A Multi-model Experiment to Develop End-to-End Speaker-Independent Automatic Speech Recognition Solutions for Dysarthric Speech in Low-Resource Languages

Khetan, Pranet (School: Shiv Nadar School)

Dysarthria is a motor-speech disorder that impairs the control of muscles required for clear speech, creating serious communication barriers that cannot be fully resolved by existing Augmentative and Alternative Communication systems. To solve this problem, 1407 audio samples totalling to 42 minutes of Hindi Dysarthric speech were collected from 28 patients with various ailments using an algorithmic data collection script as well as a WebSocket-based frontend. This data collection effort marks the largest database of Hindi Dysarthric Speech. Then, non-stationary noise reduction and silence trimming were applied to the audio files, and a Chi-Square metric was used on Mel-spectrograms to filter out unusable samples. To address the challenge of limited data, a novel two-step data augmentation technique using speed variation and synthetic sentence generation expanded the dataset to over 22.2 hours of speech data. Three state-of-the-art models were trained and evaluated in-silico: Whisper-medium, Wav2Vec2.0-MMS, and a bidirectional LSTM limited to single words. Whisper demonstrated the best accuracy at a <10% Word-Error-Rate. A low-cost, ESP32-SoC-based device was developed for practical deployment, priced under Rs. 2000 (~$25). This compact prototype records audio, transmits it to a server for Automatic Speech Recognition (ASR) processing, and synthesises intelligible speech for playback, enabling clear communication for patients in real time. This research will mark the first open-source ASR framework for Hindi dysarthric speech, overcoming challenges like speaker-dependence, vocabulary limitations, and data scarcity. The developed system significantly enhances communication for dysarthric patients, empowering them to interact effectively in their native language.

Awards Won: