Abstract

This thesis details the development and evaluation of a wavelet-based system for enhancement of speech corrupted by non-stationary and real-life noise, when only the corrupted signal is available.

First, a survey and comprehensive review of most recent and commonly used speech enhancement techniques, for both spectral and wavelet domains, are presented. After indicating the main shortcomings of these techniques, wavelet-based in particular, a new speech enhancement system for slowly varying white and coloured noise is then proposed and evaluated. The system uses two approaches. In the first, an auditory-based perceptual signal decomposition model using Bark-scaled wavelet packet is employed in conjunction with an adaptive noise suppression filter whose gain is adjusted according to an accurate estimation of per segment noise level.

The second approach is based on an adaptive wavelet thresholding technique, which used a novel multi-feature soft-thresholding gain function. The gain function is determined for each wavelet band of each speech segment based on accurate estimation of segmental signal-to-noise (SegSNR) ratio and whether the processed segment is voiced or unvoiced.

To facilitate the implementation of the above two speech enhancement approaches, the work involved the following three initial packages, as detailed in the thesis: (a) A through knowledge and understanding of the wavelet transform principles, types and implementation schemes, coupled with identification of the advantages of various wavelet decomposition models in terms of signal resolution and computational efficiency. (b) Investigation and performance evaluation of various noise power estimation algorithms, when implemented using two different wavelet decomposition schemes: the perceptual wavelet transform, and the second-generation wavelet transform. Based on this, a new efficient and accurate wavelet-based noise estimation algorithm has been formulated; and its performance evaluated in comparison to those of the other three algorithms. Based on this evaluation, our proposed speech enhancement system has been implemented using three different noise estimation algorithms, which were shown to be the most accurate; and (c) A new algorithm for classification of speech into voiced/unvoiced segments has been developed. The classification algorithm is based on estimation of the wavelet-band energy distribution of the noisy speech segment, and the per segment zero-crossing rate.