Context-responsive ASL Recommendation for Parent-Child Interaction

Ekram Hossain
ehossai2@ur.rochester.edu
University of Rochester,
USA

Merritt Cahoon
mcahoon1@samford.edu
Samford University,
USA

Yao Liu
yliu204@u.rochester.edu
University of Rochester,
USA

Chigusa Kurumada
ckuruma2@ur.rochester.edu
University of Rochester,
USA

Zhen Bai
zbai@cs.rochester.edu
University of Rochester,
USA

ABSTRACT

Parental language input in early childhood plays a critical role in lifelong neuro-cognitive and social development. Deaf and Hard of Hearing (DHH) children are often at risk of language deprivation due to hearing parents' limited knowledge of sign language - the natural language for DHH children at birth. To offer an immersive sign language environment for DHH children, we designed a novel computer-mediated communication technology named Table Top Interactive System (TIPS). It aims to provide context-responsive recommendation of American Sign Language (ASL) in real-time for hearing parents during face-to-face joint play with their DHH children. The system emphasizes supporting parent autonomy by adapting ASL recommendations using parent's speech during play, and minimizes obtrusion for face-to-face interaction through an Augmented Reality (AR) display. This paper describes the design and development of an initial working prototype of TIPS and preliminary results of the system's efficiency regarding system latency and accuracy for ASL recommendation and visualization. Next, we plan to conduct a user study to gather expert and parent feedback about the system design and ASL recommendation strategies for long-term and personalized usage.

CCS CONCEPTS

• **Human-centered computing** → **Human-centered computing**; **Accessibility**.

KEYWORDS

Computer-mediated communication, American Sign Language, Parent-Child Interaction, Augmented Reality

ACM Reference Format:
Ekram Hossain, Merritt Cahoon, Yao Liu, Chigusa Kurumada, and Zhen Bai. 2022. Context-responsive ASL Recommendation for Parent-Child Interaction. In *The 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22), October 23–26, 2022, Athens, Greece.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3517428.3550366

ASSETS '22, October 23–26, 2022, Athens, Greece

ACM ISBN 978-1-4503-9258-7/22/10.
https://doi.org/10.1145/3517428.3550366

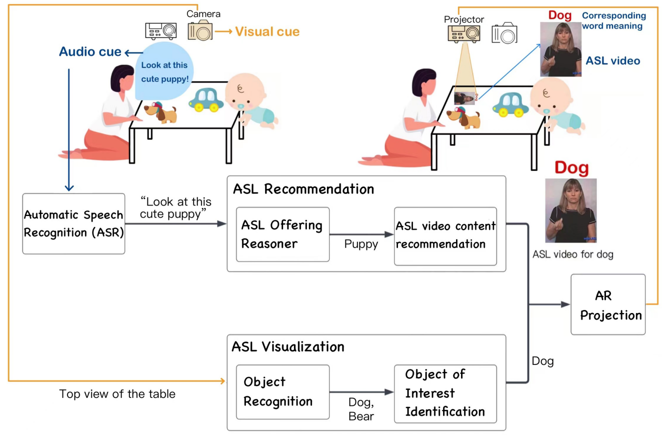
1 INTRODUCTION AND RELATED WORK

More than 90% of DHH children are born to hearing parents [31]. A lack of an immersive sign language environment at home may pose the risk of language deprivation for DHH children [13] [12] [5], leading to long-term effects on a child's ability to acquire a first language [27], cognitive delays [21], and mental health problems [14]. Therefore, there is an increasing emphasis in pediatrics research on parent and family support for sign language during the early stages of child development [18] [17].

Existing research mainly focuses on supporting sign language acquisition for hearing parents and DHH children (e.g., games [3], mobile applications [43], storybooks [25], and avatars [37] [39] [36] [1]). There remains a critical research gap of technologies that support sign language exposure for DHH children through real-time interaction with hearing parents and caregivers [43]. Emerging research on computer-mediated communication sheds light on using Natural Language Processing (NLP) technologies in monitoring parent-child conversation, and providing parents with feedback on communication strategies such as turn taking and praise when interacting with children with language and behavioral difficulties ([19] [16]). A recent system was developed to provide situation-related phrases for immigrant parents during play [22]. This system guides the parent with pre-defined phrases relevant to toys of attention, however, it may hinder parents' autonomy and fluency of play when the recommended phrase lacks correspondence to the momentary play episode. For example, the system may recommend the phrase "car moves" because it's relevant to a car toy, but if the parent says "the car is fast," then the parent may have to yield the play idea to the recommended phrase. Furthermore, visual feedback of communication strategies and language inputs in these systems are often displayed on mobile or tablet devices, which may direct away the parent's attention and impede the fluency of face-to-face interaction [45][40] [19].

