Rochester Institute of Technology
B. Thomas Golisano College
of
Computing and Information Sciences

Master of Science in Human Computer Interaction

~ Project Proposal Approval Form ~

Student Name: Name Yugo Iwamoto

Project Title: *Enhancing Inclusive Conversations: AI-Powered Speech Services for Daily Conversations*

Project Area(s): (√ primary area)

☐ Application Dev.	☐ Database	☐ Website Dev.
☐ Game Design	√ HCI	☐ eLearning
☐ Networking	☐ Project Mngt.	☐ Software Dev.
☐ Multimedia	☐ System Admin.	☐ Informatics
☐ Geospatial	☐ Other _____	

~ MS Project Committee ~

Name	Signature	Date
Tae Oh	Signed by: Tae Oh C54A3417F04F4B8...	12/20/2024
Chair		
Mick McQuaid	Signed by: C95EEC524E5545C...	12/20/2024
Committee Member		
Committee Member (optional)		

Approved: _____ Date: ____/____/____ ☐ electronic copy received

Enhancing Inclusive Conversations: AI-Powered Speech Services for Daily Conversations

By

Yugo Iwamoto

This capstone proposal is submitted in partial fulfillment of
the Master of Human Computer Interaction
in the B. Thomas Golisano College of Computing and Information Sciences
at the Rochester Institute of Technology

Date

12/18/2024

Enhancing Inclusive Conversations: AI-Powered Speech Services for Daily Conversations

YUGO IWAMOTO

Rochester Insitute of Technology, Golisano College of Computing and Information Sciences, School of Information, 152 Lomb Memorial Dr, Rochester, NY, 14623, USA
yi9686@rit.edu

Abstract: This capstone project aims to bridge communication barriers between Deaf and hard of hearing (DHH) and hearing individuals by developing an AI-powered prototype utilizing real-time speech-to-text (STT) technology. The system leverages Open AI API speech recognition capabilities to facilitate conversations, providing accessibility in environments where interpreters are limited. Combining hardware components such as a mobile phone's microphone and monitor with user-centered software design, the prototype captures and displays real-time speech transcriptions, enabling inclusive communication. The project emphasizes accessibility, usability, and adaptability, integrating user-centered design principles to ensure a positive experience for diverse users. User testing with DHH and hearing participants will evaluate the system's effectiveness, focusing on usability, accuracy, and real-world feasibility. The findings will contribute valuable insights into human-computer interaction and assistive technology, with potential applications in educational, professional, and social settings.

1. Introduction

AI-Powered Speech Recognition for Inclusive Conversations aims to investigate usability in communications between hearing and deaf/hard-of-hearing (DHH) people by building a prototype system leveraging Open AI speech recognition capabilities. The project bridges communication gaps by enabling seamless conversations in everyday situations where ASL interpreters are unavailable or limited. In addition, some people with hearing impairment may not use ASL or any other sign language to communicate, and they may benefit from this project to overcome some communication barriers. This prototype will be able to assist some individuals with hearing aids and allow accurate and effective conversations. The prototype system is an application requiring a mobile phone with a microphone and a computer monitor. A designed, user-friendly interface will assist in seamless communication. The back end utilizes Open AI/ ChatGPT resources to enable STT technology. Spoken words are captured, processed, and displayed on a screen in real-time, providing DHH users with immediate access to the conversation. DHH people will type it and display it on the interface, enabling two-way conversation. Additionally, the system emphasizes accessibility and inclusion through user-centered design principles. The project investigates the usability, effectiveness, and potential adaptability of speech recognition as a standalone assistive technology. It explores challenges such as accuracy in diverse environments, ease of use for participants, and its impact on conversation with typing inputs. This technology is not intended to replace interpreters but rather to act as a supportive tool for spontaneous or casual conversations for those familiar with English. The study will evaluate user experience, accuracy, and the overall feasibility of real-time speech recognition to foster inclusive communication.

