Voice Assistants for Speech Therapy

Ling Qiu
lingq@psu.edu
Pennsylvania State University
State College, PA, USA

Saeed Abdullah
saeed@psu.edu
Pennsylvania State University
State College, PA, USA

ABSTRACT

Many individuals with speech disorders are unable to access speech therapies due to a lack of trained professionals and available resources. To address this issue, there has been an increased focus on using eHealth technologies for remote therapy delivery. For example, there exists a number of mobile applications that aim to deliver different speech therapy exercises (e.g., listening to an audio prompt and repeating it). However, current approaches often are not able to provide personalized feedback or adapt to users' unique needs. Such limitations can impede skill development for practical communications. Furthermore, the current approaches might not support therapeutic alliance — a key factor in determining speech therapy outcomes. In this challenge paper, we explore how voice assistants (VAs) can potentially enable remote delivery of speech therapies at scale. We also describe design challenges that need to be addressed before VAs can be used to deliver speech therapies.

ACM Reference Format:
Ling Qiu and Saeed Abdullah. 2021. Voice Assistants for Speech Therapy. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3460418.3479336

1 INTRODUCTION

According to the National Institute on Deafness and Other Communication Disorders (NIDCD), approximately 7.5 million people in the United States have communication problems using their voices [30]. Many of them endure depression, social isolation and a lower quality of life [26]. Speech therapies, which are conducted by speech-language pathologists (SLPs), are proven to be an effective assessment and treatment to different types of speech disorders [8, 23, 35]. However, there is a shortage of SLPs throughout the world despite increased demand[27, 36]. In addition, some speech therapies, especially the ones for children, might require frequent interactions [25], which can be difficult to provide through traditional clinical services. Therefore, it is necessary to explore options for scalable and remote delivery of speech therapy as well as supporting at-home practices to supplement in-person therapy sessions. The COVID-19 pandemic has further demonstrated the need for flexible and remote delivery of speech therapy [39].

A number of recent studies have explored the use of telerehabilitation and mHealth as alternatives to in-person speech therapy sessions [9, 10, 18, 24, 27, 38]. These approaches could provide greater access to speech therapy services with lower costs especially for individuals living at remote locations [2, 27]. Meanwhile, the scalability and efficacy of telerehabilitation are restricted by various limitations, including not having sufficient qualified SLPs [6, 14] and adequate internet infrastructures [10, 27]. Most of the mHealth apps focus on one-way speech therapy exercises serving to promote therapeutic ends, which may not correspond to communicative goals in daily lives. In addition, the one-way interaction might fail to create and sustain a strong therapeutic alliance between users and healthcare providers. Research shows that therapeutic alliance during speech therapy is a critical determinant for treatment outcome, user engagement and satisfaction [8, 15, 21].

Recent studies have leveraged voice assistants (VAs) for a range of health care applications including enhancing pediatric clinical practices [42] and delivering supportive care interventions for women with metastatic breast cancer [33]. VAs use voice interactions to offer two-way realistic conversations between devices and users [13]. That is, VAs use speech recognition and natural language processing to understand user utterances and respond accordingly. Given the current ability of VAs to support voice interactions, we believe they can potentially be used to deliver speech therapies at scale. Since speech therapy itself is carried out as interactive conversations between clinicians and patients, VAs can lead to a more effective delivery of speech therapy compared to touch screen interactions (e.g. clicking and scrolling). Research shows that maintaining the natural flow of conversation in a speech therapy session could yield more practical outcomes (e.g. everyday communication abilities) in comparison to therapies that focus on facilitating therapeutic skills. Voice interaction can also significantly improve accessibility of technologies to this population who might live with varying levels of impairment (e.g. language, cognition, and motor function) [6, 13, 29]. Lastly, VAs could apply the in-person speech therapies' communicative techniques of building strong therapeutic alliance throughout their delivery process of therapies to facilitate user engagement and therapy efficacy. In this paper, we will briefly review the current technology-driven approaches of delivering speech therapies and the unique affordances and challenges of leveraging VAs to deliver speech therapies.

2 CURRENT TECHNOLOGY-DRIVEN APPROACHES TO DELIVER SPEECH THERAPIES

Telerehabilitation in speech therapy refers to the use of telecommunications technology for delivering speech therapies at a distance[6].

