thank Shreyas Sadagopan for helping with literature search on the ACM library. The first author would like to thank the SOURCE program at Syracuse University for partially funding this work.

REFERENCES

[1] Ali Abdolrahmani, Maya Howes Gupta, Mei-Lian Vader, Ravi Kuber, and Stacy Branham. 2021. Towards More Transactional Voice Assistants: Investigating the Potential for a Multimodal Voice-Activated Indoor Navigation Assistant for Blind and Sighted Travelers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 495, 16 pages. https://doi.org/10.1145/3411764.3445638

[2] Ali Abdolrahmani, Kevin M. Storer, Antony Rishin Mukkath Roy, Ravi Kuber, and Stacy M. Branham. 2020. Blind Leading the Sighted: Drawing Design Insights from Blind Users towards More Productivity-Oriented Voice Interfaces. *ACM Trans. Access. Comput.* 12, 4, Article 18 (Jan. 2020), 35 pages. https://doi.org/10.1145/3368426

[3] Deanne M Adams, Richard E Mayer, Andrew MacNamara, Alan Koenig, and Richard Wainess. 2012. Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of educational psychology* 104, 1 (2012), 235. https://doi.org/10.1037/a0025595

[4] Sara Aeschlimann, Marco Bleiker, Michael Wechner, and Anja Gampe. 2020. Communicative and social consequences of interactions with voice assistants. *Computers in Human Behavior* 112 (2020), 106466. https://doi.org/10.1016/j.chb.2020.106466

[5] Yoko Akama and Joyce Yee. 2016. Seeking stronger plurality: Intimacy and integrity in designing for social innovation. In *Cumulus Hong Kong 2016* (Hong Kong). Cumulus Association, Aalto, Finland, 12 pages.

[6] Mehdi Alaimi, Edith Law, Kevin Daniel Pantasdo, Pierre-Yves Oudeyer, and Hélène Sauzeon. 2020. Pedagogical Agents for Fostering Question-Asking Skills

in Children. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376776

[7] Mohammad Rafayet Ali, Seyedeh Zahra Razavi, Raina Langevin, Abdullah Al Mamun, Benjamin Kane, Reza Rawassizadeh, Lenhart K. Schubert, and Ehsan Hoque. 2020. A Virtual Conversational Agent for Teens with Autism Spectrum Disorder: Experimental Results and Design Lessons. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents* (Virtual Event, Scotland, UK) *(IVA '20)*. Association for Computing Machinery, New York, NY, USA, Article 2, 8 pages. https://doi.org/10.1145/3383652.3423900

[8] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. *Guidelines for Human-AI Interaction*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi-org/10.1145/3290605.3300233

[9] Sean Andrist, Iolanda Leite, and Jill Lehman. 2013. Fun and Fair: Influencing Turn-Taking in a Multi-Party Game with a Virtual Agent. In *Proceedings of the 12th International Conference on Interaction Design and Children* (New York, New York, USA) *(IDC '13)*. Association for Computing Machinery, New York, NY, USA, 352–355. https://doi.org/10.1145/2485760.2485800

[10] Petter Bae Bae Brandtzæg, Marita Skjuve, Kim Kristoffer Kristoffer Dysthe, and Asbjørn Følstad. 2021. When the Social Becomes Non-Human: Young People's Perception of Social Support in Chatbots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 257, 13 pages. https://doi.org/10.1145/3411764.3445318

[11] Anshul Bawa, Pranav Khadpe, Pratik Joshi, Kalika Bali, and Monojit Choudhury. 2020. Do Multilingual Users Prefer Chat-Bots That Code-Mix? Let's Nudge and Find Out! *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 041 (May 2020), 23 pages. https://doi.org/10.1145/3392846

[12] Helen Benedict. 1979. Early lexical development: comprehension and production. *Journal of Child Language* 6, 2 (1979), 183–200. https://doi.org/10.1017/S0305000900002245

[13] 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* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376344

[14] Erin Beneteau, Yini Guan, Olivia K. Richards, Mingrui Ray Zhang, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2020. Assumptions Checked: How Families Learn About and Use the Echo Dot. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 3 (March 2020), 23 pages. https://doi.org/10.1145/3380993

[15] Erin Beneteau, Olivia K. Richards, Mingrui Zhang, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication Breakdowns Between Families and Alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300473

[16] Rolf Black, Annalu Waller, Ross Turner, and Ehud Reiter. 2012. Supporting Personal Narrative for Children with Complex Communication Needs. *ACM Trans. Comput.-Hum. Interact.* 19, 2, Article 15 (July 2012), 35 pages. https://doi.org/10.1145/2240156.2240163

[17] Elena Bodrova and Deborah J Leong. 2006. *Tools of the mind*. Pearson Australia Pty Limited, Melbourne, Vic, Australia.

[18] Peter Börjesson, Wolmet Barendregt, Eva Eriksson, and Olof Torgersson. 2015. Designing Technology for and with Developmentally Diverse Children: A Systematic Literature Review. In *Proceedings of the 14th International Conference on Interaction Design and Children* (Boston, Massachusetts) *(IDC '15)*. Association for Computing Machinery, New York, NY, USA, 79–88. https://doi.org/10.1145/2771839.2771848

[19] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[20] Barbara E Bullock and Almeida Jacqueline Toribio. 2009. *Themes in the study of code-switching*. Cambridge University Press, Cambridge, UK.

