ABSTRACT

As information technology growing further, educational system is more likely using alternative media. One of the alternative educational media that has been widely used is video. The get better understanding, usually the educational video is included with caption or text that explaining what the speaker says. It will be much efficient in time and energy if the caption is generated automatically based on what speaker says. Therefore, this research will design and implement Automatic Video Captioning system using speech recognition technology.

Speech signal from video is extracted and processed with speech recognition system to generate corresponding text. In this final assessment, the speech recognition system is designed with Linear Predictive Coding as the feature extraction method and Hidden Markov Model as the feature matching method. The generated text from speech recognition system is then used as the caption for video input.

The system is tested by changing the number of data training and the HMM parameters (the number of states and clusters) to find the most optimal parameter with highest accuracy. According the test, the highest accuracy is found at 75,50% when the number of states is 6, number of clusters is 256, and the number of data training is 90 for every syllable in database.

Keywords: *Automatic Video Captioning*, *Speech recognition*, *Linear Predictive Coding* (LPC), *Hidden Markov Model* (HMM)