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3479336

Telerehabilitation has been widely adopted by SLPs in some regions of the world [14]. Research shows that telerehabilitation could yield positive clinical outcomes in a range of speech disorders, including childhood speech and language disorders [17], fluency disorders [7], and neurogeneic communication disorders [41]. However, various challenges impede the scalability and quality of current approaches. Telerehabilitation for speech therapy comprises synchronous (real-time delivery), asynchronous and hybrid deliver formats [2]. For synchronous delivery, healthcare providers need to be on the scene during the delivery process (e.g., videoconferencing between a healthcare provider and a patient) and this limits the number of patients they can support. Lack of healthcare providers could create scalability issue for synchronous telerehabilitation in some regions [27]. The real-time delivery also requires adequate internet access (i.e., high internet speed and bandwidth), which can create access barriers for some populations [10, 27]. Asynchronous delivery, on the other hand, can lead to lower lever of engagement, lack of personalized feedback, and lower perceived support [10]. As a result, asynchronous delivery can result in reduced post-treatment satisfaction, which is found to be highly correlated with treatment effectiveness [10].

A number of research and commercial mHealth apps exist to provide speech therapy to both adults and children with speech disorders. Despite serving different population, most of the existing apps focus on therapy exercises that are both phonological and semantic. For example, the Language Therapy app [37] provides SLPs-verified speech therapy exercises for aphasia. It specifically focuses on four categories of activites: reading, naming, comprehension, and writing. Articulation Station [24] — another popular iPad app — uses flashcards, matching games, short stories to teach and manage children's spontaneous speech production and reading. Duval et al. [9] created SpokeIt, a mobile app that combines offline articulation therapy games and speech recognition system capable of providing real-time feedback. These preliminary studies have shown the promising results regarding the acceptability and feasibility for individuals with speech disorders. Although some of these apps claim their efficacy of improving speech disorders, they lack a clear clinical evidence of effectiveness to support these claims [40]. In addition, no previous study has shown whether these apps could establish and sustain therapeutic alliance with patients.

3 LEVERAGING VOICE ASSISTANTS TO DELIVER SPEECH THERAPIES

3.1 Unique Affordances of VAs

Voice communication plays a significant role in face-to-face speech therapy process. It is the communication medium for almost all of the therapeutic related tasks ranging from instructing therapy exercises to planning therapeutic goals. Ferguson et al. [12] suggest that comparing to only focusing on facilitating skills to promote therapeutic ends, maintaining the natural flow of conversation in a speech therapy session could be more beneficial to enhance patients' ability to communicate outside of the session. For example, while therapy activities often involves correcting one's ("self-repairing") pronunciations and utterances, these conducts might be inappropriate or disturbing to communications in natural, non-clinical contexts [12]. Therefore, many SLPs have moved toward a more collaborative and social approach for the therapy process [16]. Voice interfaces in VAs can also adapt similar collaborative approaches and social cues to deliver speech therapies to individuals with speech disorders.

Therapeutic alliance has been proven to be a critical factor in determining the efficacy and long-term engagement of speech therapies [20, 22]. David et al. [8] found that nonspecific treatment factors such as SLP's support and encouragement may contribute to the improvement of recovery process for individuals with speech disorders. However, current mobile apps for speech therapy require users to initiate most of the interactions and they lack the same emotional and individualized support compared to in-person therapy sessions. Previous research has also suggested that leveraging human-to-human relational strategies could help to construct long-term, social-emotional relationships between users and agents [5]. To increase the user engagement and treatment efficacy, VAs can leverage their unique voice interfaces to apply tailored SLPs-verified verbal communication techniques throughout the speech therapy process. For example, some SLPs use humour to alleviate patients' embarrassment and provide light relief within the context of difficult therapy sessions [22]. Similarly, VAs could also support and encourage patients with humorous responses when they encounter challenges managing the therapy exercises.

3.2 Challenges to Use VAs for Speech Therapies

There are a number of challenges that need to be addressed before VAs can be used to effectively deliver speech therapies at scale.

3.2.1 *Ensuring Speech Recognition Accuracy.*

The Automated Speech Recognition (ASR) is the process of converting speech to text, and it is an essential factor to the usability of VAs. However, research shows that the ASR performance in commercial VAs can be sub-optimal for utterances from individuals with speech disorders [3]. This might be due to the fact that current ASR models are not trained using speech data from this population. The resultant low performance of the ASR model in a VA can lead to user frustration and low usability. To be able to use VAs for speech therapy, it is necessary to establish a benchmark ASR model accuracy for individuals with different types and levels of speech disorders. This will allow us to identify design constraints and requirements. Furthermore, the resultant data might encourage commercial VAs to train and update their ASR models to better accommodate utterances from individuals with speech disorder.

