ABSTRACT

This Thesis presents a detailed study on speech enhancement algorithm to provide robustness to the Automatic Speaker Recognizer (ASR) in real-life noisy conditions. The automatic speaker recognition technologies have developed into more and more important modern technologies required by many speech-aided applications. The main challenge for automatic speaker recognition is to deal with the variability of the environments and channels from where the speech was obtained. The main objective of this work is to attenuate the noise component of a noisy speech in order to enhance the quality of the speech processing devices and make them more robust under noisy conditions using wavelet based algorithms and to carry out a comprehensive evaluation and comparison of their performances on speaker recognition task.

The interest in the field of speech enhancement emerges from the increased usage of digital speech processing applications like mobile telephony, digital hearing aids and human-machine communication systems in our daily life. In all these speech communication systems the quality and intelligibility of speech is of utmost importance for ease and accuracy of information exchange. Both human and automatic speech communications are effective in controlled environments, due to the high quality and intelligibility of speech. However, in many situations of practical interest, speech signals are affected by the presence of undesirable background noise like multi-talker, street, station, car interior noise and so on. Thus speech enhancement subject to several background noise degradations is required to reduce the stress on the listener and ideally to increase intelligibility.
In this Thesis two new single channel wavelet based speech enhancement methods and a noise robust automatic speaker recognition are developed and reported. The proposed SE methods are capable of operating at a very low signal-to-noise ratio and to improve on existing single-microphone schemes for an extended range of noise types and noise levels, thereby making these suggested methods more suitable for speech processing applications like speaker recognition and speech coder. The proposed enhancement techniques are promising for coping with real life noise of various kinds. Firstly, a technique using Time Adaptive Discrete Wavelet Thresholding (TADWT) based on Bionic Wavelet Transform (BWT) is proposed. In this approach discrete BWT is used for speech enhancement task, the adaptive nature of the BWT is captured by introducing a time varying linear factor at each scale over time and modified soft thresholding function is used for denoising. Because of the time adaptation nature of wavelet coefficients and modified soft thresholding, significant improvement is obtained in speech quality as well as in intelligibility. This method also provides a good auditory representation (sufficient frequency resolution), good perceptual quality of speech and low computational load.

To meet the demand for quality noise reduction algorithms, capable of operating at a very low signal-to-noise ratio the second method is proposed by integrating Spectral Subtraction and Wavelet Packet based Thresholding (SSWPT) method. This hybrid approach gave improved spectral performance which reflects in better speech quality almost in all types of noise considered. In this approach, Dual Band Spectral Subtraction method (DBSS) with adaptive noise estimation is used as the pre-estimator for reducing the noise initially. Further by thresholding the wavelet packet transform (WPT)
coefficients of the preprocessed speech, improved speech quality is achieved. This results in more components representing the signal and provides more flexibility, which makes the improvement in noise reduction and spectral performance. But it is achieved by increased computational complexity to some extent.

In this thesis, basic spectral subtraction (SS), iterative Wiener filtering (IWF), Ephraim Malah filtering (EMF), Bionic wavelet based thresholding (BWT) techniques and Perceptual wavelet packet transform (PWPT) have been used as baseline methods for speech enhancement tests. Performance evaluation of proposed methods is made based on segmental signal to noise ratio (SSNR), signal to noise ratio (SNR), Itakura-Saito (IS) distance measure and minimum mean square error (MMSE) for the objective speech quality evaluation. In addition the performance of the proposed methods is evaluated using subjective quality measure by collecting MOS through listening test.

The objective quality measures showed that the proposed TADWT and SSWPT wavelet based thresholding methods provides improved performance than the base line methods considered for comparison. The two proposed algorithms are competitive to each other but based on computational scale TADWT obviously stands out as the best method than SSWPT. TADWT method seems to be better at lower SNRs for the noises like car, station, babble, HF channel than SSWPT. For the noises like air port, restaurant, exhibition noise SSWPT seems to be better than TADWT. For white, pink, train, street noise cases the performances of TADWT and SSWPT are almost similar to each other. The most difficult noises to suppress are F-16, factory and airport noise.
From the results of the MOS test, one point which stands out very clearly is that all the speech enhancement algorithms under test do improve the perceived speech quality. The algorithms that consistently scores well are TADWT and SSWPT. However it should be noted that SSWPT incurs a significant delay time and hence it is only suitable for non real time processing. On the other hand, the TADWT algorithm is suitable for real time processing of speech signals. Results based on objective measures and subjective experiments by means of a listening test, show that the system based on contributions of this thesis improves significantly over the existing ones.

The effectiveness of the proposed speech enhancement techniques is evaluated by means of speaker recognition under noisy conditions. In addition to the improvement in noise robustness, the recognition accuracy is improved by the use of combined feature extraction techniques in the proposed Automatic Speaker Recognition (ASR) system. The recognition accuracy of the suggested speaker recognition system under clean condition is found to be 96.67%. The performance of the ASR system has been evaluated in noisy environment by degrading test speech. The average recognition accuracy of the system is improved while incorporating SE methods as preprocessor while comparing with the recognition rate obtained for degraded speech. Irrespective of noise types and SNR level considered in this task, the overall average recognition accuracy of the system results as 12.8%, 40.2%, 37.4%, 47.4%, 46.9%, 57.4% and 57.65% for SS, IWF, EMF, BWT, TADWT and SSWPT respectively. The overall recognition results show that the proposed TADWT and SSWPT methods in this Thesis yield relatively higher performance than other methods.