ABSTRACT

In this globalization era, one of the effects is the formation of international languages, namely English. To deal with these problems as well as being able to provide a foundation for the younger generation, the Speech to Text application was created with the aim of introducing English. This application will be useful for introducing words in English. Besides being able to multiply vocabulary, this application is also intended to practice reciting these words in English.

This Speech to Text (STT) application uses the Hybird Hidden Markov Models (HMM) method with the Gaussian Mixture Model (GMM). The initial stage of Hidden Markov Models is sound compilation, then the sound will be recognized as a Speech Signal. Then using the extraction feature that is Mel-cepstral frequency coefficient (MFCC) the signal is stored in the frames and the cepstral coefficient value is searched. Furthermore, each vector is quantized to produce an output symbol (codebook). Every unknown word will be modeled with HMM / GMM so that it gets a word model. For the word recognition process, the probability of similarity of patterns from each HMM / GMM model will be calculated with the results of the observations. The best maximum probability results are then defined as words that are recognized.

This test is done by changing the value of MFCC features and GMM mixture values. System performance is measured based on the accuracy obtained from the WER (Word Error Rate) parameter. After testing the system with several scenarios, the best accuracy is 100% in recognizing 10 words. This accuracy was obtained from the test results with MFCC 13 Feature and GMM 6 mixture.

Key Word: *Speech to Text, Hidden Markov Model, Mel-frequency cepstral coefficient, Gaussian Mixture Model*