3.2.2 *Identifying Attributes and Cues in VAs for Therapeutic Alliance.*

SLPs use various communicative strategies to develop and manage therapeutic alliance with patients in face-to-face therapy sessions. For example, some of these strategies at the beginning of a therapy include recognizing patients' personhood via the use of self-disclosure, helping patients to make realistic expectations, and encouraging patients to take ownership of their rehabilitation [22]. Research shows that VAs' personalities may impact users' perception and communication satisfaction [32, 34]. For example, Poushneh [32] found that users feel more in control when they interact with VAs that demonstrates functional intelligence, sincerity, and creativity. VAs can also use random backchanneling to support active listening [28]. However, there has not been much work in identifying VA attributes and cues that might lead to create and sustain therapeutic alliance. Future work should explore how to effectively design interactions in VAs to ensure therapeutic allaince during speech therapy.

3.2.3 Adapting Speech Therapy Content for VAs. A speech therapy often comprises three parts: evaluation, treatment, and at-home programs [1]. The evaluation process of speech disorders might not be feasible for commercial VAs to carry out because it requires quantitative analysis of voice signals [31] and most commercial VAs do not allow collecting raw audio data for privacy reasons. Therapy sessions and at-home programs varies greatly from different types of speech disorders. For example, treatments for apraxia mainly target at articulation or prosody and treatments for phonemic paraphasia focuse on improving sentence repetition and self-monitoring [19]. The content of these therapy exercises should be transformed to be appropriate for VAs with a design emphasis on leveraging the two-way voice interaction. Furthermore, HCI researchers could potentially conduct interviews or co-design activities with SLPs and patients to determine the therapy contents.

3.2.4 Enabling Family and Group Interactions. Group treatment, where multiple individuals with speech disorders participate in treatment together, is an integral element in the speech therapy practice [4]. Research shows that the social environment of group treatment encourages patients to initiate more communication than one-on-one speech therapies [11]. Future work should explore how to better integrate group treatments in the VAs for speech therapies. Similarly, VAs can also better facilitate family interactions during at-home exercises (e.g., allowing parents to schedule and adapt practice content depending on the unique need of a child).

3.2.5 Ethical Challenges in the Design. Researchers also need to address ethical challenges in using voice assistants to deliver speech therapies. For example, current voice assistants only provide users with homogeneous and standard communication styles. This might pose both efficacy and ethical issues for the minorities with different accents and language styles. It is noteworthy that a similar issue also exists in clinical-based speech therapies, where the SLPs often have a very homogeneous social demographic profile (white, middle class, female) [12]. Secondly, the content of speech therapies, especially the conversational part, should be critically designed and selected to avoid creating any biases.

4 CONCLUSION

VAs can potentially be used to significantly improve access to speech therapies for individuals with speech disorders, especially for the population living in rural and remote areas. However, there are considerable design and implementation challenges before VAs can be used to deliver speech therapies as we have described above. This challenge paper aims to create a discussion within the UbiComp community regarding the opportunity and potential challenges of using VAs for this unique application domain, which can have significant impact on individuals with speech disorder. Specifically, we hope to leverage the workshop discussion to solicit feedback and identify future steps regarding: 1) creation of a VA and ASR performance benchmark for individuals with speech disorders, 2) developing VAs to create and sustain therapeutic alliance.