2. Motivation and Goals

Approximately 3.7 % of Rochester, NY's population is DHH. Rochester Institute of Technology (RIT) is home to the National Technical Institute for the Deaf (NTID), making it a prime environment for technology that addresses accessibility challenges. [14] Students, faculty, and staff at RIT are required to communicate with DHH individuals. However, those who do not know ASL, like myself, often struggle to communicate effectively without the assistance

of an interpreter, writing on paper, or typing on the phone. TigerChat [12] is available as a communication aid on campus, but there are some usability issues that are not widely used across the RIT community.

2.1. Problem Statement

Communication barriers between DHH and hearing individuals persist due to limited resources, including the shortage of sign language interpreters and the lack of proficiency in sign language among hearing individuals. Existing STT solutions are promising but often fail to meet the accuracy, latency, and usability requirements for real-time, inclusive conversations.

2.2. Goals

This project seeks to provide a better understanding of the capability of AI speech recognition and these usabilities in the daily conversations between two parties. Designing the prototype with a combination of software and hardware allows us to investigate AI speech recognition's benefits, limitations, and future potentials with the combination of typing inputs. The impact extends beyond the academic setting, with potential applications in workplaces, public service, and social environments. This project is designed to address the needs of the DHH community and contribute to the growing body of research in human-computer interaction (HCI) and assistive technology.

2.3. Research Questions

Designing the prototype will allow me to investigate the effectiveness of AI speech recognition as an assistive technology to enable daily conversations between hearing and DHH individuals. The primary research question is: *How effective is AI-powered speech recognition technology with the combination of typing inputs in enabling accurate and accessible conversations between DHH and hearing users?* The primary question seeks to determine the effectiveness of AI speech recognition in the conversation. To support this, usability and feasibility are critical aspects to examine. The following questions will guide the study: *What are the user experiences and perceptions of real-time AI-driven STT technology regarding accessibility, usability, convenience, and speed?* The question looks into the effectiveness of speech-to-text by itself to support conversation between hearing and DHH people. Then, *What is the accuracy rate of Open AI speech-to-text system, and how does it perform under different environmental conditions such as background noise, speaker accents, and varying speech speeds?* which investigate the accuracy of capturing the spoken language as accuracy is another key aspect of information delivery. These research questions aim to evaluate the system's effectiveness in promoting inclusive communication, focusing on real-world applicability and user satisfaction.

2.4. Objectives

The primary objective is to investigate the effectiveness and usability of AI-powered STT technology in facilitating daily conversations between hearing and DHH people. This study assesses how Open AI's speech recognition capabilities can overcome communication barriers in real-time scenarios. Key objectives include:

- *Effectiveness Assessment:* Measuring the accuracy and reliability of STT conversion in real-time conversations with typing inputs, mainly focusing on diverse environment conditions such as background noise, speaker accents, and speech speed variations.
- *User Experience Evaluation:* Assessing user satisfaction, accessibility, usability, and the overall conversational flow during interactions mediated by the prototype.

- *Accuracy of AI Speech Recognition:* Conducting initial tests to evaluate the performance of Open AI's speech recognition in accurately transcribing speech under varying conditions. This will include factors such as speech clarity, environmental distractions, and microphone quality.

- *Feedback Gathering:*Observing and collecting feedback from DHH and hearing participants during user studies to identify strengths and limitations of the prototype. Insights will guide further improvements and adaptations.

This project aims to create a practical and accessible communication tool for spontaneous conversations, enhancing inclusivity and usability for people in real-world settings.