[21] Harry Bunt. 2009. The DIT++ taxonomy for functional dialogue markup. In *Proceedings of 8th International Conference on Autonomous Agents and Multiagent Systems* (Budapest, Hungary) *(AAMAS 2009)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, USA, 13–24.

[22] Canadian Institutes of Health Research. 2006. *Randomized controlled trials registration/application checklist (12/2006)*. Technical Report. Government of Canada. http://www.cihr-irsc.gc.ca/e/documents/rct_reg_e.pdf

[23] Fabio Catania, Micol Spitale, Giulia Cosentino, and Franca Garzotto. 2020. Conversational Agents to Promote Children's Verbal Communication Skills. In *International Workshop on Chatbot Research and Design*. Springer, Cham, Switzerland, 158–172. https://doi.org/10.1007/978-3-030-68288-0_11

[24] Theodora Chaspari, Chi-Chun Lee, and Shrikanth Narayanan. 2012. Interplay between verbal response latency and physiology of children with autism during ECA interactions. In *Thirteenth Annual Conference of the International Speech Communication Association* (Portland, OR, USA) *(INTERSPEECH 2012)*. International Speech Communication Association, Grenoble, France, 1319–1322.

[25] Ying-Yu Chen, Ziyue Li, Daniela Rosner, and Alexis Hiniker. 2019. Understanding Parents' Perspectives on Mealtime Technology. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1, Article 5 (March 2019), 19 pages. https://doi.org/10.1145/3314392

[26] Leigh Clark, Benjamin R. Cowan, Abi Roper, Stephen Lindsay, and Owen Sheers. 2020. Speech Diversity and Speech Interfaces: Considering an Inclusive Future through Stammering. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) *(CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 24, 3 pages. https://doi.org/10.1145/3405755.3406139

[27] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, et al. 2019. The state of speech in HCI: Trends, themes and challenges. *Interacting with Computers* 31, 4 (2019), 349–371. https://doi.org/10.1093/iwc/iwz016

[28] Leigh Clark, Abdulmalik Ofemile, Svenja Adolphs, and Tom Rodden. 2016. A Multimodal Approach to Assessing User Experiences with Agent Helpers. *ACM Trans. Interact. Intell. Syst.* 6, 4, Article 29 (Nov. 2016), 31 pages. https://doi.org/10.1145/2983926

[29] Leigh Clark, Abdulmalik Ofemile, and Benjamin R Cowan. 2021. Exploring Verbal Uncanny Valley Effects with Vague Language in Computer Speech. In *Voice Attractiveness*. Springer, Singapore, 317–330. https://doi.org/10.1007/978-981-15-6627-1_17

[30] Lynn Schofield Clark. 2011. Parental mediation theory for the digital age. *Communication theory* 21, 4 (2011), 323–343. https://doi-org/10.1111/j.1468-2885.2011.01391.x

[31] Eric Corbett and Astrid Weber. 2016. What Can I Say? Addressing User Experience Challenges of a Mobile Voice User Interface for Accessibility. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 72–82. https://doi.org/10.1145/2935334.2935386

[32] Benjamin R Cowan, Holly P Branigan, Mateo Obregón, Enas Bugis, and Russell Beale. 2015. Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human- computer dialogue. *International Journal of Human-Computer Studies* 83 (2015), 27–42. https://doi.org/10.1016/j.ijhcs.2015.05.008

[33] Nicola Dell and Neha Kumar. 2016. The Ins and Outs of HCI for Development. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2220–2232. https://doi.org/10.1145/2858036.2858081

[34] Eric Deng, Bilge Mutlu, and Maja J Mataric. 2019. Embodiment in Socially Interactive Robots. *Foundations and Trends® in Robotics* 7, 4 (2019), 251–356. https://doi.org/10.1561/2300000056

[35] Rita Kaur Dhamoon. 2011. Considerations on mainstreaming intersectionality. *Political Research Quarterly* 64, 1 (2011), 230–243.

[36] Joao Dias and Ana Paiva. 2005. Feeling and reasoning: A computational model for emotional characters. In *Portuguese conference on artificial intelligence*. Springer, Berlin, Germany, 127–140. https://doi.org/10.1007/11595014_13

[37] Tawanna R. Dillahunt, Xinyi Wang, Earnest Wheeler, Hao Fei Cheng, Brent Hecht, and Haiyi Zhu. 2017. The Sharing Economy in Computing: A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 38 (Dec. 2017), 26 pages. https://doi.org/10.1145/3134673

[38] Philip R Doyle, Leigh Clark, and Benjamin R. Cowan. 2021. What Do We See in Them? Identifying Dimensions of Partner Models for Speech Interfaces Using a Psycholexical Approach. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 244, 14 pages. https://doi.org/10.1145/3411764.3445206

[39] Philip R. Doyle, Justin Edwards, Odile Dumbleton, Leigh Clark, and Benjamin R. Cowan. 2019. Mapping Perceptions of Humanness in Intelligent Personal Assistant Interaction. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) *(MobileHCI '19)*. Association for Computing Machinery, New York, NY, USA, Article 5, 12 pages. https://doi.org/10.1145/3338286.3340116

