Methods of bangla speech recognition

English continuous speech can be recognized with accuracy rate more than 90%. Unfortunately, in Bangla, works on speech recognition is still in the very preliminary stage. Research work to recognize Bangla speech has been started around 2000. In early 2000s, researchers tended to use different signal processing techniques to extract the features of speech signal and used different classifiers for recognition. Since then different attempts have been made to recognize Bangla speech with high accuracy. Reference has attempted continuous Bangla speech recognition using HMM for pattern recognition with the assistance of stochastic language model. They have achieved 85% accuracy but their grammar is built using only 100 words. Muhammad et al. also used HMM with Mel-frequency cepstral coefficient (MFCC) features to classify only Bangla digits. Even with this small number of classes, it fails to recognize digits with similar pronunciations like 6 (six), 7 (seven), 8 (eight) and 9 (nine). This system fails to detect these similar digits because it depends on the spectrogram as feature and the spectrogram of those digits are very similar. Linear Predictive Coding (LPC) and Artificial Neural Networks (ANNs) was used in. The main limitation of this work is it has been done using only four Bangla words uttered by only four persons (two Male And two Female). The accuracy they got is also not clear from their paper. In LPC, GMM and MFCC was used for feature extraction and Dynamic Time Warping for matching. Their best model has got highest 84% recognition rate for one hundred Bangla word in normal room environment. Though ANN has been used in, it has been done for only ten (10) digits in a noise free environment. The average accuracy is 92% for speaker independent setting. In very recent time, has proposed a Bangla Phoneme Recognition system to help in the development of a Bangla Speech

Recognizer using a new Linear Predictive Cepstral Coefficient-based feature. Though it can detect phonemes with high accuracy (99%) this method is not an end to end speech recognizer. In several ANNs architecture was experimented for Bangla words, characters and digits recognition using end-to-end approach.

All the previous methods for Bangla speech recognition were developed for a noise free well controlled speech data and none of the methods experimented with continuous speech. They did not consider the variability in various parameters like speed, noise, regional dialect and loudness. The size of the data set for evaluation were also very small for all the works. In conclusion, not a single Bangla speech to text system has been found in the published literature that works very well with real life large scale data.