Abstract

This thesis presents the overall of the retrieval of Thai broadcast news' voice. The corpus is taken from NECTEC for creating LVCSR to recognize the speech-based data into keyword. Later, the keyword has been searched for transcribe news. The result will display in the pattern of audio files containing with the keywords. Furthermore, this project also develops train-set data which is suitable for test-set data by using LOTUS Corpus for training acoustic model and BN Corpus for training language model. From the experiment, it reveals that correctness is 78.32% and accuracy is 74.34% which is the appropriate result. Moreover, the 10-best and sub keyword is brought to accelerate the better result.