[40] 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* (Stanford, California, USA) *(IDC '17)*. Association for Computing Machinery, New York, NY, USA, 595–600. https://doi.org/10.1145/3078072.3084330

[41] Daniel A. Epstein, Clara Caldeira, Mayara Costa Figueiredo, Xi Lu, Lucas M. Silva, Lucretia Williams, Jong Ho Lee, Qingyang Li, Simran Ahuja, Qiuer Chen, Payam Dowlatyari, Craig Hilby, Sazeda Sultana, Elizabeth V. Eikey, and Yunan Chen. 2020. Mapping and Taking Stock of the Personal Informatics Literature. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 126 (Dec. 2020), 38 pages. https://doi.org/10.1145/3432231

[42] Julia Fink, Séverin Lemaignan, Pierre Dillenbourg, Philippe Rétornaz, Florian Vaussard, Alain Berthoud, Francesco Mondada, Florian Wille, and Karmen Franinović. 2014. Which Robot Behavior Can Motivate Children to Tidy up Their

Toys? Design and Evaluation of "Ranger". In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (Bielefeld, Germany) *(HRI '14)*. Association for Computing Machinery, New York, NY, USA, 439–446. https://doi.org/10.1145/2559636.2559659

[43] Joel E Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for Voice Interface Design. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (Dublin, Ireland) *(CUI '19)*. ACM, New York, NY, USA, Article 26, 8 pages. https://doi.org/10.1145/3342775.3342788

[44] Daniel Fitton, Janet C Read, Gavin Sim, and Brendan Cassidy. 2018. Co-Designing Voice User Interfaces with Teenagers in the Context of Smart Homes. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (Trondheim, Norway) *(IDC '18)*. Association for Computing Machinery, New York, NY, USA, 55–66. https://doi.org/10.1145/3202185.3202744

[45] Centers for Disease Control and Prevention. 2020. Child Development. Retrieved July 7, 2021 from https://www.cdc.gov/ncbddd/childdevelopment/positiveparenting/index.html

[46] Jon Froehlich, Leah Findlater, and James Landay. 2010. The Design of Eco-Feedback Technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1999–2008. https://doi.org/10.1145/1753326.1753629

[47] James J Gallagher and Mary Jane Aschner. 1963. A preliminary report on analyses of classroom interaction. *Merrill-Palmer Quarterly of Behavior and Development* 9, 3 (1963), 183–194.

[48] Adriana Galván. 2017. *The neuroscience of adolescence.* Cambridge University Press, Cambridge, UK.

[49] Ge Gao, Sun Young Hwang, Gabriel Culbertson, Susan R. Fussell, and Malte F. Jung. 2017. Beyond Information Content: The Effects of Culture on Affective Grounding in Instant Messaging Conversations. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 48 (Dec. 2017), 18 pages. https://doi.org/10.1145/3134683

[50] Radhika Garg and Subhasree Sengupta. 2019. "When You Can Do It, Why Can't I?": Racial and Socioeconomic Differences in Family Technology Use and Non-Use. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 63 (Nov. 2019), 22 pages. https://doi.org/10.1145/3359165

[51] Radhika Garg and Subhasree Sengupta. 2020. Conversational Technologies for In-Home Learning: Using Co-Design to Understand Children's and Parents' Perspectives. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376631

[52] Radhika Garg and Subhasree Sengupta. 2020. He Is Just Like Me: A Study of the Long-Term Use of Smart Speakers by Parents and Children. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 11 (March 2020), 24 pages. https://doi.org/10.1145/3381002

[53] David Antonio Gómez Jáuregui, Léonor Philip, Céline Clavel, Stéphane Padovani, Mahin Bailly, and Jean-Claude Martin. 2013. Video Analysis of Approach-Avoidance Behaviors of Teenagers Speaking with Virtual Agents. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (Sydney, Australia) *(ICMI '13)*. Association for Computing Machinery, New York, NY, USA, 189–196. https://doi.org/10.1145/2522848.2522891

[54] Michal Gordon and Cynthia Breazeal. 2015. Designing a Virtual Assistant for In-Car Child Entertainment. In *Proceedings of the 14th International Conference on Interaction Design and Children* (Boston, Massachusetts) *(IDC '15)*. Association for Computing Machinery, New York, NY, USA, 359–362. https://doi.org/10.1145/2771839.2771916

[55] Betty Hart and Todd R Risley. 1995. *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing, Baltimore, MD, USA.

[56] Geert Hofstede. 1983. National cultures in four dimensions: A research-based theory of cultural differences among nations. *International Studies of Management & Organization* 13, 1-2 (1983), 46–74.

[57] Geert Hofstede. 2011. Dimensionalizing cultures: The Hofstede model in context. *Online readings in psychology and culture* 2, 1 (2011), 2307–0919.