2.5. Potential Contributions & Benefits

This project will contribute to the fields of HCI and assistive technology by leveraging Open AI's enabling real-time speech recognition to address communication barriers between hearing and DHH parties. By focusing on speech recognition as an assistive technology, the project aims to create an effective and accessible solution that fosters inclusion and facilitates accurate communication. Beyond addressing immediate needs within the RIT community, the broader impact of this project includes its adaptability to various settings, such as workplaces, educational institutions, public services, and social environments. By enabling real-time conversations in diverse scenarios, this technology has the potential to empower DHH individuals to participate more fully in daily interactions. Through its emphasis on inclusivity and accessibility, it contributes to ongoing efforts to bridge the communication divide and promotes empathy and understanding in technological development. By empowering DHH individuals to engage on equal terms, this project enhances personal agency and supports a cultural shift toward universal access and inclusivity in communication technologies. In addition, the project will provide valuable user experience data that can inform the future development of speech recognition systems, ensuring they are tailored to meet the needs of diverse users. By focusing on usability, effectiveness, and adaptability, this work sets the foundation for scalable solutions in assistive technology.

3. Related Work

The project focuses on leveraging AI-powered speech recognition to address communication barriers between DHH and hearing people. The prototype I aim to design is very similar to Shezi et al., [13], But I will have a monitor to output the speech and typing. Below is a review of existing work related to speech recognition and the communication challenges faced by the DHH community.

3.1. Speech Recognition

Speech recognition is the major frame of this project, providing a real-time communication solution for DHH individuals. Accurate automatic speech recognition (ASR) is critical for success, particularly in accessibility-focused applications. Systems such as Google Speech API and Microsoft Azure Speech Services have demonstrated high accuracy and reliability in converting spoken language into text in real-time [3]. Microsoft Azure Speech Services, in particular, offers robust features like speaker diarization, real-time transcription, and customizable models to handle domain-specific vocabularies. These features make it a suitable candidate for applications targeting inclusivity and accessibility. Artificial intelligence will contribute a lot to speech recognition and conversation using STT. Emotion is one of the key factors of communication, and Luo [7] investigated emotional recognition using deep learning. Although emotion recognition within speech is still under development, the environment is a great factor

that affects the accuracy of speech recognition, such as noise. In a public setting, noise can be louder and affect the accuracy; Liu et al., [5] investigate wave signals to improve the accuracy of speech recognition in public settings. Research by Kushalnagar et al. [4] emphasizes the importance of latency and readability in STT systems, as optimized displays can significantly enhance user experiences for DHH individuals. Utilizing ASR and training language model tried to aim the human intelligence level of language recognition despite the accents [2]. ASR is very flexible and adaptable to different scenarios and situations. ASR can be a great assistive technology for individuals with speech disorders [6]. Liu et al. investigated the approach of multiple ASR models to improve the recognition and error rates in Aster, one of the speech disorders.

3.2. Communication Barriers in the DHH Community

The DHH community faces a range of pervasive communication barriers that hinder their full participation in societal, educational, and professional contexts. One significant challenge is the reliance on traditional communication aids, such as interpreters, which are often unavailable. This shortage is particularly evident in high-demand fields such as healthcare, education, and legal services [9]. The dependency on certified ASL interpreters creates a systemic gap, leaving DHH individuals without access to critical conversations or services. Another issue lies in the lack of ASL proficiency among hearing individuals, which contributes to limited interaction and a broader lack of understanding of the DHH community's needs. This gap perpetuates feelings of frustration, isolation, and exclusion, particularly in educational environments where inclusive teaching practices are critical for student success. According to Napier et al., [8], interpreter shortages are exacerbated by the increasing demand for inclusive education, highlighting the need for technological alternatives. Societal attitudes further exacerbate these challenges. A lack of awareness and understanding of the unique needs of DHH individuals often leads to communication avoidance or exclusionary practices, creating additional hurdles for meaningful engagement. Research by Shezi et al., [13] indicates that accessibility technologies such as STT can reduce the barriers in communications by utilizing mobile devices.