To address these challenges, we designed a novel computer-mediated communication technology named TIPS (Tabletop Interactive Play System) (Figure 1). It aims to provide context-responsive recommendation of American Sign Language (ASL) in real-time for hearing parents during face-to-face joint play with their DHH children. The system emphasizes on (1) supporting parent autonomy by recommending ASL in alignment with the semantic contents of the parent's speech during play, and (2) minimizing obtrusion for face-to-face interaction through Augmented Reality (AR) display, which

(a) An illustration of the TIPS setup - the parent (left) and child (right) sit across the table. TIPS monitors the parent's speaking, and projects context-appropriate ASL near the object of interest.



(b) Overview of system pipeline for TIPS system

Figure 1: System Overview of TIPS: Tabletop Interactive Play System

is promising in eliminating attention diversion in face-to-face interaction ([20] [32]). TIPS adopts a projection-based AR display that visualizes the recommended ASL video near the object of interest in a tabletop setting. A recent study demonstrates that the design concept of projection-based AR display is most preferred for ASL visualization during face-to-face interaction as compared to tablets, smart watches, and smart glasses, due to its high glanceability, unobtrusiveness, and visibility of facial gestures (an integral aspect of ASL) [2].

This paper describes the design and development of an initial working prototype of TIPS, which includes two main modules: **vocabulary-based ASL recommendation** and **ASL AR visualization**. A small scale system performance evaluation shows satisfactory efficiency regarding system latency (on average 0.28 seconds) and visualization with a 75% accuracy rate for detecting the correct toy of interest. Next, we plan to conduct a user study to gather expert and parent feedback about the system design and ASL recommendation strategies for long-term and personalized usage.

2 SYSTEM DESIGN

The goal of the **ASL Recommendation Module** (ARM) is to analyze the parents' utterances and provide context-responsive ASL content directly derived from the utterances. ARM consists of three components: **Automatic Speech Recognition**, **ASL Recommendation Reasoner**, and **ASL Video Retrieval**. ARM captures the verbal cues using an ASR system which transcribes the speech into text. Then the "ASL recommendation reasoner" conducts semantic analysis of the text, and extracts the appropriate vocabulary based on pre-defined ASL recommendation strategies.

The current ASL recommendation reasoner focuses on vocabulary-based language input over phrases and sentences. According to usage-based theory, children learn low-scope structures based on individual words or morphemes and gradually develop more complex and abstract linguistic representations [23]. The reasoner prioritizes content words (nouns, main verbs, adjectives, and adverbs) over function words (e.g., that, the). Research shows that content words are more helpful to promote children to engage during interactions [23], and there is a high percentage of content words in young children's expressive vocabulary [10]. The current reasoner serves as a technology probe to allow co-design with key stakeholders about advanced ASL recommendation strategies, and personalized approach to adopt to individual's ASL communication needs (see the "future user study plan" section).

The goal of the **ASL AR Visualization Module** is to dynamically project the recommended ASL video next to the toy of interest. The **"Object of interest identification module"** (TIIM) captures both verbal cues and real-time visual information (top view of tabletop) to determine where to project the ASL content. ARM provides the current context, and TIIM identifies the most relevant toy associated with the parent's utterance. Then, the AR projection module projects the video next to the toy on the table. According to dual-coding learning theory, linking physical objects with appropriate semantic information may improve language acquisition [28].

3 SYSTEM DEVELOPMENT

3.1 ASL Recommendation Module (ARM)

3.1.1 Automatic Speech Recognition Framework. We used the state-of-the-art offline speech recognition framework Wav2vec2 [46] to accomplish the ASR task. This framework is being used in different real-time context-aware application such as valence-arousal estimation [30], mispronunciation detection [34], emotion estimation [29], autism classification [6], and sentiment analysis [26]. The Word Error Rate (WER) of this framework is 1.8/3.3 (in%) [46] on the noisy/clean test sets of Librispeech [33] which is better than other existing ASR system such as Discrete-BERT [15], ContextNet [15], Conformer [11]. Considering the performance in terms of WER and it's practical usage in different context-aware application, we choose this framework for ASR.