[58] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé, Miro Dudik, and Hanna Wallach. 2019. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3290605.3300830

[59] Inseok Hwang, Youngki Lee, Chungkuk Yoo, Chulhong Min, Dongsun Yim, and John Kim. 2019. Towards Interpersonal Assistants: Next-Generation Conversational Agents. *IEEE Pervasive Computing* 18, 2 (2019), 21–31. https://doi.org/10.1109/MPRV.2019.2922907

[60] Lilly Irani, Janet Vertesi, Paul Dourish, Kavita Philip, and Rebecca E. Grinter. 2010. *Postcolonial Computing: A Lens on Design and Development.* Association for Computing Machinery, New York, NY, USA, 1311–1320. https://doi-org/10.1145/1753326.1753522

[61] Alexis D. Jacq, Séverin Lemaignan, Fernando Garcia, Pierre Dillenbourg, and Ana Paiva. 2016. Building Successful Long Child-Robot Interactions in a Learning Context. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (Christchurch, New Zealand) *(HRI '16)*. IEEE Press, New York, NY, USA, 239–246.

[62] Kristiina Jokinen and Michael McTear. 2009. Spoken Dialogue Systems. *Synthesis Lectures on Human Language Technologies* 2, 1 (2009), 1–151.

[63] Peter H Kahn, Batya Friedman, Deanne R Perez-Granados, and Nathan G Freier. 2006. Robotic pets in the lives of preschool children. *Interaction Studies* 7, 3 (2006), 405–436.

[64] Peter H. Kahn Jr, Heather E. Gary, and Solace Shen. 2013. Children's Social Relationships With Current and Near-Future Robots. *Child Development Perspectives* 7, 1 (2013), 32–37. https://doi.org/10.1111/cdep.12011

[65] Naveena Karusala, Apoorva Bhalla, and Neha Kumar. 2019. Privacy, Patriarchy, and Participation on Social Media. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 511–526. https://doi.org/10.1145/3322276.3322355

[66] Saba Kawas, Ye Yuan, Akeiylah DeWitt, Qiao Jin, Susanne Kirchner, Abigail Bilger, Ethan Grantham, Julie A Kientz, Andrea Tartaro, and Svetlana Yarosh. 2020. Another Decade of IDC Research: Examining and Reflecting on Values and Ethics. In *Proceedings of the Interaction Design and Children Conference* (London, UK) *(IDC '20)*. Association for Computing Machinery, New York, NY, USA, 205–215. https://doi.org/10.1145/3392063.3394436

[67] Saba Kawas, Ye Yuan, Akeiylah DeWitt, Qiao Jin, Susanne Kirchner, Abigail Bilger, Ethan Grantham, Julie A Kientz, Andrea Tartaro, and Svetlana Yarosh. 2020. Another Decade of IDC Research: Examining and Reflecting on Values and Ethics. In *Proceedings of the Interaction Design and Children Conference* (London, UK) *(IDC '20)*. Association for Computing Machinery, New York, NY, USA, 205–215. https://doi.org/10.1145/3392063.3394436

[68] Minkyung Kim, Soonhyung Yi, and Donghun Lee. 2019. Between living and nonliving: Young children's animacy judgments and reasoning about humanoid robots. *PloS one* 14, 6 (2019), e0216869.

[69] Walter Kintsch. 1988. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological review* 95, 2 (1988), 163.

[70] Raina Langevin, Ross J Lordon, Thi Avrahami, Benjamin R. Cowan, Tad Hirsch, and Gary Hsieh. 2021. Heuristic Evaluation of Conversational Agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 632, 15 pages. https://doi.org/10.1145/3411764.3445312

[71] Richard S Lazarus. 1991. *Emotion and Adaptation.* Oxford University Press, New York, NY, USA.

[72] Gaël Lejeune, François Rioult, and Bruno Crémilleux. 2016. Highlighting Psychological Features for Predicting Child Interjections During Story Telling. In *Interspeech 2016* (San Francisco, CA, USA). International Speech Communication Association, Grenoble, France, 2056–2059. https://doi.org/10.21437/Interspeech.2016-527

[73] Oliver Lemon. 2012. Conversational interfaces. In *Data-Driven Methods for Adaptive Spoken Dialogue Systems.* Springer, New York, NY, USA, 1–4. https://doi.org/10.1007/978-1-4614-4803-7_1

[74] James C. Lester, Sharolyn A. Converse, Susan E. Kahler, S. Todd Barlow, Brian A. Stone, and Ravinder S. Bhogal. 1997. The Persona Effect: Affective Impact of Animated Pedagogical Agents. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '97)*. Association for Computing Machinery, New York, NY, USA, 359–366. https://doi.org/10.1145/258549.258797

[75] Alessandro Liberati, Douglas G Altman, Jennifer Tetzlaff, Cynthia Mulrow, Peter C Gøtzsche, John PA Ioannidis, Mike Clarke, Philip J Devereaux, Jos Kleijnen, and David Moher. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of clinical epidemiology* 62, 10 (2009), e1–e34. https://doi.org/10.1016/j.jclinepi.2009.06.006

[76] Silvia Lindtner, Ken Anderson, and Paul Dourish. 2012. Cultural Appropriation: Information Technologies as Sites of Transnational Imagination. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 77–86. https://doi.org/10.1145/2145204.2145220

[77] Gustavo Lopez and Luis A. Guerrero. 2017. Awareness Supporting Technologies Used in Collaborative Systems: A Systematic Literature Review. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 808–820. https://doi.org/10.1145/2998181.2998281