4. Methodology

4.1. Prototype (Technical Details)

The prototype is a critical component of this research, combining hardware and software to deliver an STT solution tailored for communication between hearing and DHH individuals. The hardware setup includes a microphone on mobile devices to capture speech and a monitor to display the transcribed text. The software leverages Open AI for real-time speech recognition and transcription. On the other hand, DHH user will type on their mobile phone allowing access to the conversation, and the feedback will be on the monitor. The interface will be designed with HTML, CSS, and JavaScript, emphasizing accessibility and user-centered design principles. The monitor will display transcriptions of spoken words in real time, enabling a seamless conversation experience. The system architecture ensures minimal latency and high readability to accommodate the needs of diverse users. Utilizing a QR code will allow users to get to the chat easily. Figure 1 illustrates the expected hardware setup.

4.2. User Testing

To evaluate the prototype's usability and effectiveness, structured user testing will involve a minimum of 16 participants, comprising eight DHH individuals and eight hearing individuals. Since qualitative data mainly focuses on collecting data, justifying the number of participants with *Local Standards for Sample Size at CHI*. [1] Participants will be selected based on a pre-screening

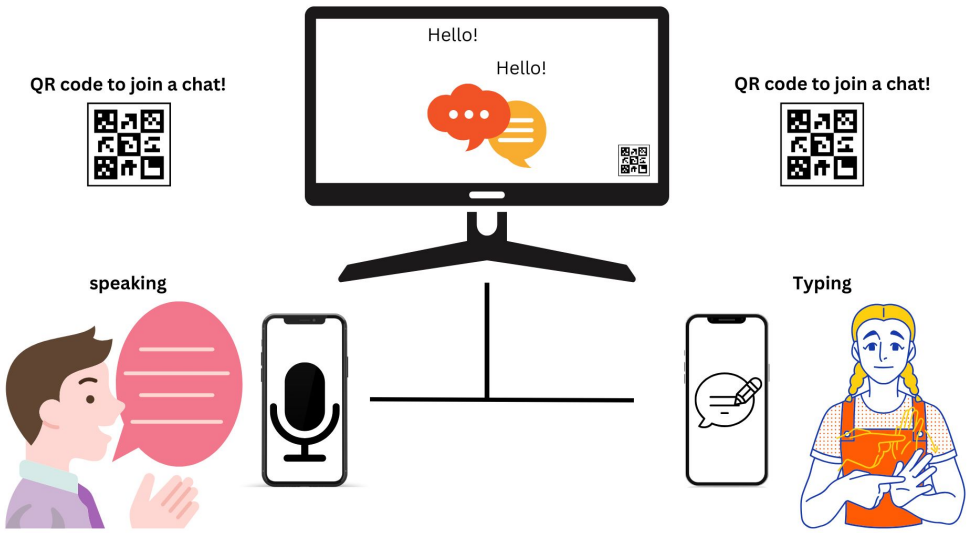

Fig. 1. Showing the hardware setup as a monitor is in the middle, and speech will be captured by microphone on the phone, and DHH people will type their input on their phone.

questionnaire to ensure diversity in demographic backgrounds and conversational preferences. The testing process will include the following steps:

1. Agreement and Setup: Participants fill out a consent form and complete a short orientation.
2. Introductions: Participants engage in a two-minute introductory conversation using the prototype.
3. Main Conversation: A pair of hearing and DHH participants engage in a 10-minute conversation using the prototype (refer to Section 4.4).
4. NASA TLX Survey: Participants complete the NASA Task Load Index (TLX) survey online to assess cognitive and physical effort [10].
5. Post-Completion Interviews: Participants provide feedback on their experience in an interview session, which may be conducted online for convenience (refer to Section 4.3).

The conversation topics will be pre-determined and relevant to the participant's interests, ensuring natural engagement during testing.

4.3. Post-Completion Survey

An Example of a post-completion survey is listed below.

1. What is your overall experience of this conversation?
2. Do you believe that the conversation is smooth? Why or why not?
3. Were there any moments during the interaction when you felt the conversation was interrupted or it was difficult to understand the script?
4. Did you encounter any challenges while using the system? If so, please describe them.