3.1.2 ASL Recommendation Reasoner. To suggest vocabulary from the transcribed speech, we used Stanford Dependency Parser [35] to extract content words. To remove function/stop words, we used the NLTK package [4].

3.1.3 ASL Video Retrieval. The ASL Video Retrieval System searches for the selected word's exact or the semantically relevant ASL video content. We used BERT embedding [7] for semantic search because such embeddings are dynamically calculated according to sentence-level context [42]. The difference between static and dynamic embedding is that dynamic embedding can capture the sense [44] of a word in a sentence. For example, the word "fly" can be used in two different senses, as a "bug" or as "moving through the air." BERT will compute different embeddings for "fly" based on the contextual senses in a sentence. For semantic vocabulary search, first, we pre-calculated the embeddings of each word in our ASL dictionary. The algorithm uses cosine similarity to provide the exact or the most relevant ASL video content against the selected word. This design will support parents' diverse semantically relevant linguistic input. For example, a parent addresses a "dog" toy as "puppy" or "doggy." In both cases, our systems provide ASL for "dog."

3.2 ASL Video Databases:

To build an extensive and comprehensive ASL video corpus, we used the open-sourced ASL video data from three university ASL databases: the University of Rochester, Rochester Institute of Technology, and Gallaudet University. In total, we have 6472 number of ASL videos.

3.3 ASL AR Visualization Module

A camera and a projector are requirements to implement projection-based augmented reality 1b, and projector-camera calibration is recommended for the best outcome of AR projection. However, for this study, we only did a camera calibration following the suggestion of system implementation in PATI [9]. To get the exact location of the toy on a table from the top view, the camera-captured image should be transformed into the - "birds-eye-view" image. To calculate this, first we need to apply perspective transformation on the pre-selected four points to calculate the perspective matrix M. Using this matrix, we applied warp transformations on the camera-captured image to obtain a "birds-eye-view" image of the table surface. We chose a projector (Optoma ML1050ST+) with a short throw ratio for better projection because our projector is situated around 3 feet above the table. The other factors in choosing this projector were size, weight, screen width, and brightness.

3.4 The "Object of interest identification module" (TIIM)

To project ASL video content near the toy, the CV module detects the current toy of interest through a multimodal approach based on the verbal cue (parent's input) and visual input (top view image of tabletop). First, to locate the current toys on the table, we choose the YOLO object detection model [24] for its robustness for real-time performance and detecting small objects. We fine-tuned the existing model with 13 toys. The next task is to localize which toy is the most relevantly referred to. We use BERT to calculate the word embedding of the selected word extracted by the ASL offering reasoner and the current-table-top detected toys using the labeled class name. Then, the cosine similarity is calculated to find the most relevantly referred toy against the selected word from the parent's utterance. After the AR module projects the video follows the current toy of interest.

4 PRELIMINARY EVALUATION

We conducted an internal preliminary evaluation with two researchers on the team (both are native English speakers) to evaluate the system's efficiency. We placed five toys on the table first one at a time, then two toys at a time (bus, firetruck, red bus, policeman, girl doll), and asked each speaker to play with these toys for a total of 10 minutes. The pairings of toys were the bus with the firetruck, policeman with firetruck, and girl doll with bus. We limited the amount of instructions given to the evaluators as to not limit their toy play and sentence generation. The only instructions were to not use complex sentences or questions, and that the sentences should either be about only one toy or two toys interacting. We also let the speakers know that the sentences do not have to be directly about the object or mention the object, and that they can use the toy to represent other objects of interest. The total number of sentences collected during the evaluation was 41. The average length of these sentence was 5.225. We calculated the latency of the system using the sentences uttered by the speakers. The system's overall latency is composed of the duration of speech to text, the time that it takes for the ASL recommendation algorithm to select the appropriate word, the retrieval of the corresponding ASL video content from the ASL video dictionary, and the time to visualize ASL video through the projector. ASL recommendation latency refers to the time the system takes to generate the POS (parts of speech) tags from a transcribed sentence, and to suggest an appropriate word from the extracted POS tags. Video retrieval latency refers to the time the system takes to perform the semantic search over the ASL dictionary to retrieve the relevant ASL video content for the selected word. The amount of time from the retrieval of the ASL video content to projector visualization is negligible. That is why we did not consider this latency to calculate the overall system's latency. The average display latency between the finish of the sentence and the appearance of the ASL video was 0.28 seconds (*SD = 0.084*). Based of off [38], [8], and [41] our system qualifies as being instantaneous due to being under 300 milliseconds. Our system also showed 75% accuracy rate of identification of toy of interest. TIPS captures the audio and video stream simultaneously. If the "Object of interest identification module" (TIIM) displays the retrieved ASL video content that matches with the intended object that the parent refers to, we consider that our system successfully identified the toy of interest. We calculated this by dividing the total number of successful identification of toy of interest by the total number of identification attempts.

