# REVOICE: A Real-Time Speech-to-Text Text-to-Speech Destuttering Pipeline

Chen, Cynthia (School: Walter Payton College Preparatory High School)

Purpose or Problem: Millions of people suffer from stammering, and adult severe stammering patients have a low chance of recovery. Other temporary alleviation methods are not consistent or affordable. This project aims to design and create a real-time destuttering system for public presentation using audio stutter classification and speech-to-text audio cleanup to 1) create an easy speech therapy aid and 2) create the baseline for a future portable destuttering device for everyday life. Procedure: REVOICE consists of four parts: a server-client workflow, a custom-trained and tuned StutterNet classification model, the Destutterer class, and Retrieval-Based Voice Conversion (RVC). First, custom STT and TTS clients and a server were coded using the WhisperSTT and MeloTTS models as basis. Next, the SEP-28k+FluencyBank stuttering dataset was used to train and fine-tune the StutterNet model found in the interspeech2024 repository. Critical interspeech2024 pipeline errors were corrected and PR curve + beta value analysis was used to select the best threshold and beta values. Then, the Destutterer class was created to utilize StutterNet and destutter real-time via server input. Afterwards, a modified version of RVC was applied with additional tuning to preserve the user's original voice. Finally, latency optimization with queues and parallelization was applied. Conclusion: Overall, all types of stuttering met or exceeded design criteria, achieving destuttering rates as high as 95.45%, latencies as low as 2.72s, and little negative speech naturalness deviancy (largest magnitude 5.8%) . Future improvements include optimizing server-client lag, prosody, and combining the software with a Raspberry pi and an ANC system for a portable device.

**Awards Won:**
Third Award of $1,200