[78] Silvia Lovato and Anne Marie Piper. 2015. "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* (Boston, Massachusetts) *(IDC '15)*. Association for Computing Machinery, New York, NY, USA, 335–338. https://doi.org/10.1145/2771839.2771910

[79] Silvia B. Lovato, Anne Marie Piper, and Ellen A. Wartella. 2019. 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* (Boise, ID, USA) *(IDC '19)*. Association for Computing Machinery, New York, NY, USA, 301–313. https://doi.org/10.1145/3311927.3323150

[80] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[81] Michal Luria, Rebecca Zheng, Bennett Huffman, Shuangni Huang, John Zimmerman, and Jodi Forlizzi. 2020. Social Boundaries for Personal Agents in the Interpersonal Space of the Home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376311

[82] EM Mahone and HE Schneider. 2012. Assessment of attention in preschoolers. *Neuropsychology review* 22, 4 (2012), 361–383.

[83] Richard E Mayer, Kristina Sobko, and Patricia D Mautone. 2003. Social cues in multimedia learning: Role of speaker's voice. *Journal of educational Psychology* 95, 2 (2003), 419.

[84] Erica F McClure. 1977. *Aspects of code-switching in the discourse of bilingual Mexican-American children.* Technical Report. University of Illinois at Urbana-Champaign, Champaign, IL, USA.

[85] Emily McReynolds, Sarah Hubbard, Timothy Lau, Aditya Saraf, Maya Cakmak, and Franziska Roesner. 2017. Toys That Listen: A Study of Parents, Children, and Internet-Connected Toys. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 5197–5207. https://doi.org/10.1145/3025453.3025735

[86] Gail F Melson, Peter H Kahn Jr, Alan Beck, Batya Friedman, Trace Roberts, Erik Garrett, and Brian T Gill. 2009. Children's behavior toward and understanding of robotic and living dogs. *Journal of Applied Developmental Psychology* 30, 2 (2009), 92–102.

[87] Ellen Moran, David Warden, Lindsey Macleod, Gillian Mayes, and John Gillies. 1997. Stranger-danger: What do children know? *Child Abuse Review: Journal of the British Association for the Study and Prevention of Child Abuse and Neglect* 6, 1 (1997), 11–23.

[88] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100. https://doi.org/10.1109/MRA.2012.2192811

[89] Emily Mower, Chi-Chun Lee, James Gibson, Theodora Chaspari, Marian E Williams, and Shrikanth Narayanan. 2011. Analyzing the nature of ECA interactions in children with autism. In *Twelfth Annual Conference of the International Speech Communication Association (INTERSPEECH 2011)*. International Speech Communication Association, Grenoble, France, 2989–2992.

[90] Christine Murad, Cosmin Munteanu, Benjamin R. Cowan, and Leigh Clark. 2021. Finding a New Voice: Transitioning Designers from GUI to VUI Design. In *CUI 2021 - 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 22, 12 pages. https://doi.org/10.1145/3469595.3469617

[91] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of Social Issues* 56, 1 (2000), 81–103.

[92] Amy Ogan, Samantha Finkelstein, Elijah Mayfield, Claudia D'Adamo, Noboru Matsuda, and Justine Cassell. 2012. "Oh Dear Stacy!": Social Interaction, Elaboration, and Learning with Teachable Agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 39–48. https://doi.org/10.1145/2207676.2207684

[93] Andrew Ortony, Gerald L Clore, and Allan Collins. 1990. *The Cognitive Structure of Emotions.* Cambridge University Press, Cambridge, UK.

[94] Luiza Superti Pantoja, Kyle Diederich, Liam Crawford, and Juan Pablo Hourcade. 2019. Voice Agents Supporting High-Quality Social Play. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID, USA) *(IDC '19)*. Association for Computing Machinery, New York, NY, USA, 314–325. https://doi.org/10.1145/3311927.3323151

[95] Sunjeong Park and Youn-kyung Lim. 2020. Investigating User Expectations on the Roles of Family-Shared AI Speakers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376450

[96] Jean Piaget. 1976. Piaget's Theory. In *Piaget and His School: A Reader in Developmental Psychology*, Bärbel Inhelder, Harold H. Chipman, and Charles Zwingmann (Eds.). Springer, Berlin, Germany, 11–23. https://doi.org/10.1007/978-3-642-46323-5_2

[97] Luke Pittaway, Maxine Robertson, Kamal Munir, David Denyer, and Andy Neely. 2004. *Networking and Innovation in the UK: On the Systematic Literature Review.* Technical Report. Advanced Institute for Management Research.