5 FUTURE USER STUDY PLAN

We are currently working on improving the system's performance. We plan to conduct a larger scale usability study to evaluate the efficiency of the improved system in terms of display latency, accuracy rate of identification of toy of interest, and precision of ASL video projection. We will then conduct two consecutive user studies. The first study is a co-design study with key stakeholders (language acquisition researcher, ASL educator, early childhood educator and hearing parents with DHH children, novice ASL learners, and DHH individuals). This study aims to gather feedback on the current system and brainstorm ASL recommendation strategies for long-term and personalized usage. The second study will be with hearing parents and DHH children. This study aims to evaluate the effectiveness of the working prototype in supporting hearing parents to deliver context-responsive ASL on the fly during joint play with DHH children.

6 CONCLUSION

In this research, we developed a working prototype of the TIPS system to improve sign language exposure for DHH children, through recommending hearing parents context-responsive ASL during joint play. TIPS adopts a speech-driven approach to support the autonomy of the parent by selecting a video of the ASL they wish to sign in line with an ongoing play episode, and by implementing the projection-based AR display to minimize obtrusion during face-to-face interactions. Results of a small scale usability evaluation demonstrate satisfactory system efficiency. We discussed the plan for future studies to iterate the system development and conduct quantitative evaluation of the effectiveness of TIPS in augmenting hearing parents' communication in ASL with their DHH children.

ACKNOWLEDGMENTS

We would like to thank Google Inclusive Research Award for supporting this research. We also thank the NSF REU program for paving the way for collaboration for this project.

REFERENCES

[1] Sedeeq Al-Khazraji, Larwan Berke, Sushant Kafle, Peter Yeung, and Matt Huenerfauth. 2018. Modeling the speed and timing of American Sign Language to generate realistic animations. In *Proceedings of the 20th international ACM SIGACCESS conference on computers and accessibility*. 259–270.

[2] ZHEN BAI, ELIZABETH M CODICK, ASHELY TENESACA, WANYIN HU, XIURONG YU, PEIRONG HAO, CHIGUSA KURUMADA, and WYATTE HALL. 2022. Signing-on-the-Fly: Technology Preferences to Reduce Communication Gap between Hearing Parents and Deaf Children. (2022).

[3] Dhruva Bansal, Prerna Ravi, Matthew So, Pranay Agrawal, Ishan Chadha, Ganesh Murugappan, and Colby Duke. 2021. Copycat: Using sign language recognition to help deaf children acquire language skills. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–10.

[4] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

[5] Naomi K Caselli, Wyatte C Hall, and Jonathan Henner. 2020. American Sign Language interpreters in public schools: An illusion of inclusion that perpetuates language deprivation. *Maternal and Child Health Journal* 24, 11 (2020), 1323–1329.

[6] Nathan A Chi, Peter Washington, Aaron Kline, Arman Husic, Cathy Hou, Chloe He, Kaitlyn Dunlap, and Dennis Wall. 2022. Classifying Autism from Crowdsourced Semi-Structured Speech Recordings: A Machine Learning Approach. *arXiv preprint arXiv:2201.00927* (2022).

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[8] Rina A Doherty and Paul Sorenson. 2015. Keeping users in the flow: mapping system responsiveness with user experience. *Procedia Manufacturing* 3 (2015), 4384–4391.

[9] Yuxiang Gao and Chien-Ming Huang. 2019. PATI: a projection-based augmented table-top interface for robot programming. In *Proceedings of the 24th international conference on intelligent user interfaces*. 345–355.

[10] Judith C Goodman, Philip S Dale, and Ping Li. 2008. Does frequency count? Parental input and the acquisition of vocabulary. *Journal of child language* 35, 3 (2008), 515–531.

