

ABSTRACT

Speech is a natural mode of communication among human beings. There is no need for speech processing for communication among human beings, but a human machine interface needs speech processing as it does not have the knowledge of speech perception. Speech processing can be divided into speech recognition, speech enhancement and speech analysis.

Speech recognition, deals with the analysis of the linguistic content of the speech signal. There are many aspects of speech recognition that are already well understood, yet the desired quality of speech recognition is not achieved. There are many aspects of speech recognition that are already well understood, yet the human quality of speech recognition is still not achieved. Modern speech recognition systems for sentence recognition use various combinations of standard techniques over the basic approach. One such combination is the hybrid model. Two hybrid models are proposed in this thesis, which consist of Hidden Markov Model (HMM) as the front end that extracts features from the speech signal. The training and recognition systems constitute the back end of the hybrid model. The Radial Basis Function Neural Network (RBFNN) and the Fuzzy system serve as efficient methods for the back end. RBFNN is

ideally suited for pattern recognition applications. The two hybrid models HMM/RBFNN and HMM/Fuzzy are assessed for the sentences from TIMIT database. The performance measures used are recognition accuracy, False Acceptance Rate (FAR) and False Rejection Rate (FRR). 800 sentences (50% male and 50% female) from TIMIT database are used for the testing. The recognition accuracy obtained for the two hybrid models are higher than the highest accuracy results for sentence recognition reported so far 63.8% by researchers. The recognition accuracy for HMM/RBFNN is 93.3% and 89.8% for HMM/Fuzzy. The FAR and FRR are 33.3 and 3.7 respectively for HMM/RBFNN hybrid and 50.98 and 5.66 respectively for HMM/Fuzzy hybrid. These hybrid models are suitable for recognition systems involving large vocabulary speaker independent continuous sentence recognition.

Speech enhancement improves the perceptual quality of the speech signal by removing the destructive effects of noise. In applications, which require a microphone, the signal of interest is contaminated by background noise and reverberations. The presence of noise degrades the quality and intelligibility of speech. De-noising is performed before it is communicated or stored. Therefore, it is essential to have effective speech enhancement techniques to extract the desired speech signal from the corrupted noise. Noise reduction techniques are applied to many real world

applications from hearing aids, mobile phones, video-conferencing equipment to voice controlled systems involving human-machine interface.

This research focuses on the development of a modified speech enhancement algorithm for signals through a single channel microphone. Quality of the enhanced signals is evaluated using subjective objective and composite measures. Two existing algorithms namely multiband spectral subtraction algorithm and logMMSE statistical model based algorithm, which estimate the magnitude spectrum in the mean square sense, are taken for the comparison of performance. Data from NOIZEUS, a noisy speech corpus is used for evaluation.

A new noise suppression technique for speech enhancement is proposed using Neuro fuzzy classifier. It is a hybrid technique which has the ability to learn and is also good at making decisions based on linguistic rules. The performance of the technique can be adjusted by tuning the rules. A first-order Sugeno type fuzzy inference system with a five layer structure is adopted. The Adaptive Neuro Fuzzy Inference System (ANFIS) integrates the interpretability of a fuzzy inference system with the adaptability of a neural network. In the training phase, the hybrid learning algorithm is used to adapt to all parameters. Each step of the iterative learning algorithm has two parts. In the forward pass, the inputs go forward till layer 4 and the consequent parameters are identified by the least square

estimate, while the premise parameters are fixed. In the backward pass, the input patterns are propagated again. In each iteration, the error rates defined as the derivative of the squared error with respect to each node's output propagate backward and the premise parameters are updated by speeding up conjugate gradient descent while the consequent parameters remain fixed.

In the existing techniques like Spectral subtraction method, Statistical algorithm, Karhonen-Loeve Transform and Wiener filtering, speech quality is improved rather than speech intelligibility. These techniques do not estimate the signal to noise ratio. The proposed Neuro fuzzy classifier algorithm decomposes the input signal into Time-Frequency (TF) units and makes decisions using the neuro fuzzy classifier as to whether each TF unit is dominated by the target class, target-dominated class, masker-dominated class or the masker class. Considering it as a pattern recognition problem, three phases of operation are identified - the pre-processing, feature extraction and classification phases. The neuro fuzzy classifier takes up the training phase and enhancement phase. Initially, features are extracted from speech signal and are trained by neuro fuzzy classifier. In the enhancement phase, the trained classifier is used to classify the TF units of the noise-masked signal into four classes. Individual TF units of the noise-masked signal are multiplied with the

corresponding weight of the class and subsequently the enhanced speech waveform is reconstructed.

The performance of the proposed method is evaluated through comparisons of subjective, objective and composite measures. Validation is done by computing Pearson's correlation coefficient between subjective and objective measures as well as subjective and composite measures. To establish its potential for practical applications, the enhancement technique is demonstrated with speech recorded from nine persons who are implanted with voice prosthesis. This recording process has taken a period of four months.

This research also focuses on speech analysis where the paralinguistic features are determined. A dysfunction in the vocal cord causes this characteristic feature of a person to vary. The extent of degeneration can be assessed from the deviation of the parameters from that of a normal voice. This study also compares the alaryngeal voice with a normal voice to determine the extent to which the prosthetic voice resembles the normal voice.