ABSTRACT

While speaker dependent speech recognition systems have achieved close to 90% accuracy, the speaker independent speech recognition systems have poorer efficiency. Speech recognition systems used in real time applications involve complex algorithms for faithful recognition. In this thesis, we describe a Speaker Independent Speech Recognition System for Indian English, American English and British English speeches. The entire procedure is divided into four stages: the initial stage deals with the general processing of the speech input, the second stage deals with preprocessing of the input speech and learning of the sound units. The third stage performs phoneme recognition using two-level neural networks, Probabilistic Neural Network and Recurrent Neural Network. The fourth stage executes word recognition and text recognition from the string of phonemes employing Hidden Markov Model. The system is trained by Indian English speech consisting of 300 words uttered by 60 speakers. The test samples comprised 300 words spoken by a different set of 40 speakers. The recognition accuracy is found to be 89% on an average which is better than the previous results. The Speaker Independent Speech Recognition system was tested for 10 Indian English speakers live and showed a recognition rate of 76.8%, the higher error rate due to ambient noise. The Speaker Independent Speech Recognition system was also tested for isolated digits from 0 to 9 uttered by 3 speakers live and achieved a recognition rate of 98.3%.
Then, the Speaker Independent Speech Recognition System is trained by American English speech consisting of 250 words uttered by 50 speakers. The test samples comprised 250 words spoken by a different set of 30 speakers. The recognition accuracy is found to be 89.1% on an average which is better than the previous results. Further, the Speaker Independent Speech Recognition System is trained by British English speech consisting of 200 words uttered by 30 speakers. The test samples comprised 200 words spoken by a different set of 20 speakers. The recognition accuracy is found to be 92.3% on an average which is well above the previous results.