[98] Kaundefinedka Porayska-Pomsta, Alyssa M. Alcorn, Katerina Avramides, Sandra Beale, Sara Bernardini, Mary Ellen Foster, Christopher Frauenberger, Judith Good, Karen Guldberg, Wendy Keay-Bright, Lila Kossyvaki, Oliver Lemon, Marilena Mademtzi, Rachel Menzies, Helen Pain, Gnanathusharan Rajendran, Annalu Waller, Sam Wass, and Tim J. Smith. 2018. Blending Human and Artificial Intelligence to Support Autistic Children's Social Communication Skills. *ACM Trans. Comput.-Hum. Interact.* 25, 6, Article 35 (Dec. 2018), 35 pages. https://doi.org/10.1145/3271484

[99] Martin Porcheron. 2021. What's in a Name and Does CUI Matter?. In *CUI 2021 - 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 24, 3 pages. https://doi.org/10.1145/3469595.3469619

[100] Martin Porcheron, Joel E Fischer, and Stuart Reeves. 2021. Pulling Back the Curtain on the Wizards of Oz. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 243 (jan 2021), 22 pages. https://doi.org/10.1145/3432942

[101] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 ACM Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 640, 12 pages. https://doi.org/10.1145/3173574.3174214

[102] Martin Porcheron, Joel E Fischer, and Sarah Sharples. 2017. "Do Animals Have Accents?": Talking with Agents in Multi-Party Conversation. In *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '17)*. ACM, New York, NY, USA, 207–219. https://doi.org/10.1145/2998181.2998298

[103] Barry M Prizant, Amy M Wetherby, Emily Rubin, and Amy C Laurent. 2003. The SCERTS model: A transactional, family-centered approach to enhancing communication and socioemotional abilities of children with autism spectrum disorder. *Infants & Young Children* 16, 4 (2003), 296–316.

[104] Emilee Rader, Margaret Echelbarger, and Justine Cassell. 2011. Brick by Brick: Iterating Interventions to Bridge the Achievement Gap with Virtual Peers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2971–2974. https://doi.org/10.1145/1978942.1979382

[105] Ashwin Ram. 1991. A theory of questions and question asking. *Journal of the Learning Sciences* 1, 3-4 (1991), 273–318. https://doi.org/10.1080/10508406.1991.9671973

[106] Amon Rapp, Lorenzo Curti, and Arianna Boldi. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies* 151 (2021), 102630. https://doi.org/10.1016/j.ijhcs.2021.102630

[107] Amon Rapp, Lorenzo Curti, and Arianna Boldi. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies* 151 (2021), 102630. https://doi.org/10.1016/j.ijhcs.2021.102630

[108] Karen Renaud and Judy van Biljon. 2008. Predicting Technology Acceptance and Adoption by the Elderly: A Qualitative Study. In *Proceedings of the 2008 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries: Riding the Wave of Technology* (Wilderness, South Africa) *(SAICSIT '08)*. Association for Computing Machinery, New York, NY, USA, 210–219. https://doi.org/10.1145/1456659.1456684

[109] Minjin Rheu, Ji Youn Shin, Wei Peng, and Jina Huh-Yoo. 2021. Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human–Computer Interaction* 37, 1 (2021), 81–96. https://doi.org/10.1080/10447318.2020.1807710

[110] Simon Robinson, Jennifer Pearson, Shashank Ahire, Rini Ahirwar, Bhakti Bhikne, Nimish Maravi, and Matt Jones. 2018. Revisiting "Hole in the Wall" Computing: Private Smart Speakers and Public Slum Settings. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3173574.3174072

[111] Everett M Rogers. 2010. *Diffusion of innovations.* Simon and Schuster, New York, NY, USA.

[112] Daniel Rough and Benjamin Cowan. 2020. Don't Believe The Hype! White Lies of Conversational User Interface Creation Tools. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) *(CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 17, 3 pages. https://doi.org/10.1145/3405755.3406140

[113] Sherry Ruan, Jiayu He, Rui Ying, Jonathan Burkle, Dunia Hakim, Anna Wang, Yufeng Yin, Lily Zhou, Qianyao Xu, Abdallah AbuHashem, Griffin Dietz, Elizabeth L. Murnane, Emma Brunskill, and James A. Landay. 2020. Supporting Children's Math Learning with Feedback-Augmented Narrative Technology. In *Proceedings of the Interaction Design and Children Conference* (London, UK) *(IDC '20)*. Association for Computing Machinery, New York, NY, USA, 567–580. https://doi.org/10.1145/3392063.3394400

[114] Stuart Russell and Peter Norvig. 2002. *Artificial Intelligence: A Modern Approach* (2 ed.). Prentice Hall, Hoboken, NJ, USA.

[115] Kyle-Althea Santos, Ethel Ong, and Ron Resurreccion. 2020. Therapist Vibe: Children's Expressions of Their Emotions through Storytelling with a Chatbot. In *Proceedings of the Interaction Design and Children Conference* (London, UK) *(IDC '20)*. Association for Computing Machinery, New York, NY, USA, 483–494. https://doi.org/10.1145/3392063.3394405

[116] Ari Schlesinger, W. Keith Edwards, and Rebecca E. Grinter. 2017. Intersectional HCI: Engaging Identity through Gender, Race, and Class. In *Proceedings of the*

2017 CHI Conference on Human Factors in Computing Systems (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5412–5427. https://doi.org/10.1145/3025453.3025766

[117] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 857–868. https://doi.org/10.1145/3196709.3196772

[118] Subhasree Sengupta and Radhika Garg. 2019. Impact of Voice-based Interaction on Learning Practices and Behavior of Children.. In *IUI Workshops*. CEUR Workshops, Aachen, Germany, 3 pages.

