Fisseha Welday Atsbaha, Study of stop landmark durations for speaker recognition, M. Tech. Thesis, Department of Electrical Engineering, Indian Institute of Technology Bombay, 2009.

Supervisor(s): P. C. Pandey

Abstract **-** The objective of this project is to investigate the use of stop landmark durations for improving speaker recognition. The variations of stop closure and burst durations across speakers are studied using variance tests. The results indicate that stop closure and burst durations may be used in combination with the spectral features for improving speaker recognition. Rate-of-rise (ROC) of mel-filtered squared magnitude spectrum is investigated for locating spectral transitions in the speech signal. Mel-filtering along the frequency axis improves landmark detection by enhancing the perceptually significant spectral transitions and smoothing harmonic structure and noise. Two automated methods for detecting stop closure, burst and frication offset landmarks are developed based on the ROC of mel-filtered spectrum. In the first method, the ROC peaks are selected using peak picking algorithm based on local threshold, and spectral slope, Wiener entropy and average magnitude spectrum are used as additional features to detect stop landmarks. In the second method, closure intervals in the speech signal are located based the product of Wiener entropy and log energy, and stop landmarks are detected by picking ROC peaks around the end points of the closures. The Wiener entropy and log energy are computed from the magnitude spectrum, while the spectral slope is computed from the mel-filtered squared magnitude spectrum. Landmark detection tests are carried out on VCV syllables and TIMIT sentences. Stop landmarks in VCVs are detected at rates of 53%, 75%, 90%, 95% and 97% respectively within 3, 5, 10, 15 and 30 ms of the manually labeled landmarks. The detection rates for TIMIT sentences are 52%, 69%, 83%, 87%, 90% and 93% within 3, 5, 10, 15, 20 and 30 ms of the manual landmarks, respectively.

Text-independent speaker recognition tests are conducted using Gaussian mixture modeling of closure and burst durations, and MFCC parameters. The performance of the duration features alone is not satisfactory, but an improvement of up to 4% is obtained using combination of MFCC and duration features. The results indicate that stop closure and burst durations convey speaker-dependent information and they could be potential candidates for improving speaker recognition.