A Speech-To-Text Interface Design for Effective Communication Between Physicians and Deaf Patients

Amitta Southivong[1] and Dr. Dezhi Wu[2*]

[1, 2] Department of Integrated Information Technology, College of Engineering and Computing, University of South Carolina, Columbia, SC 29309, USA

[*] Corresponding Author: dezhiwu@cec.sc.edu

Introduction/Background

- Prior studies show that **deafness and hard-of-hearing (HH) individuals amongst patients has a disproportionate effect on their access to equitable healthcare, leaving them prone to higher rates of health risks** [1, 2]**.**
- Often, deaf or hard-of-hearing patients **leave medical consultations feeling confused or uncertain on what their diagnosis was** or what preventative measures to take **due to lack of interpreters** [3].
- We propose a speech-to-text application that translates spoken languages into written text, **allowing deaf/HH patients to have real-time captioning for conversations.**
- This research **explores the effectiveness of integrating Augmented Reality (AR) and automatic speech recognition** (ASR) for deaf patients [4,5]

AR Introduction

- **AR** is an ongoing developing technology that superimposes images onto a user's viewing screen.
- The **OpenAI Whisper model** used in our initial prototype (Fig. 1) is trained for speech recognition with an average word error rate of 10% - 12.5% and a 95% accuracy rate [6, 7].
- AR/ASR is **more cost effective** as opposed to expensive hearing aids that can cost individuals up to $10,000 when hospitals can adopt this proposed application that can serve thousands of deaf/hard-of-hearing patients [8].
- Augmented reality has been proven to also **positively impact deaf individual's education** [9], and other scenarios where voice conversations are expected such as cash registers [10].

Novelty

Numerous ASR and speech-to-text applications are already made specifically for deaf individuals, this prototype proposes novel UI deign in two areas:

1) Capability of recognizing and translating for accents and English dialects [11].

2) Color indicator when there is a low-confidence of the translation to minimize any mistakes in a diagnosis [12].

Proposed User Study

- ***Participants***: We plan to recruit 10 deaf and/or HH individuals who are the ages of 18 and over.
- ***Qualitative***: Recruited users will be asked to read a script on the prototype in [Fig 1], providing feedback with a pre-survey google doc form and a post-survey.
- ***Quantitative***: User's will be asked to note how many times the prototype had an error each time they read a script. This will be used to account for the error rate and effectiveness.

Research Questions

- **What user needs** are required to engage deaf/HH patients in order to foster trust with their doctors?
- **How** can we develop an AR application that will be accurate enough to translate an entire conversation without leaving room for errors that could possibly lead to more user confusion?
- **Why** do the deaf/ HH patients use the proposed AR system when there are advanced hearing aids?

Prototype in Development

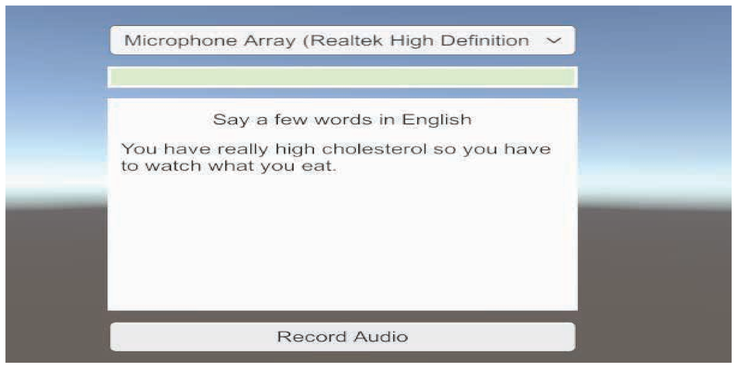

Fig 1. Early-stage prototype UI

Projected Outcomes

There are 3 main potential outcomes of this application's use and development

- ***Accessibility*** due to it being more cost effective and affordable than expensive hearing aids.
- ***Equitable healthcare*** because of the establishment of trust between a doctor and their patient when accurate communication is ensured.
- ***Health safety*** will also be ensured as the patient should feel more confident in their diagnosis or correct dosage/medication to take.

Future Development

- Although OpenAI Whisper model has a 95.3% accuracy rate, further development will use red to indicate low confidence translation, yellow for medium confidence, green text for high confidence translation to account for accuracy.
- Train the data model to understand accents and English Dialects.
- Develop the UI further to overlay the camera's view so that it will appear as a scrolling chat-box conversation.

Acknowledgements

We acknowledge the generous funding support provided by the University of South Carolina (USC) (PI: Dr. Dezhi Wu, Grant 80002838), and to the McNair Junior Fellows program at USC for sponsoring this research.

References

Available upon request