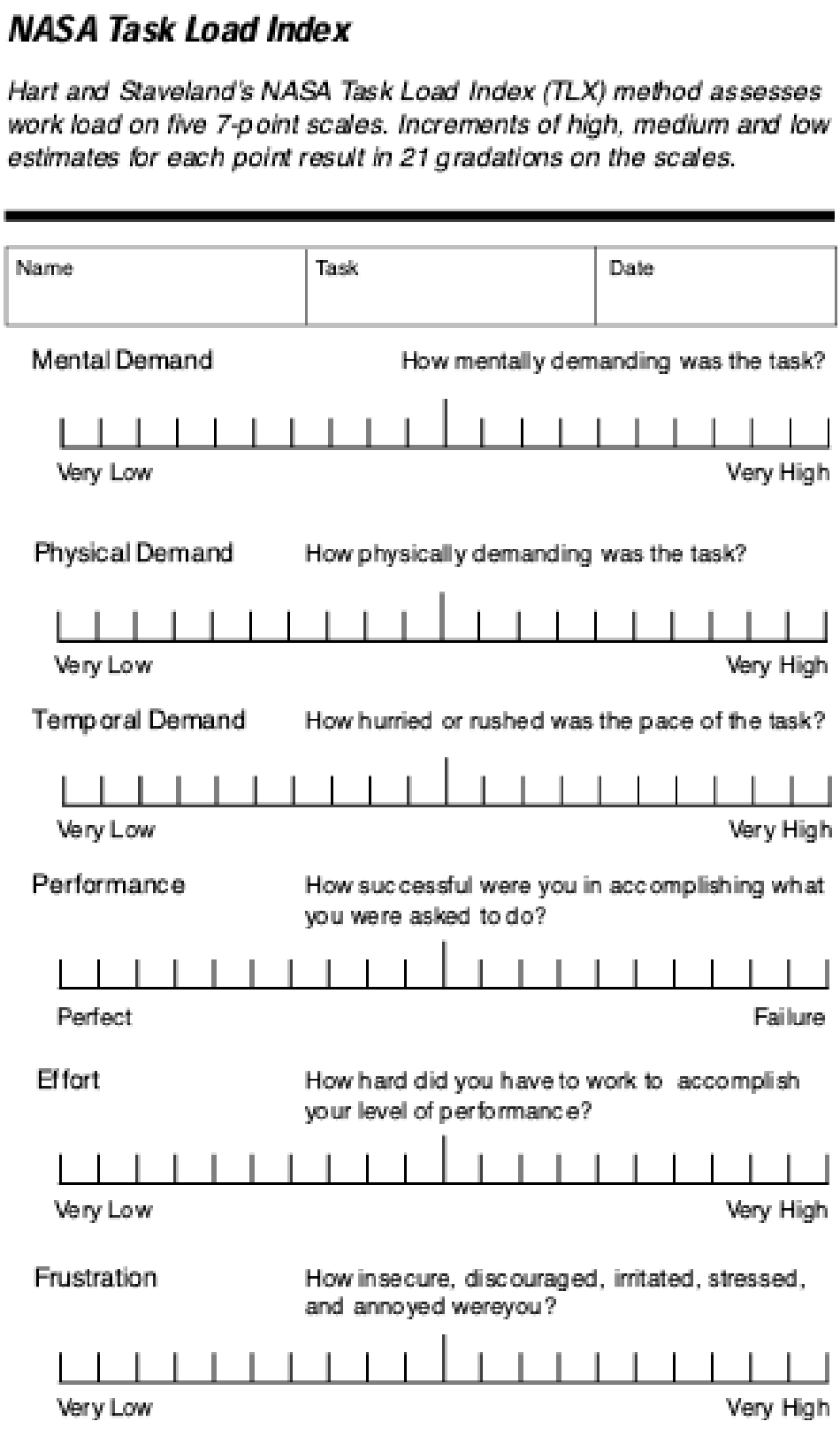

Fig. 2. NASA TLX is the list of measurements, including mental state.

