

Few-shot Dysarthric Speech Recognition with Text-to-Speech Data Augmentation

Enno Hermann, Mathew Magimai.-Doss

Speakers with dysarthria could particularly benefit from assistive speech technology, but are underserved by current automatic speech recognition (ASR) systems. The differences of dysarthric speech pose challenges, while recording large amounts of training data can be exhausting for patients. In this paper, we synthesise dysarthric speech with a FastSpeech 2-based multi-speaker text-to-speech (TTS) system for ASR data augmentation. We evaluate its few-shot capability by generating dysarthric speech with as few as 5 words from an unseen target speaker and then using it to train speaker-dependent ASR systems. The results indicated that, while the TTS output is not yet of sufficient quality, this could allow easy development of personalised acoustic models for new dysarthric speakers and domains in the future.