Priyanka Bansal

Department of Electronics & Communication Engineering, Faculty of Engineering & Technology, Jamia Millia Islamia, New Delhi

Topic: Performance Evaluation of Speaker Recognition in a Continuous Speech Recognition System using Some Novel Techniques

Supervisor: Prof. M. T. Beg

Keywords: Diarization, turbulence, structured, feature generation, real time speech

ABSTRACT

Speaker recognition technique was earlier used as the biometric system. In the recent years, speaker recognition systems are used in various complicated applications including emotion recognition, multi-speaker recognition, person confidence recognition etc. The accuracy of speaker recognition systems depends on the quality of speech. The capturing device, environmental disturbance and background noise are some of the factors that can affect the quality of speech. In this present work, an effective framework is presented to recognize the speaker from multi-speaker speech or conversation.

The main contribution of this research is to recognize the individual from multi-speaker speech. A multi-featured framework is designed to handle the environmental and real time challenges to recognize the speaker. The framework is defined with three integrated process stages. In first stage, the Gaussian Mixture Model and wavelet decomposition are applied to rectify the speech against background noise and instrumentation noise. Other than speech level impurities, the main challenge in such system is diarization. The participation contribution is extracted using frequency specific, dictionary specific and channel compensation scoring method. After this stage, distributed speech segment of each speaker is collected as individual signal block. These collected speaker specific segments are considered as actual input to the framework for speaker recognition. The training and testing data sets are obtained in the form of segmented speech. The multi-aspect driven feature processors are applied on this normalized speaker segmented dataset to generate the effective and relevant features. At this stage, wavelet shrinkage, i-vector scoring, acoustic scoring, log-likelihood measure and MFCC methods are applied to transform the segmented speech to feature set. This feature processor is made to work in frequency and time domain to acquire the effective features. In the final stage, weight evaluation method is applied on generated features and processed under probabilistic neural network to recognize the speaker.

Another contribution of this research work is the sub-model to recognize the individual speaker. This sub-model is defined to recognize the speaker based on character, word or sentence based speech. The model is defined to process the real time acquired speech to recognize the speaker. The acquired speech can be affected by various noise vectors including interference, background noise and disturbances. A wider speech rectification stage is applied to remove these different kinds of impurities. The DWT integrated band pass filter is applied to remove the high level background noise. The spectral subtraction method is applied to remove the instrumentation noise. The probabilistic Linear Predictive Coding is integrated for reduction of low level acoustic turbulence from speech signal. The low level interference noise is removed using Gaussian distance scoring method. After obtaining filtered speech, the multiple feature generators are applied collectively to generate the wider feature set. In this feature processing stage, frequency, structure and probabilistic features are generated. The block adaptive signal peak analysis is applied to generate the frequency specific features. To generate the structural features, the Independent Component Analysis based statistical measure is applied. The rule specific probabilistic evaluation is performed using Hidden Markov Model for effective feature generation. These composite feature processors are applied to transform the speech data set to feature data set. The feature generator is applied on both training and testing data set. At the final stage, the classifier is applied on this feature adaptive training set to generate the classification rules. At last, the fuzzy integrated SVM method is applied for effective rule formulation. These rules are applied on test set to recognize the speaker.

Both the models are analyzed and implemented on multiple training and testing sets. The comparative results are generated against Probabilistic Neural Network, Neural Network and Hidden Markov Model-Principal Component Analysis methods. As the training set is reduced to 25 speakers, the lower accuracy rate of 74.44% is achieved. The comparative results identified that the higher accuracy of 87.8% is achieved for a training set of 100 speakers. For the individual speaker recognition 78% accuracy is achieved while verifying the speaker based on character speech. For the sentence based data set, the model has provided the comparatively higher results. As the speeches are acquired from real environment, it is robust against various interferences and noise that infect the speech signal. The overall evaluation identified that the proposed work has improved the speaker recognition for both individual and multi-speaker speech.