[11] Anmol Gulati, James Qin, Chung-Cheng Chiu, Niki Parmar, Yu Zhang, Jiahui Yu, Wei Han, Shibo Wang, Zhengdong Zhang, Yonghui Wu, et al. 2020. Conformer: Convolution-augmented transformer for speech recognition. *arXiv preprint arXiv:2005.08100* (2020).

[12] Matthew L Hall, Wyatte C Hall, and Naomi K Caselli. 2019. Deaf children need language, not (just) speech. *First Language* 39, 4 (2019), 367–395.

[13] Wyatte C Hall. 2017. What you don't know can hurt you: The risk of language deprivation by impairing sign language development in deaf children. *Maternal and child health journal* 21, 5 (2017), 961–965.

[14] Wyatte C Hall, Leonard L Levin, and Melissa L Anderson. 2017. Language deprivation syndrome: A possible neurodevelopmental disorder with sociocultural origins. *Social psychiatry and psychiatric epidemiology* 52, 6 (2017), 761–776.

[15] Wei Han, Zhengdong Zhang, Yu Zhang, Jiahui Yu, Chung-Cheng Chiu, James Qin, Anmol Gulati, Ruoming Pang, and Yonghui Wu. 2020. Contextnet: Improving convolutional neural networks for automatic speech recognition with global context. *arXiv preprint arXiv:2005.03191* (2020).

[16] Bernd Huber, Richard F Davis III, Allison Cotter, Emily Junkin, Mindy Yard, Stuart Shieber, Elizabeth Brestan-Knight, and Krzysztof Z Gajos. 2019. SpecialTime: Automatically detecting dialogue acts from speech to support parent-child interaction therapy. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 139–148.

[17] Tom Humphries, Poorna Kushalnagar, Gaurav Mathur, Donna Jo Napoli, Carol Padden, Christian Rathmann, and Scott Smith. 2016. Avoiding linguistic neglect of deaf children. *Social Service Review* 90, 4 (2016), 589–619.

[18] Tom Humphries, Poorna Kushalnagar, Gaurav Mathur, Donna Jo Napoli, Carol Padden, Christian Rathmann, and Scott R Smith. 2012. Language acquisition for deaf children: Reducing the harms of zero tolerance to the use of alternative approaches. *Harm Reduction Journal* 9, 1 (2012), 1–9.

[19] Inseok Hwang, Chungkuk Yoo, Chanyou Hwang, Dongsun Yim, Youngki Lee, Chulhong Min, John Kim, and Junehwa Song. 2014. TalkBetter: family-driven mobile intervention care for children with language delay. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 1283–1296.

[20] Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E Froehlich. 2015. Head-mounted display visualizations to support sound awareness for the deaf and hard of hearing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 241–250.

[21] Ajay Kumar, Michael E Behen, Piti Singsoonsud, Amy L Veenstra, Cortney Wolfe-Christensen, Emily Helder, and Harry T Chugani. 2014. Microstructural abnormalities in language and limbic pathways in orphanage-reared children: a diffusion tensor imaging study. *Journal of child neurology* 29, 3 (2014), 318–325.

[22] Taeahn Kwon, Minkyung Jeong, Eon-Suk Ko, and Youngki Lee. 2022. Captivate! Contextual Language Guidance for Parent–Child Interaction. In *CHI Conference on Human Factors in Computing Systems*. 1–17.

[23] Emily Laubscher and Janice Light. 2020. Core vocabulary lists for young children and considerations for early language development: A narrative review. *Augmentative and Alternative Communication* 36, 1 (2020), 43–53.

[24] Kaliappan Madasamy, Vimal Shanmuganathan, Vijayalakshmi Kandasamy, Mi Young Lee, and Manikandan Thangadurai. 2021. OSDDY: embedded system-based object surveillance detection system with small drone using deep YOLO. *EURASIP Journal on Image and Video Processing* 2021, 1 (2021), 1–14.

[25] Melissa Malzkuhn and Melissa Herzig. 2013. Bilingual storybook app designed for deaf children based on research principles. In *Proceedings of the 12th International Conference on Interaction Design and Children*. 499–502.

[26] Huisheng Mao, Ziqi Yuan, Hua Xu, Wenmeng Yu, Yihe Liu, and Kai Gao. 2022. M-SENA: An Integrated Platform for Multimodal Sentiment Analysis. *arXiv preprint arXiv:2203.12441* (2022).

[27] Rachel I Mayberry and Robert Kluender. 2018. Rethinking the critical period for language: New insights into an old question from American Sign Language. *Bilingualism: Language and Cognition* 21, 5 (2018), 886–905.

