Abstract:

Title: Automatic detection of fluency disorder in children

The main goal of this project is to develop an automatic system for the detection of speech fluency disorder in children. The broader perspective of this work is to reach out to the remote and rural population via electronic device/ gadgets useful for the detection of speech fluency disorder at early stage of the childhood.

Speech fluency disorder mainly termed as stuttering or stammering is characterized by repetition of sounds, syllables, or words; prolongation of sounds; and interruptions in speech known as blocks. The speech sound is produced by the coordinated muscle movements involving breathing, phonation, and articulation i.e. movement of the throat, palate, tongue, and lips. Muscle movements are controlled by the brain. The main problem in stuttering /stammering is the failure of one or more sequence of articulatory movements responsible to produce a fluent speech as per the instruction by brain. These struggle behaviours reflected in speech sound produced by a person. Stuttering can make it difficult to communicate with other people, which then affects a person's quality of social life and interpersonal relationships.

In this work we focused on the automatic detection of speech fluency disorder via two staged approach. We trained GMM-HMM acoustic models with normal and speech disordered speech (disfluent and filler phones). We used this models at the first stage to segment the test speech utterances at phoneme levels using trained acoustic models. This provides phoneme level segmented speech with crude boundary. These phoneme boundaries can further be refined with boundary specific acoustic features to get more accurate timing information. The second stage is very crucial and is the detection and evaluation of fluency disorder. For the detection of speech fluency disorder, we used the orthographic transcription of the test prompts as a reference and the recognized phone segments are compared with DTW to detect the repetition of phoneme/words. Further the segment level analysis is carried out to detect prolongation of sounds and blocks. The main features used in this automatic detection are the comparison of segment duration, speech rate, stop gap duration, VC transition duration, frication duration, VOT, vowel duration, formants, signal energy level with the standard normal speech segments.