The first stage of speech recognition is compressing a speech signal into streams of acoustic feature vectors, referred to as speech feature vectors. The extracted vectors are assumed to have sufficient information and to be compact enough for efficient recognition [12].The concept of feature extraction is actually divided into two parts: first is transforming the speech signal into feature vectors; secondly is to choose the useful features which are insensitive to changes of environmental conditions and speech variation [11].

The present system is based on converting the source wave file into a speech signal and then extracting first 12 MFCCs for each frames of the converted signal. But only the first 12 of the DCT coefficients are kept. This is as a result of the upper DCT coefficients mean quick changes within the filterbank energies and it seems that these quick changes truly degrade ASR performance, therefore higher classification results will get by dropping them.