5. Would you use this system in real-world scenarios? Why or why not?
6. What improvements or features would you recommend adding to this system?

4.4. Conversation Topic Ideas

Here are some ideas for conversation topics, but this is not limited to these topics, and we will investigate topics based on the prescreening during recruitment.

- Favorite activity to do on campus at RIT
- Your favorite Rochester restaurant
- Your favorite city in this world

4.5. Data Collection and Analysis

The study employs a mixed-methods approach to collect both qualitative and quantitative data.

- *Qualitative Data:* Observations during user testing will document participants' interactions with the prototype, focusing on the conversational flow, ease of use, and any noticeable challenges. Video recordings will be used to ensure a detailed and accurate analysis of nonverbal cues such as body language and facial expressions.
- *Quantitative Data:* The NASA Task Load Index (TLX) will provide metrics on perceived workload, including mental demand, effort, and frustration. These scores will serve as quantitative measures of the system's usability and effectiveness.

Post-completion surveys will gather participants' subjective feedback, including their opinions on the system's strengths, limitations, and suggestions for improvement. Open-ended questions will allow for detailed responses, capturing nuanced insights into the user experience. This dual approach will provide a comprehensive understanding of the prototype's usability and potential for real-world application.

5. Challenges/Barriers & Limitations

5.1. Challenges/Barriers

Several challenges arise in achieving the goals of this research. While OpenAI speech recognition is robust, it is still a challenge to achieve high accuracy in diverse environments (e.g., noisy backgrounds, varying accents, and speech speeds). Ensuring consistent and reliable transcriptions is critical to the system's usability. Secondly, the conversation will be scripted on the monitor; participants will not look at each other, which will most likely result in missing nonverbal behaviors in communication. Thirdly, user privacy is paramount as the system captures and processes audio data in real time, so addressing ethical concerns about data privacy is important. Clear data storage, processing, and deletion policies will be implemented to protect user confidentiality and comply with ethical standards. Lastly, hardware limitations ensuring the microphone's sensitivity and the monitor's readability for diverse users may present hardware challenges. The setup must be cost-effective and accessible while maintaining high performance.

5.2. Limitations

This project also faces inherent limitations. Firstly, it is dependent on technology. The reliance on Open AI speech recognition requires stable internet connectivity, which may limit its use in environments with poor network access. Secondly, environmental constraints on speech recognition accuracy may be affected by environmental factors such as background noise,

Date	
12/10/2024	complete proposal
12/15/2024	Submit IRB
	Work on prototype and finish testing prototype
1/30/2025	Hope to get IRB approval by!
2/1/2025	Start collecting data
2/20/2025	Data analysis and work on deliverable
4/30/2025	defending capstone
5/5/2025	final deliverable submit

Fig. 3. The timeline of the capstone project and aim to deliver the final products by May 5, 2025.

overlapping speech, or the quality of audio capture. Thirdly, the generalization of the system may not accommodate all user accents or speech patterns effectively. Lastly, not every DHH user is familiar with speaking language, and those who are not familiar with speaking language will be limited to utilizing or participating in this study. Although this study aims to expand and improve accessibility and efficiency, there will be some limitations that prevent some individuals from communicating. Customizing the language model to handle diverse linguistic nuances requires further research and development. In addition, the limited number of participants (16) in the user testing phase may restrict the generalizability of findings. A larger sample size would provide more robust insights.

6. Timeline of this Project & delivery method

The timeline of this capstone project is presented in Figure 3, outlining critical milestones from prototype development to final testing and documentation. The final product will come with the document and prototype with user testing data.

References

1. Caine, K. 2016. Local Standards for Sample Size at CHI. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI' 16), ACM, New York, NY, USA, 981–992. https://doi.org/10.1145/2858036.2858498.

2. Ghorbani, Shahram, and Hansen, John H.L. 2022. Domain Expansion for End-to-End Speech Recognition: Applications for Accent/Dialect Speech. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31, 762–774. IEEE Press. https://doi.org/10.1109/TASLP.2022.3233238.