REFERENCES

[1] [n.d.]. Pediatric Speech Therapy Evaluation and Treatment Flow Chart. https://openlinesny.com/pediatric-speech-therapy-evaluation-and-treatment-flow-chart/
[2] Speech Pathology Australia. 2014. Position Statement Telepractice in Speech Pathology. *Melbourne, Australia: The Speech Pathology Association of Australia Ltd* (2014).
[3] Fabio Ballati, Fulvio Corno, and Luigi De Russis. 2018. Assessing virtual assistant capabilities with Italian dysarthric speech. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 93–101.
[4] Christine Baron, Molly Holcombe, and Candace van der Stelt. 2018. Providing effective speech-language pathology group treatment in the comprehensive inpatient rehabilitation setting. In *Seminars in speech and language*, Vol. 39. Thieme Medical Publishers, 053–065.
[5] Timothy W Bickmore and Rosalind W Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 2 (2005), 293–327.
[6] David M Brennan and Linsey M Barker. 2008. Human factors in the development and implementation of telerehabilitation systems. *Journal of telemedicine and telecare* 14, 2 (2008), 55–58.
[7] Brenda Carey, Sue O'Brian, Mark Onslow, Susan Block, Mark Jones, and Ann Packman. 2010. Randomized controlled non-inferiority trial of a telehealth treatment for chronic stuttering: The Camperdown Program. *International journal of language & communication disorders* 45, 1 (2010), 108–120.
[8] Rachel David, Pam Enderby, and David Bainton. 1982. Treatment of acquired aphasia: speech therapists and volunteers compared. *Journal of Neurology, Neurosurgery & Psychiatry* 45, 11 (1982), 957–961.
[9] Jared Duval, Zachary Rubin, Elena Márquez Segura, Natalie Friedman, Milla Zlatanov, Louise Yang, and Sri Kurniawan. 2018. SpokeIt: building a mobile speech therapy experience. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12.
[10] Maryam Eslami Jahromi, Jamileh Farokhzadian, and Leila Ahmadian. 2021. Two-Sided Perspective on Tele-speech Therapy: Experiences of stuttering patients, and their parents. *Assistive Technology* just-accepted (2021).
[11] Mackenzie E Fama, Christine R Baron, Brooke Hatfield, and Peter E Turkeltaub. 2016. Group therapy as a social context for aphasia recovery: a pilot, observational study in an acute rehabilitation hospital. *Topics in stroke rehabilitation* 23, 4 (2016), 276–283.
[12] Alison Ferguson and Elizabeth Armstrong. 2004. Reflections on speech—language therapists' talk: implications for clinical practice and education. *International Journal of Language & Communication Disorders* 39, 4 (2004), 469–507.
[13] Asbjørn Følstad and Petter Bae Brandtzæg. 2017. Chatbots and the new world of HCI. *interactions* 24, 4 (2017), 38–42.
[14] Raymond Fong, Chun Fung Tsai, and Oi Yan Yiu. 2021. The implementation of telepractice in speech language pathology in Hong Kong during the COVID-19 pandemic. *Telemedicine and e-Health* 27, 1 (2021), 30–38.
[15] Anneka Freckmann, Monique Hines, and Michelle Lincoln. 2017. Clinicians' perspectives of therapeutic alliance in face-to-face and telepractice speech–language pathology sessions. *International Journal of Speech-Language Pathology* 19, 3 (2017), 287–296.
[16] Margaret Freeman. 2004. SLT talk and practice knowledge: a response to Ferguson and Armstrong. *International journal of language & communication disorders* 39, 4 (2004), 481–6.
[17] Sue Grogan-Johnson, Robin Alvares, Lynne Rowan, and Nancy Creaghead. 2010. A pilot study comparing the effectiveness of speech language therapy provided by telemedicine with conventional on-site therapy. *Journal of Telemedicine and Telecare* 16, 3 (2010), 134–139.
[18] Susan Grogan-Johnson, Rodney M Gabel, Jacquelyn Taylor, Lynne E Rowan, Robin Alvares, and Jason Schenker. 2011. A pilot exploration of speech sound disorder intervention delivered by telehealth to school–age children. *International journal of telerehabilitation* 3, 1 (2011), 31.
[19] Tyson G Harmon, Lucy Hardy, and Katarina L Haley. 2018. Proactive social validation of methods and procedures used for training speech production in aphasia. *Aphasiology* 32, 8 (2018), 922–943.
[20] Adam O Horvath, AC Del Re, Christopher Flückiger, and Dianne Symonds. 2011. Alliance in individual psychotherapy. (2011).
[21] Michelle Lawton, Gillian Haddock, Paul Conroy, Laura Serrant, and Karen Sage. 2018. People with aphasia's perception of the therapeutic alliance in aphasia rehabilitation post stroke: a thematic analysis. *Aphasiology* 32, 12 (2018), 1397–1417.
[22] Michelle Lawton, Karen Sage, Gillian Haddock, Paul Conroy, and Laura Serrant. 2018. Speech and language therapists' perspectives of therapeutic alliance construction and maintenance in aphasia rehabilitation post-stroke. *International Journal of Language & Communication Disorders* 53, 3 (2018), 550–563.
[23] Nadina B Lincoln, GP Mulley, AC Jones, E McGuirk, W Lendrem, and JRA Mitchell. 1984. Effectiveness of speech therapy for aphasic stroke patients: a randomised controlled trial. *The Lancet* 323, 8388 (1984), 1197–1200.

