Abstract

Automatic Speech Recognition (ASR) has ability to make a computer recognize what was said by someone just from their voice signal. With that ability it can be used to recognize when someone pronounce word inaccurately. The method used in this thesis are Mel frequency cepstral coefficient (MFCC) for feature extraction. It will change the value of the amplitude into a row of frames which will be processed using mel-filterbank that adapting the workings of the human auditory to forming coefficient values that become feature represent each word. Results of MFCC are then quantized into a codebook that would become an input for Hidden Markov Models (HMM) to be modeled. Feature extraction result of test data are then quantized to be identified using a model that obtained before. Tests carried out by using 10 pairs of words with high degree of similarity and often confusedly pronounced. From the test results, the obtained average accuracy rate for each pair of words is 78.3% using 3-state HMM models and 81.67% using 5-state HMM models.

Keywords : Automatic Speech Recognition, MFCC, HMM