

ABSTRACT

Speech signal information has a unique characteristic. Many researches in the speech signal processing field has done. One of the results produced is speech-to-text application. This application is a branch of speech recognition application which processes a speech signal, recognizes it, and changes it into textual representation. Generally, phonemes used in the speech-to-text system as a recognizer unit, but the problem is located in the certain phonemes classes which has a short duration. In order to solve this problem, speech recognition method using diphone (phonemes combination) database as an alternate recognition unit is developed.

In this final project, diphones used as a reference database in the speech recognition system is being explored. The diphone itself is easier to be recognized than the phoneme, since it gives more acoustics information. The classification method used here are Hybrid Hidden Markov Model and Genetic Algorithm (HMMGA). Three different methods used in the speech signal feature extraction, those are LPC (Linier Predictive Coding), MFCC (Mel Frequency Cepstral Coefficient) and HFCC (Human Factor Cepstral Coefficient).

The testing used three speech signal feature extraction methods on the HMM system for word recognizing; these methods resulting 85.56 % (LPC), 57.78 % (MFCC), and 66.67 % (HFCC) as the best accuracy, while for the diphone recognizing is 92.22 % (LPC), 76.67 % (MFCC), and 80.00 % (HFCC). On the other hand, the testing used hybrid HMMGA system showing the accuracy increasing about 3-15 % measured from the worst HMM.

Key words: speech recognition, speect-to-text, hybrid Hidden Markov Model and Genetic Algorithm (HMMGA), Linier Predictive Coding (LPC), Mel Frequency Cepstral Coefficient (MFCC), Human Factor Cepstral Coefficient (HFCC).