[24] Little Bee Speech. [n.d.]. *Articulation Station Pro.* https://apps.apple.com/us/app/articulation-station-pro/id491998279
[25] Edwin Maas, Christina E Gildersleeve-Neumann, Kathy J Jakielski, and Ruth Stoeckel. 2014. Motor-based intervention protocols in treatment of childhood apraxia of speech (CAS). *Current developmental disorders reports* 1, 3 (2014), 197–206.
[26] Ray M Merrill, Nelson Roy, and Jessica Lowe. 2013. Voice-related symptoms and their effects on quality of life. *Annals of Otology, Rhinology & Laryngology* 122, 6 (2013), 404–411.
[27] Haritha S Mohan, Ayesha Anjum, and Prema KS Rao. 2017. A survey of telepractice in speech-language pathology and audiology in India. *International journal of telerehabilitation* 9, 2 (2017), 69.
[28] Nasim Motalebi, Eugene Cho, S Shyam Sundar, and Saeed Abdullah. 2019. Can Alexa be your Therapist? How Back-Channeling Transforms Smart-Speakers to be Active Listeners. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*. 309–313.
[29] Clifford Ivar Nass and Scott Brave. 2005. *Wired for speech: How voice activates and advances the human-computer relationship.* MIT press Cambridge, MA.
[30] NIDCD. 2016. Statistics on voice, speech, and language. (2016).
[31] Juan Rafael Orozco-Arroyave, Juan Camilo Vásquez-Correa, Jesús Francisco Vargas-Bonilla, Raman Arora, Najim Dehak, Phani S Nidadavolu, Heidi Christensen, Frank Rudzicz, Maria Yancheva, H Chinaei, et al. 2018. NeuroSpeech: An open-source software for Parkinson's speech analysis. *Digital Signal Processing* 77 (2018), 207–221.
[32] Atieh Poushneh. 2021. Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors. *Journal of Retailing and Consumer Services* 58 (2021), 102283.
[33] Ling Qiu, Bethany Kanski, Shawna Doerksen, Renate Winkels, Kathryn H Schmitz, and Saeed Abdullah. 2021. Nurse AMIE: Using Smart Speakers to Provide Supportive Care Intervention for Women with Metastatic Breast Cancer. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 302, 7 pages. https://doi.org/10.1145/3411763.3451827
[34] Pejman Sajjadi, Laura Hoffmann, Philipp Cimiano, and Stefan Kopp. 2019. A personality-based emotional model for embodied conversational agents: Effects on perceived social presence and game experience of users. *Entertainment Computing* 32 (2019), 100313.
[35] Ann Bosma Smit. 2004. *Articulation and phonology resource guide for school-age children and adults.* Cengage Learning.
[36] Katie Squires. 2013. Addressing the Shortage of Speech-Language Pathologists in School Settings. *Journal of the American Academy of Special Education Professionals* 131 (2013), 137.
[37] Brielle C Stark and Elizabeth A Warburton. 2018. Improved language in chronic aphasia after self-delivered iPad speech therapy. *Neuropsychological rehabilitation* 28, 5 (2018), 818–831.
[38] Tactus Therapy Solutions Ltd. [n.d.]. *Conversation Therapy.* https://apps.apple.com/us/app/conversation-therapy/id620456076
[39] Seyed Abolfazl Tohidast, Banafshe Mansuri, Rasool Bagheri, and Hadi Azimi. 2020. Provision of speech-language pathology services for the treatment of speech and language disorders in children during the COVID-19 pandemic: Problems, concerns, and solutions. *International journal of pediatric otorhinolaryngology* 138 (2020), 110262.
[40] Atiyeh Vaezipour, Jessica Campbell, Deborah Theodoros, and Trevor Russell. 2020. Mobile Apps for Speech-Language Therapy in Adults With Communication Disorders: Review of Content and Quality. *JMIR mHealth and uHealth* 8, 10 (2020), e18858.
[41] Kristen Weidner and Joneen Lowman. 2020. Telepractice for adult speech-language pathology services: a systematic review. *Perspectives of the ASHA Special Interest Groups* 5, 1 (2020), 326–338.
[42] Jayme L Wilder, Devin Nadar, Nitin Gujral, Benjamin Ortiz, Robert Stevens, Faye Holder-Niles, John Lee, and Jonathan M Gaffin. 2019. Pediatrician Attitudes toward Digital Voice Assistant Technology Use in Clinical Practice. *Applied clinical informatics* 10, 2 (2019), 286.