[119] Ovidiu Serban, Mukesh Barange, Sahba Zojaji, Alexandre Pauchet, Adeline Richard, and Emilie Chanoni. 2017. Interactive Narration with a Child: Impact of Prosody and Facial Expressions. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (Glasgow, UK) *(ICMI '17)*. Association for Computing Machinery, New York, NY, USA, 23–31. https://doi.org/10.1145/3136755.3136797

[120] Yedendra B. Shrinivasan, Mohit Jain, Deva P. Seetharam, Abhishek Choudhary, Elaine M. Huang, Tawanna Dillahunt, and Jennifer Mankoff. 2013. *Deep Conservation in Urban India and Its Implications for the Design of Conservation Technologies*. Association for Computing Machinery, New York, NY, USA, 1969–1978. https://doi.org/10.1145/2470654.2466261

[121] Catherine Snow. 1983. Literacy and language: Relationships during the preschool years. *Harvard educational review* 53, 2 (1983), 165–189.

[122] Micol Spitale, Silvia Silleresi, Giulia Cosentino, Francesca Panzeri, and Franca Garzotto. 2020. "Whom Would You like to Talk with?": Exploring Conversational Agents for Children's Linguistic Assessment. In *Proceedings of the Interaction Design and Children Conference* (London, UK) *(IDC '20)*. Association for Computing Machinery, New York, NY, USA, 262–272. https://doi.org/10.1145/3392063.3394421

[123] Statista.com. 2020. Smart Speaker Ownership Worldwide by Country. Retrieved June 7, 2021 from https://www.statista.com/statistics/1130208/smart-speaker-ownership-worldwide-by-country

[124] Reed Stevens and Lori Takeuchi. 2011. *The new coviewing: Designing for learning through joint media engagement.* Technical Report. The Joan Ganz Cooney Center.

[125] Kevin M. Storer, Tejinder K. Judge, and Stacy M. Branham. 2020. "All in the Same Boat": Tradeoffs of Voice Assistant Ownership for Mixed-Visual-Ability Families. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376225

[126] Ann Swidler. 1986. Culture in action: Symbols and strategies. *American sociological review* 51, 2 (apr 1986), 273–286.

[127] Anuj Tewari and John Canny. 2014. What Did Spot Hide? A Question-Answering Game for Preschool Children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1807–1816. https://doi.org/10.1145/2556288.2557205

[128] Pablo Vicente Torres-Carrión, Carina Soledad González-González, Silvana Aciar, and Germania Rodríguez-Morales. 2018. Methodology for systematic literature review applied to engineering and education. In *2018 IEEE Global engineering education conference (EDUCON)*. IEEE, New York, NY, USA, 1364–1373. https://doi.org/10.1109/EDUCON.2018.8363388

[129] Daniel C. Tozadore and Roseli A. F. Romero. 2020. Multimodal Fuzzy Assessment for Robot Behavioral Adaptation in Educational Children-Robot Interaction. In *Companion Publication of the 2020 International Conference on Multimodal Interaction* (Virtual Event, Netherlands) *(ICMI '20 Companion)*. Association for Computing Machinery, New York, NY, USA, 392–399. https://doi.org/10.1145/3395035.3425201

[130] Esko Ukkonen. 2009. Maximal and minimal representations of gapped and non-gapped motifs of a string. *Theoretical Computer Science* 410, 43 (2009), 4341–4349.

[131] Patti M Valkenburg, Marina Krcmar, Allerd L Peeters, and Nies M Marseille. 1999. Developing a scale to assess three styles of television mediation:"Instructive mediation,""restrictive mediation," and "social coviewing". *Journal of broadcasting & electronic media* 43, 1 (1999), 52–66.

[132] Mike van Duuren and Michael Scaife. 1996. "Because a robot's brain hasn't got a brain, it just controls itself"—Children's attributions of brain related behaviour to intelligent artefacts. *European Journal of Psychology of Education* 11, 4 (1996), 365–376.

[133] Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS quarterly* 27, 3 (sept 2003), 425–478. https://doi.org/10.2307/30036540

[134] Michael Villano, Charles R. Crowell, Kristin Wier, Karen Tang, Brynn Thomas, Nicole Shea, Lauren M. Schmitt, and Joshua J. Diehl. 2011. DOMER: A Wizard of Oz Interface for Using Interactive Robots to Scaffold Social Skills for Children with Autism Spectrum Disorders. In *Proceedings of the 6th International Conference on Human-Robot Interaction* (Lausanne, Switzerland) *(HRI '11)*. Association for Computing Machinery, New York, NY, USA, 279–280. https://doi.org/10.1145/1957656.1957770

[135] Voicebot.AI. 2020. U.S. Smart Speaker Growth Flat Lined in 2020. Retrieved June 7, 2021 from https://voicebot.ai/2021/04/14/u-s-smart-speaker-growth-flat-lined-in-2020/

[136] Lev Semenovich Vygotsky. 1980. *Mind in society: The development of higher psychological processes.* Harvard University Press, Cambridge, MA, USA.

