# Evaluating Accuracy of Open-Source Automatic Speech Recognition (ASR) Models With Acoustic Analysis

Choate, Audrey (School: Oviedo High School)

Augmentative and alternative communication (AAC) aids people with speech difficulties, especially ASD; however, limited research exists regarding forms of AAC designed specifically for semi-verbal autistic people. ASR technology provides real-time auditory transcription, which could aid verbal skills while also enabling communication for semi-verbal people. This project aims to evaluate ASR models under adverse conditions using acoustics in order to determine their suitability for AAC. Audio distortion was done by waveform clipping at -5 dB intervals. Whisper, RealtimeSTT, Wav2vec2, and SpeechRecognition all transcribed public domain recordings under increased distortion, and the WER was calculated by comparison with an official transcription. Whisper and Wav2vec2 had a significantly lower increase in WER with increased distortion across all speech samples ($p < .001$ using ANOVA.) RealtimeSTT was inconsistent in its performance, but generally performed better than SpeechRecognition. Three Fast Fourier Transforms were done on each speech at the 0 and -20dB marks to determine the effects of clipping on different frequency bands, showing that higher frequency signals were more significantly affected. A graph of the signal-to-noise ratio and the WER showed a slight negative correlation for RealtimeSTT and Whisper, with no clear trend for the other models. This study shows that models such as Whisper and Wav2vec2 perform the best in adverse conditions, and thus may be more suited to AAC applications. More research is needed regarding different distortion methods and ASR error sources. More research is needed to determine if the effect of clipping on higher frequencies may affect the transcription accuracy of individuals with higher-pitched voices.

**Awards Won:**