

Abstract

This thesis investigates the methods for English speech evaluation, which requires the techniques from audio signal processing and speech recognition. In order to develop an appropriate and consistent English speech evaluative system, we design three steps for our English speech evaluation. The first one is “Utterance Verification”, the second one is “Forced Alignment”, and the final one is “English Speech Evaluation”.

“Utterance Verification” can be seen as the firewalls for English speech evaluation system. The purpose of utterance verification is to determine if the test utterance is close enough to the target sentence. If the confidence measure is too low, we may reject the utterance directly since further similarity computation may not be meaningful.

“Forced Alignment” is a method based on Viterbi decoding, which can obtain time duration of each phoneme from a given utterance. In this part, we use a large-vocabulary speaker-independent recognition engine and Viterbi algorithm to obtain the duration of each phoneme.

“English Speech Evaluation” is the core of our evaluation system, which can compute the similarity between the reference and test utterances. We use magnitude contour, pitch contour, rhythm, and HMM log-likelihood as the features to generate a similarity score. The parametric scoring function is optimized by down-hill Simplex method to minimize the difference between the computed scores and those by human experts.