[137] Wayne Ward, Ronald Cole, Daniel Bolaños, Cindy Buchenroth-Martin, Edward Svirsky, Sarel Van Vuuren, Timothy Weston, Jing Zheng, and Lee Becker. 2011. My Science Tutor: A Conversational Multimedia Virtual Tutor for Elementary School Science. *ACM Trans. Speech Lang. Process.* 7, 4, Article 18 (Aug. 2011), 29 pages. https://doi.org/10.1145/1998384.1998392

[138] Zhuxiaona Wei and James A. Landay. 2018. Evaluating Speech-Based Smart Devices Using New Usability Heuristics. *IEEE Pervasive Computing* 17, 2 (2018), 84–96. https://doi.org/10.1109/MPRV.2018.022511249

[139] Joseph Weizenbaum. 1983. ELIZA — a Computer Program for the Study of Natural Language Communication between Man and Machine. *Commun. ACM* 26, 1 (Jan. 1983), 23–28. https://doi.org/10.1145/357980.357991

[140] Iain Werry, Kerstin Dautenhahn, Bernard Ogden, and William Harwin. 2001. Can Social Interaction Skills Be Taught by a Social Agent? The Role of a Robotic Mediator in Autism Therapy. In *International Conference on Cognitive Technology (CT 2001)*. Springer, Berlin, Germany, 57–74. https://doi.org/10.1007/3-540-44617-6_6

[141] Jacqueline M. Kory Westlund, Hae Won Park, Randi Williams, and Cynthia Breazeal. 2018. Measuring Young Children's Long-Term Relationships with Social Robots. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (Trondheim, Norway) *(IDC '18)*. Association for Computing Machinery, New York, NY, USA, 207–218. https://doi.org/10.1145/3202185.3202732

[142] James Wilson and Daniel Rosenberg. 1988. Rapid prototyping for user interface design. In *Handbook of human-computer interaction*. Elsevier, Amsterdam, Netherlands, Chapter 39, 859–875. https://doi.org/10.1016/B978-0-444-70536-5.50044-0

[143] Terry Winograd, Fernando Flores, and Fernando F Flores. 1986. *Understanding Computers and Cognition: A New Foundation for Design*. Addison Wesley, Reading, MA, USA.

[144] Julia Woodward, Zari McFadden, Nicole Shiver, Amir Ben-hayon, Jason C. Yip, and Lisa Anthony. 2018. Using Co-Design to Examine How Children Conceptualize Intelligent Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174149

[145] 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* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 34, 9 pages. https://doi.org/10.1145/3379503.3403563

[146] Susan Wyche, Tawanna R. Dillahunt, Nightingale Simiyu, and Sharon Alaka. 2015. "If God Gives Me the Chance i Will Design My Own Phone": Exploring Mobile Phone Repair and Postcolonial Approaches to Design in Rural Kenya. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 463–473. https://doi.org/10.1145/2750858.2804249

[147] Ying Xu and Mark Warschauer. 2020. A Content Analysis of Voice-Based Apps on the Market for Early Literacy Development. In *Proceedings of the Interaction Design and Children Conference* (London, UK) *(IDC '20)*. Association for Computing Machinery, New York, NY, USA, 361–371. https://doi.org/10.1145/3392063.3394418

[148] 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* (London, UK) *(IDC '20)*. Association for Computing Machinery, New York, NY, USA, 216–228. https://doi.org/10.1145/3392063.3394417

[149] 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*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376416

[150] Svetlana Yarosh, Stryker Thompson, Kathleen Watson, Alice Chase, Ashwin Senthilkumar, Ye Yuan, and A. J. Bernheim Brush. 2018. Children Asking Questions: Speech Interface Reformulations and Personification Preferences. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (Trondheim, Norway) *(IDC '18)*. Association for Computing Machinery, New York, NY, USA, 300–312. https://doi.org/10.1145/3202185.3202207

[151] Serdar Yildirim, Shrikanth Narayanan, and Alexandros Potamianos. 2011. Detecting emotional state of a child in a conversational computer game. *Computer Speech & Language* 25, 1 (2011), 29–44. https://doi.org/10.1016/j.csl.2009.12.004

Affective Speech in Real-Life Interactions.

[152] Jason C. Yip, Kiley Sobel, Xin Gao, Allison Marie Hishikawa, Alexis Lim, Laura Meng, Romaine Flor Ofiana, Justin Park, and Alexis Hiniker. 2019. *Laughing is Scary, but Farting is Cute: A Conceptual Model of Children's Perspectives of Creepy Technologies*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300303

[153] Charles Young and Richard Horton. 2005. Putting clinical trials into context. *The Lancet* 366, 9480 (2005), 107–108. https://doi.org/10.1016/S0140-6736(05)67370-9

[154] Ye Yuan, Stryker Thompson, Kathleen Watson, Alice Chase, Ashwin Senthilkumar, AJ Bernheim Brush, and Svetlana Yarosh. 2019. Speech interface reformulations and voice assistant personification preferences of children and parents. *International Journal of Child-Computer Interaction* 21 (2019), 77–88. https://doi.org/10.1016/j.ijcci.2019.04.005

[155] Bieke Zaman, Marije Nouwen, Jeroen Vanattenhoven, Evelien De Ferrerre, and Jan Van Looy. 2016. A qualitative inquiry into the contextualized parental mediation practices of young children's digital media use at home. *Journal of Broadcasting & Electronic Media* 60, 1 (2016), 1–22. https://doi-org/10.1080/08838151.2015.1127240