3. Kuhn, K., Kersken, V., Reuter, B., Egger, N., & Zimmermann, G. 2024. Measuring the Accuracy of Automatic Speech Recognition Solutions. ACM Transactions on Accessible Computing (TACCESS), 16, 4, Article 25, 23 pages. https://doi.org/10.1145/3636513.

4. Kushalnagar, R. S., Behm, G. W., Kelstone, A. W., & Ali, S. 2015. Tracked Speech-To-Text Display: Enhancing Accessibility and Readability of Real-Time Speech-To-Text. Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS' 15), ACM, 223–230. https://doi.org/10.1145/2700648.2809843.

5. Liu, Tiantian, Wang, Chao, Li, Zhengxiong, Huang, Ming-Chun, Xu, Wenyao, & Lin, Feng. 2024. Wavoice: An mmWave-Assisted Noise-Resistant Speech Recognition System. *ACM Transactions on Sensor Networks (ACM Trans. Sen. Netw.)*, 20(4), Article 86, 29 pages. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3597457.

6. Liu, Yi, Li, Yuekang, Deng, Gelei, Juefei-Xu, Felix, Du, Yao, Zhang, Cen, Liu, Chengwei, Li, Yeting, Ma, Lei, and Liu, Yang. 2024. Aster: Automatic Speech Recognition System Accessibility Testing for Stutterers. In *Proceedings of the 38th IEEE/ACM International Conference on Automated Software Engineering (ASE '23)*, Echternach, Luxembourg, 510–521. IEEE Press. https://doi.org/10.1109/ASE56229.2023.00107.

7. Luo, Sandra. 2024. Navigating the Diverse Challenges of Speech Emotion Recognition: A Deep Learning Perspective. In *Proceedings of the 27th International Academic Mindtrek Conference (Mindtrek '24)*, Tampere, Finland, 133–146. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3681716.3681725.

8. Napier, J., & McKee, R. (2018). Interpreting and Accessibility in Education: Addressing the Shortages. Journal of Inclusive Education, 22(5), 451–467. https://doi.org/10.1080/13603116.2017.1375334.

9. National Deaf Center. 2024. The ASL Interpreter Shortage and Its Impact on Accessibility in College Settings. National Deaf Center. Retrieved November 16, 2024, from https://nationaldeafcenter.org/news-items/the-asl-interpreter-shortage-and-its-impact-on-accessibility-in-college-settings/.

10. NASA Human Systems Integration Division. 2024. NASA Task Load Index (TLX). Retrieved November 16, 2024, from https://humansystems.arc.nasa.gov/groups/tlx/.

11. Nikolaraizi, M., & Argyropoulos, V. 2015. The Learning and Communication Barriers of Deaf and Hard of Hearing Students in Higher Education. 7th International Conference on Education and New Learning Technologies (EDULEARN15 Proceedings), IATED, Barcelona, Spain, 4130–4134. https://library.iated.org/view/NIKOLARAIZI2015LEA. Accessed: November 16, 2024.

12. Rochester Institute of Technology. 2024. TigerChat: Real-Time Speech-to-Text Communication Tool. Retrieved November 29, 2024, from https://www.rit.edu/accesstechnology/tigerchat.

13. Shezi, M., Ade-Ibijola, A., and Mtsweni, S. 2020. Deaf Chat: A Speech-to-Text Communication Aid for Hearing Deficiency. *Advances in Science, Technology and Engineering Systems Journal*, 5(5), 826–833. ASTES Journal.

14. WXXI News. 2022. Resources and culture make Rochester a draw for many in the Deaf community. Inclusion Desk, WXXI News, January 24, 2022. https://www.wxxinews.org/inclusion-desk/2022-01-24/resources-culture-make-rochester-a-draw-for-many-in-deaf-community. Accessed: November 16, 2024.