

Abstract

Speech Recognition can be performed by representing speech signal into wave form or into parameter form. In this Final Project, speech signal is represented into parameter form where feature vector speech signal are extracted to retrieve the information signal then create the model for each word that wants to be recognized.

Each signal has different characteristic. This can be observed if we zoom in the signal wave around 10-30 ms. The short speech signals are stored into frames and then they are analyzed to retrieve the characteristic using LPC (Linear Predictive Coding).

The feature vectors are quantized using k-means cluster method to produce the sequence of observation. Each sequence observation builds different sequent states. Each word that wants to be recognized is modeled by HMM architecture. For testing, each word is evaluated to each model. The model which is closest to the word is the model that we are looking for. The result shows that the proper selection of codebook size can provide data accuracy above 90%, while the number of state had little influence on the accuracy of the data.

Key Words: *speech-to-text, Linear Predictive Coding (LPC), Hidden Markov Model (HMM), codebook size, number of state*