Thesis Project Portfolio

Lost in Compression: Who's Heard and Who's Blurred in Digital Voice Communication?

(Technical Report)

From Voice to Data: How Lack of Transparency Shapes Speech Recognition Performance

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Madison Sullivan

Spring, 2025

Department of Systems and Information Engineering

Table of Contents

Sociotechnical Synthesis

Digital communication has become essential - especially post COVID-19, but processes including speech recognition and Voice Over Internet Protocol (VoIP) produce disproportionate results depending on the user. Speech recognition is used in a multitude of applications for dictation and translation purposes. However, the data used to build these models tends to be proprietary, making it difficult for users to know whether the software is appropriate for their use case, introducing potentially biased results. Similarly, the transmission of audio data using VoIP is done through an encoding and decoding (codec) process, which varies in the amount of data transmitted depending on the bitrate of the codec. My technical project sought to analyze three widely used codecs (OPUS, AMR, CODEC2) at varying bitrates to see where and how bias emerges. The general problem underlying both my STS and technical research is the issue of unequal representation of voices using digital communication processes, leading to ineffective communication. This problem matters as it hinders equitability in voice recognition and minimizes inclusivity in the communication sphere.

For my technical project, my teammates and I investigated the problem of biases in vocal transmission in VoIP, due to differences in degradation level depending on the codec and bitrate. We analyzed over 2,900 speakers with varying demographic backgrounds using 14 audio quality metrics to showcase differences between sex in OPUS, AMR, and CODEC2 across their bitrate ranges. The results showed that bias was present within all codecs. OPUS showed a shift in the direction of bias at the mid-level bitrates while AMR and CODEC2 consistently had more metrics favoring male voices to females. This highlights an important issue of disproportionate voice degradation by sex, potentially hindering the vocal perception of females compared to

males. Similar to speech recognition, there needs to be transparency between creators and users in order to promote equitable voice representation.

My STS research focused on the issue that speech recognition algorithms have minimal transparency from creator to user, causing improper usage and varying success rates, despite their growing popularity. I analyzed this problem by looking at the composition, utilization, and evaluation of speech recognition software. This included the training data and its selection process, case studies showing failure points, and metrics commonly used to measure algorithmic success. Through my analysis, I found that those creating a speech recognition software fail to include data representative of all potential use cases. This in turn can generate poor results for the user if they are one of these cases. Additionally, I found that there is an overall lack of transparency in this technology, as creators do not disclose how their model should be used and by whom. This matters because many dictation and transcription errors have occurred due to this failure. As a result, implications have ranged from incomprehensible automatic captions to improper documentation within a patient's medical note, ultimately causing their death.

My technical project was successful in proving that bias occurs within these three common codecs and that vocal compression tends to favor males based on the metrics used. However, the analysis only considered 14 objective audio quality measures. Results could be improved by expanding the number of metrics tested against or through subjective listening tests to determine if the discovered bias is perceivable to the human ear. My STS research was not as successful at pinpointing one solution to making speech recognition more equitable, but rather explored the process and identified possible improvements. Much of the training data used to build popular speech recognition softwares was unavailable for my analysis, which made it difficult to identify if specific demographic groups face more bias than others, as done in my

technical research. Future work could include speech recognition testing with a known training dataset on users that range in one demographic factor, such as sex, to see how success rates change depending on how representative the training data is of the user base.

I would like to thank my capstone team members, Elizabeth Recktenwald, Catherine Nguyen, and Lucas Vallarino, as well as our advisors, Jad Atweh and Matthew Bolton, for their time and dedication to this project. I would also like to thank my STS professor, Caitlyn Wylie, for her guidance and feedback throughout my research process.