[28] Richard E Mayer and Valerie K Sims. 1994. For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of educational psychology* 86, 3 (1994), 389.

[29] Liyu Meng, Yuchen Liu, Xiaolong Liu, Zhaopei Huang, Wenqiang Jiang, Tenggan Zhang, Yuanyuan Deng, Ruichen Li, Yannan Wu, Jinming Zhao, et al. 2022. Multimodal Emotion Estimation for in-the-wild Videos. *arXiv preprint arXiv:2203.13032* (2022).

[30] Liyu Meng, Yuchen Liu, Xiaolong Liu, Zhaopei Huang, Wenqiang Jiang, Tenggan Zhang, Chuanhe Liu, and Qin Jin. 2022. Valence and Arousal Estimation Based on Multimodal Temporal-Aware Features for Videos in the Wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2345–2352.

[31] Ross E Mitchell and Michaela Karchmer. 2004. Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States. *Sign language studies* 4, 2 (2004), 138–163.

[32] Alex Olwal, Kevin Balke, Dmitrii Votintcev, Thad Starner, Paula Conn, Bonnie Chinh, and Benoit Corda. 2020. Wearable subtitles: Augmenting spoken communication with lightweight eyewear for all-day captioning. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 1108–1120.

[33] Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. 2015. Librispeech: an asr corpus based on public domain audio books. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 5206–5210.

[34] Linkai Peng, Yingming Gao, Binghuai Lin, Dengfeng Ke, Yanlu Xie, and Jinsong Zhang. 2022. Text-Aware End-to-end Mispronunciation Detection and Diagnosis. *arXiv preprint arXiv:2206.07289* (2022).

[35] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.

[36] Lorna Quandt. 2020. Teaching ASL signs using signing avatars and immersive learning in virtual reality. In *The 22nd international ACM SIGACCESS conference on computers and accessibility*. 1–4.

[37] Brian Scassellati, Jake Brawer, Katherine Tsui, Setareh Nasihati Gilani, Melissa Malzkuhn, Barbara Manini, Adam Stone, Geo Kartheiser, Arcangelo Merla, Ari Shapiro, et al. 2018. Teaching language to deaf infants with a robot and a virtual human. In *Proceedings of the 2018 CHI Conference on human Factors in computing systems*. 1–13.

[38] Steven C Seow. 2008. *Designing and engineering time: The psychology of time perception in software*. Addison-Wesley Professional.

[39] Qijia Shao, Amy Sniffen, Julien Blanchet, Megan E Hillis, Xinyu Shi, Themistoklis K Haris, Jason Liu, Jason Lamberton, Melissa Malzkuhn, Lorna C Quandt, et al. 2020. Teaching american sign language in mixed reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–27.

[40] Ashely Tenesaca, Jung Yun Oh, Crystal Lee, Wanyin Hu, and Zhen Bai. 2019. Augmenting Communication Between Hearing Parents and Deaf Children. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 431–434.

[41] Niraj Tolia, David G Andersen, and Mahadev Satyanarayanan. 2006. Quantifying interactive user experience on thin clients. *Computer* 39, 3 (2006), 46–52.

[42] Yile Wang, Leyang Cui, and Yue Zhang. 2019. How Can BERT Help Lexical Semantics Tasks? *arXiv preprint arXiv:1911.02929* (2019).

[43] Kimberly A Weaver and Thad Starner. 2011. We need to communicate! helping hearing parents of deaf children learn american sign language. In *The proceedings of the 13th international ACM SIGACCESS Conference on Computers and Accessibility*. 91–98.

[44] Gregor Wiedemann, Steffen Remus, Avi Chawla, and Chris Biemann. 2019. Does BERT make any sense? Interpretable word sense disambiguation with contextualized embeddings. *arXiv preprint arXiv:1909.10430* (2019).

[45] Kristin Williams, Karyn Moffatt, Jonggi Hong, Yasmeen Faroqi-Shah, and Leah Findlater. 2016. The cost of turning heads: A comparison of a head-worn display to a smartphone for supporting persons with aphasia in conversation. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. 111–120.

[46] Cheng Yi, Jianzhong Wang, Ning Cheng, Shiyu Zhou, and Bo Xu. 2020. Applying wav2vec2. 0 to speech recognition in various low-resource languages. *arXiv preprint arXiv:2012.12121* (2020).